

Stichproben nach § 42 RSAV

Gutachten im Auftrag des Bundesversicherungsamtes

Endfassung unter Berücksichtigung der Ergebnisse der Anhörung des Spitzenverbandes Bund der Krankenkassen sowie der Prüfdienste des Bundes und der Länder

Prof. Dr. Thomas Schäfer
Oberuhldingen, im April 2013

Inhaltsverzeichnis

Zusammenfassung	1
1 Hintergrund und Aufgabenstellung	7
1.1 Hintergrund	7
1.2 Aufgabenstellung	8
1.3 Zu prüfende Angaben (Satzarten)	8
1.3.1 Prüfung nach § 42 Abs. 1 S. 1 Nr. 1 RSAV – Versichertenzeiten	8
1.3.2 Prüfung nach § 42 Abs. 1 S. 1 Nr. 2 RSAV – Morbiditätsdaten	8
1.4 Verwendung der zu prüfenden Angaben im Morbi-RSA	9
1.5 Konsequenz der Prüfungen nach § 42 RSAV	11
2 Mögliche Stichprobenverfahren und Struktur des Gutachtens	13
2.1 Grundbegriffe der Stichprobentheorie	13
2.2 Taxonomie der Stichprobenverfahren	14
2.3 Die einfache Zufallsauswahl	14
2.4 Die geschichtete Zufallsauswahl	15
2.5 Sequentielle Zufallsauswahlverfahren	15
2.6 Für Stichproben nach § 42 RSAV ungeeignete Auswahlverfahren	17
2.7 Struktur des Gutachtens	18
3 Planung und Hochrechnung einer einfachen Zufallsstichprobe	19
3.1 Bezeichnungen	19
3.2 Mittlerer Korrekturbetrag und Summe der Korrekturbeträge	19
3.3 Zur Verteilung des Korrekturbetrages	20
3.4 Benötigter Stichprobenumfang zur Erreichung einer vorgegebenen Präzision der Schätzung des mittleren Korrekturbetrages	23
3.5 Anteil der Versicherten mit Fehlern in den übermittelten Daten	28
3.5.1 Relative Fehler bei der Schätzung ausgewählter Kennziffern: Nr. 1-Stichproben in der Simulation	30
3.5.2 Relative Fehler bei der Schätzung ausgewählter Kennziffern: Nr. 2-Stichproben in der Simulation	31
3.6 Erwarteter Stichprobenumfang bei sequentieller Schätzung der Fehlerquote	32

4	Planung und Hochrechnung einer geschichteten Zufallsstichprobe	35
4.1	Bezeichnungen	35
4.2	Schätzung des Mittelwertes und der Varianz	35
4.3	Potentielle Schichtungsvariablen	36
4.4	Wahl der Schichtenzahl und der Schichtgrenzen	37
4.5	Aufteilung des Gesamtstichprobenumfangs auf die Schichten	40
4.6	Abschätzung des erreichbaren Schichtungseffekts.....	42
4.7	Nachträgliche Schichtung.....	44
4.8	Vorteile einer Schichtung mit Neymann-Allokation bei Erfassung teurer HMGs.....	45
5	Partielle Sequentialverfahren: Zweistufige Stichprobenziehung	51
5.1	Steins Zweistichprobenplan zur Schätzung des Mittelwertes einer Normalverteilung mit vorgegebener Präzision.....	51
5.2	Der Zweistichprobenplan von Cox zur Schätzung eines Anteilswertes mit einem vorgegebenem Variationskoeffizienten.....	52
5.3	Anscombes zentraler Grenzwertsatz für Sequentialverfahren	54
6	Überprüfung der Stichproben auf Qualität.....	55
7	Resümee und Verfahrensvorschlag.....	59
7.1	Planungsrelevante Vorkenntnisse	59
7.2	Verfahrensimplizante Kernprobleme der Prüfungen nach § 42 RSAV.....	59
7.3	Zweistufiges Prüfverfahren mit Schwellenwert für die Fehlerquote.....	60
7.4	Verfahrensvorschlag zur Vorbereitung der erstmaligen Prüfung der Krankenkassen („Pilotphase“).....	62
7.4.1	Nr. 1-Stichproben: Auswertung der Ergebnisse der Prüfungen nach § 15a RSAV	62
7.4.2	Nr. 2-Stichproben: Vorbereitungsstichproben im Rahmen simulierter Prüfverfahren für ausgewählte Beispielskassen	62
7.5	Planung und Durchführung des ersten Prüfzyklus	63
7.6	Laufende Überprüfung und Weiterentwicklung der Verfahren in nachfolgenden Prüfzyklen	66

8	Anhang	67
8.1	Beispielskassen für die Stichprobenplanung nach § 42 Abs. 1 S. 1 Nr. 1 RSAV	67
8.2	Beispielskassen für die Stichprobenplanung nach § 42 Abs. 1 S. 1 Nr. 2 RSAV	69
8.3	Prüfungsaffine Kennziffern für die Beispielskassen und alle im Ausgleichsjahr 2009 am Morbi-RSA beteiligten Krankenkassen.....	72
	8.3.1 Grundgesamtheiten für die Nr. 1-Stichproben	72
	8.3.2 Grundgesamtheiten für die Nr. 2-Stichproben	73
8.4	Stichprobenumfangsplanung in Abhängigkeit vom Variationskoeffizienten..	74
9	Literatur/Quellen	75

Tabellenverzeichnis

Tabelle 1-1	Verwendung der zu prüfenden Angaben im Morbi-RSA	10
Tabelle 1-2	Konsequenz der Prüfungen nach § 42 Abs. 1 S. 1 Nr. 1 RSAV	11
Tabelle 1-3	Konsequenz der Prüfungen nach § 42 Abs. 1 S. 1 Nr. 2 RSAV	12
Tabelle 3-1	Wichtige Kennziffern der Grundgesamtheit und analoge Stichprobengrößen	19
Tabelle 3-2	Abhängigkeit des Variationskoeffizienten V des Korrekturbetrages von der Fehlerquote und dem Variationskoeffizienten q innerhalb von VG_F ...	23
Tabelle 3-3	Variationskoeffizient des Korrekturbetrages aus der Simulation I in den zehn Beispielskassen	26
Tabelle 3-4	Mittlerer Variationskoeffizient des Korrekturbetrages aus Simulationen II bis IV in den zehn Beispielskassen	26
Tabelle 3-5	Erforderliche Stichprobenumfänge in den fünf Beispielskassen für die Nr. 1-Stichproben zur Schätzung des KBs in Abhängigkeit von der Fehlerquote p_F und der gewünschten (relativen) Genauigkeit ε ($q=0,6$)...	27
Tabelle 3-6	Erforderliche Stichprobenumfänge in den fünf Beispielskassen für die Nr. 2-Stichproben zur Schätzung des KBs in Abhängigkeit von der Fehlerquote p_F und der gewünschten (relativen) Genauigkeit ε ($q=0,6$) ..	27
Tabelle 3-7	Erforderliche Stichprobenumfänge in den fünf Beispielskassen für die Nr. 1-Stichproben in Abhängigkeit vom zu schätzenden Fehleranteil p_F und der gewünschten (relativen) Genauigkeit ε	29
Tabelle 3-8	Erforderliche Stichprobenumfänge in den fünf Beispielskassen für die Nr. 2-Stichproben in Abhängigkeit vom zu schätzenden Fehleranteil p_F und der gewünschten (relativen) Genauigkeit ε	29
Tabelle 3-9	Relative Abweichung der aus der Stichprobe von den aus der Grundgesamtheit berechneten ausgewählten Kennziffern (Nr. 1- Stichprobe)	31
Tabelle 3-10	Relative Abweichung der aus der Stichprobe von den aus der Grundgesamtheit berechneten ausgewählten Kennziffern (Nr. 2- Stichprobe)	32
Tabelle 4-1	Korrelation der mittleren Zuweisung mit dem Mittelwert ausgewählter Variablen (als Merkmale der 20 Schichten) für die Versicherten der GKV-insgesamt, Nr. 1-Stichprobe	37

Tabelle 4-2	Korrelation der mittleren Zuweisung mit dem Mittelwert ausgewählter Variablen (als Merkmale der 20 Schichten) für die Versicherten der GKV-insgesamt, Nr. 2-Stichprobe	37
Tabelle 4-3	GKV-Gesamt (in Bezug auf Nr. 1- Stichprobe), neu geschichtet	40
Tabelle 4-4	GKV-Gesamt (in Bezug auf Nr. 2- Stichprobe), neu geschichtet	40
Tabelle 4-5	Stichprobenumfänge nach Neymann-Allokation für die (neu) geschichteten Stichproben bei der Zuweisungssumme proportionalem Korrekturbetrag.....	43
Tabelle 4-6	Schichtungseffekte im Best-Case-Szenario für die Planung	43
Tabelle 4-7	Stichprobenumfänge nach Neymann-Allokation zur Erhebung des in der Simulation I erzeugten Korrekturbetrages	44
Tabelle 4-8	Schichtungseffekte bei einer Verteilung des Korrekturbetrages gemäß Simulation I	44
Tabelle 4-9	Die zehn teuersten HMGs im Klassifikationsmodell 2009, nach absteigender Zuschlagshöhe geordnet.....	45
Tabelle 4-10	Die zehn seltensten HMGs im Klassifikationsmodell 2009, nach aufsteigender Häufigkeit geordnet.....	46
Tabelle 4-11	Häufigkeit der zehn teuersten HMGs (in absteigender Zuschlagshöhe) und ausgewählte Charakteristika in den Beispielskassen F bis J	46
Tabelle 4-12	Häufigkeit der zehn teuersten HMGs (in absteigender Zuschlagshöhe) und ausgewählte Charakteristika in den einfachen Zufallsstichproben aus den Beispielskassen F bis J (geplant für $p_F=0,05$ und $\varepsilon =10\%$)	47
Tabelle 4-13	Schichtgrenzen der fünf Schichten bei Schichtung der Stichproben aus den Beispielskassen F bis J nach Zuweisungshöhe	47
Tabelle 4-14	Häufigkeit der zehn teuersten HMGs (in absteigender Zuschlagshöhe) und ausgewählte Charakteristika in den nach der Zuweisungssumme geschichteten Stichproben mit Neymann-Allokation aus den Beispielskassen F bis J (geplant für $p_F=0,05$ und $\varepsilon =10\%$)	48
Tabelle 4-15	Ausschöpfungsquoten der zehn teuersten HMGs in den nach der Zuweisungssumme geschichteten Stichproben mit Neymann-Allokation aus den Beispielskassen F bis J (geplant für $p_F=0,05$ und $\varepsilon =10\%$)	48
Tabelle 4-16	Relative Abweichungen ausgewählter, aus den Stichproben geschätzter Charakteristika der Beispielskassen von den entsprechenden Werten der Grundgesamtheiten	49

Tabelle 6-1	Ausgewählte Anteilswerte in den Nr. 1-Grundgesamtheiten der Beispielskassen und relative Abweichungen bei Schätzung aus der Simulation III-Stichprobe	57
Tabelle 6-2	Erweiterte Zuweisungen für zwei ausgewählte HMGs in den Nr. 1-Grund-gesamtheiten der Beispielskassen und relative Abweichungen bei Schätzung aus der Simulation III-Stichprobe	57
Tabelle 6-3	Ausgewählte Anteilswerte in den Nr. 2-Grundgesamtheiten der Beispielskassen und relative Abweichungen bei Schätzung aus der Simulation III-Stichprobe	58
Tabelle 6-4	Erweiterte Zuweisungen für zwei ausgewählte HMGs in den Nr. 2-Grund-gesamtheiten der Beispielskassen und relative Abweichungen bei Schätzung aus der Simulation III-Stichprobe	58
Tabelle 8-1	Beispielskasse A nach ausgewählten Kennziffern	67
Tabelle 8-2	Beispielskasse B nach ausgewählten Kennziffern	67
Tabelle 8-3	Beispielskasse C nach ausgewählten Kennziffern	68
Tabelle 8-4	Beispielskasse D nach ausgewählten Kennziffern	68
Tabelle 8-5	Beispielskasse E nach ausgewählten Kennziffern	69
Tabelle 8-6	Beispielskasse F nach ausgewählten Kennziffern	69
Tabelle 8-7	Beispielskasse G nach ausgewählten Kennziffern	70
Tabelle 8-8	Beispielskasse H nach ausgewählten Kennziffern	70
Tabelle 8-9	Beispielskasse I nach ausgewählten Kennziffern	71
Tabelle 8-10	Beispielskasse J nach ausgewählten Kennziffern	71
Tabelle 8-11	Beispielskassen und GKV-Kassen nach prüfungsaffinen Kennziffern, Nr. 1-Stichproben	72
Tabelle 8-12	Beispielskassen und GKV-Kassen nach prüfungsaffinen Kennziffern, Nr. 2-Stichproben	73
Tabelle 8-13	Erforderliche Stichprobenumfänge in den fünf Beispielskassen für die Nr. 1-Stichproben in Abhängigkeit vom Variationskoeffizient (V) des Korrekturbeitrags und der gewünschten (relativen) Genauigkeit ϵ	74
Tabelle 8-14	Erforderliche Stichprobenumfänge in den fünf Beispielskassen für die Nr. 2-Stichproben in Abhängigkeit vom Variationskoeffizient (V) des Korrekturbeitrags und der gewünschten Genauigkeit ϵ	74

Abbildungsverzeichnis

Abbildung 4-1 Zuordnungsschema I für die Zusammenfassung der 20 Schichten 39

Abbildung 4-2 Zuordnungsschema II für die Zusammenfassung der 20 Schichten 39

Zusammenfassung

(1) Nach dem neu in die RSAV eingefügten § 42 sind die Datenmeldungen für den Morbi-RSA mindestens alle zwei Jahre auf ihre Richtigkeit zu prüfen. Die Prüfungen umfassen – grob zusammengefasst – einerseits a) die allgemeine und spezielle Anspruchsberechtigung im Ausgleichsjahr und andererseits b) die Korrektheit der Zuweisung zu den Risikogruppen des versichertenbezogenen Klassifikationssystems im Jahr vor dem Ausgleichsjahr. In beiden Fällen handelt es sich um Stichprobenprüfungen. Zur Reduzierung des notwendigen Prüfumfangs werden die Stichproben für die Prüfungen nach b) auf die Grundgesamtheit der Versicherten beschränkt, die einen prüfungsrelevanten Zuschlag erhalten haben. Die erstmalige Anwendung des Verfahrens bezieht sich auf das Ausgleichsjahr 2009. Das Bundesversicherungsamt (BVA) soll für jede der im Ausgleichsjahr 2009 am Morbi-RSA beteiligten Krankenkassen aus der Menge der gemeldeten Pseudonyme für die beiden Grundgesamtheiten nach a) und b) jeweils eine Stichprobe ziehen und Pseudonyme der Stichprobenversicherten den Prüfdiensten einschließlich der zu prüfenden Angaben übermitteln. Die Prüfdienste prüfen dann die Korrektheit der Angaben vor Ort bei den Krankenkassen und melden gefundene Fehler an das BVA zurück, das aus den Stichprobenergebnissen einen Korrekturbetrag auf den Gesamtbestand der geprüften Krankenkasse hochrechnet und einfordert.

(2) Die erstmalige Festlegung des Stichproben- und Hochrechnungsverfahrens soll nach § 42 Abs. 6 RSAV auf der Grundlage eines wissenschaftlichen Gutachtens erfolgen, das insbesondere Vorschläge für die Stichprobenmethodik, für die Sicherstellung der Stichprobengüte, für ein Hochrechnungsverfahren sowie für eine Weiterentwicklung dieser Verfahren in den nachfolgenden Prüfzyklen unterbreiten soll. Die Endfassung dieses Gutachtens wird hiermit vorgelegt.

(3) Da die Planung einer Stichprobe immer auf eine konkrete Grundgesamtheit hin erfolgt, über die einige Eckdaten bekannt sein müssen, kann das Gutachten nach § 42 Abs. 6 RSAV nur ein Rahmengutachten sein, das die Fragestellung exemplarisch für einige Beispielskassen löst. Das BVA hat hierfür fünf Krankenkassen ausgewählt, die nach Auffassung des Amtes in ihrer Gesamtheit das Spektrum der im Ausgleichsjahr 2009 am Morbi-RSA beteiligten Krankenkassen der GKV gut abbilden. Aus dem Versichertenbestand jeder dieser Krankenkassen wurden vom BVA zum Zwecke der Anonymisierung zwei Stichproben mit einem Umfang von ca. 80% gezogen, in der ausgewählte Versichertenmerkmale dokumentiert sind (eine bildet die Grundgesamtheit für die Prüfungen nach a) die andere für die Prüfungen nach b)). Die Daten wurden dem Gutachter in zwanzig Schichten aggregiert übergeben, die keine versichertenbezogenen Merkmale enthalten, sondern durch die zugehörigen Anteilswerte, Mittelwerte und Standardabweichungen charakterisiert sind.

(4) Zwei Größen stehen für die Planung der Stichproben im Vordergrund des Interesses. Die erste ist der Anteil der Pseudonyme einer Kasse mit Fehlern in den an das BVA übermittelten Datensätzen (im Folgenden kurz auch als Fehlerquote und mit p_F bezeichnet). Die zweite hängt eng mit dem Korrekturbetrag zusammen. Es ist zweckmä-

ßig, ein versicherten-(bzw. pseudonym-)bezogenes Merkmal „Korrekturbetrag“ zu definieren, das für alle Pseudonyme ohne Fehler in den übermittelten Datensätzen eine 0 als Merkmalsausprägung zugewiesen bekommt.

(5) Die Hochrechnung des Gesamtbetrages, der sich aufgrund der Korrekturen in den Stichproben ergibt, auf die jeweilige Grundgesamtheit, erfolgt in zwei Schritten: Im ersten Schritt wird der Mittelwert des Korrekturbetrages in der Grundgesamtheit aus der Stichprobe designadäquat geschätzt. Im zweiten Schritt wird der Mittelwert mit der Zahl der Versicherten in der jeweiligen Grundgesamtheit multipliziert.

(6) Aus den Prüfungen nach § 15a RSAV liegen Ergebnisse über die Fehlerquoten vor, die mit gewissen Einschränkungen als Vorkenntnisse für die in Ziffer (1) unter a) genannten Prüfungen gelten können. Was die Planung der Stichprobe und insbesondere des Stichprobenumfangs für die in Ziffer (1) unter b) genannten Prüfungen betrifft, so gibt es gesichertes Vorwissen weder über p_F , noch über die Verteilung des Merkmals „Korrekturbetrag“. Man ist auf Annahmen angewiesen oder auf Simulationen, wobei letzteren wiederum Annahmen zugrunde gelegt werden müssen, von denen man auch nicht weiß, ob sie die Realität zutreffend abbilden. Eine Annahme, die plausibel erscheint und über die vermutlich Einigkeit erzielt werden kann, ist diejenige, dass p_F vergleichsweise klein sein wird (z. B. $< 10\%$). Im empirischen Teil des vorliegenden Gutachten wurde zumeist von $p_F = 5\%$ ausgegangen.

(7) Im Wesentlichen kommen nur zwei Stichprobenverfahren mit jeweils in natürlicher Weise verbundenen Schätzverfahren in Betracht: die einfache und die geschichtete Zufallsstichprobe. Vollständig sequentielle Verfahren scheiden aus organisatorischen Gründen aus, weil sie nur mit einer größeren Zahl von Kommunikationsakten zwischen dem BVA und den Prüfdiensten zu realisieren wären. Aus praktischen Erwägungen heraus wurde jedoch ein zweistufiges Sequentialverfahren in das Stichprobenkonzept mit aufgenommen, dessen theoretischen Grundlagen für kleine Stichproben weniger gut erforscht sind, als die der klassischen Stichprobenverfahren mit vorgegebenem festen Stichprobenumfang, aber dessen Anwendung für die Stichproben nach § 42 RSAV unter Bezugnahme auf asymptotische Eigenschaften als gerechtfertigt erscheint.

(8) In Rahmen der Planung einer einfachen Zufallsstichprobe ist es üblich, den Stichprobenumfang so zu festzusetzen, dass der Mittelwert des Erhebungsmerkmals (hier der Korrekturbetrag) mit 95% Wahrscheinlichkeit mit einer maximalen relativen Abweichung von ε geschätzt werden kann (z. B. $\varepsilon = 20\%$). Der hierfür benötigte Stichprobenumfang n hängt wesentlich von dem Variationskoeffizienten V der Verteilung des Erhebungsmerkmals ab ($V = \text{Standardabweichung}/\text{Mittelwert}$, eine dimensionslose Größe). Er wächst näherungsweise mit dem Quadrat von V . So ergibt sich beispielsweise für eine Krankenkasse mit 1,5 Mio. Versicherten, wenn $\varepsilon = 20\%$ gewählt wird, $n=864$ für $V=3$, aber $n=3.449$ für $V=6$.

(9) Es gibt Gründe für die Annahme, dass der Variationskoeffizient des Korrekturbetrages hoch ist. Das Merkmal „Korrekturbetrag“ ist im statistischen Sinn hochgradig pathologisch. Wenn etwa $p_F = 5\%$ gilt, so haben nur 5% der Pseudonyme eine von Null ver-

schiedene Merkmalsausprägung. Damit ist z. B. der Median = 0 €. Die extreme Schiefe der Verteilung resultiert in einer erheblichen Aufblähung der Varianz, was schließlich auch zu einem ungewöhnlich hohen Variationskoeffizienten führt. In Simulationen mit Daten der Beispielskassen wurden Werte zwischen 6 und 9 für den Variationskoeffizienten des Korrekturbetrages gefunden. Darüber hinaus konnte eine Beziehung zwischen der Fehlerquote und dem Variationskoeffizienten etabliert werden, welche für kleine Fehlerquoten unter 3% hohe Variationskoeffizienten in einer Größenordnung zwischen 7 bis 11 bestätigt.

(10) Durch Schichtung der Grundgesamtheit kann die Genauigkeit der Mittelwertschätzung eventuell gesteigert und der Stichprobenumfang reduziert werden. Voraussetzung ist, dass ein Merkmal gefunden wird (das. sog. Schichtungsmerkmal), welches eine ausreichende Korrelation mit dem Korrekturbetrag aufweist und dessen Verteilung in der Grundgesamtheit bekannt ist. Im empirischen Teil des Gutachtens wurde die versichertenbezogene Zuweisung als ein a priori plausibles Schichtungsmerkmal untersucht. In der Simulation ergaben sich für fünf in Bezug auf die Gradienten der Mittelwerte und Standardabweichungen abgegrenzte Schichten beachtliche Schichtungseffekte, die bei optimaler Verteilung des Gesamtstichprobenumfangs auf die Schichten („Neymann-Allokation“) je nach Beispielskasse zwischen -24% und -58% schwankten, interpretierbar als prozentuale Abzüge vom berechneten Stichprobenumfang einer einfachen Zufallsstichprobe, wenn eine optimal geschichtete Zufallsstichprobe gezogen wird. Dabei hat sich die Neymann-Allokation – vor allem auch zur angemessenen Berücksichtigung der Versicherten mit hohen Kosten – gegenüber einer Schichtung als vorteilhaft erwiesen, welche den Gesamtumfang der Stichprobe proportional zum Anteil der Schichten in der Grundgesamtheit verteilt („proportionale Allokation“).

(11) Beruht die Planung der Stichproben für die Prüfungen nach § 42 RSAV ausschließlich auf empirisch nicht fundierten Annahmen, wie im vorliegenden Gutachten, so läuft man Gefahr, entweder erhebliche Ressourcen zu verschwenden, weil die Stichprobe zu groß, oder man handelt sich enorme Fehlerbreiten ein, weil sie zu klein angelegt wurde. Das Stichprobenverfahren soll den Stichprobenfehler so gering wie möglich halten. Es darf dabei aber die Gesamtprüfkapazität der Prüfdienste nicht überschritten werden. Daher scheint zwingend geboten, die erforderlichen Planungsdaten in einer dem ersten Prüfzyklus vorgeschalteten Pilotphase zu erarbeiten, weil anders die Planungsaufgabe nicht zu lösen ist.

(12) In Hinblick auf die in Ziffer (1) unter a) genannten Prüfungen sollten in der Pilotphase die Ergebnisse der ähnlichen Prüfungen nach § 15a RSAV ausgewertet werden. Für die unter b) genannten Prüfungen werden Vorbereitungsstichproben aus einer geringen Zahl von gut ausgesuchten Beispielskassen vorgeschlagen. In diesen sollte das Prüfverfahren auf der Basis von einfachen Zufallsstichproben erprobt werden, deren Stichprobenumfänge so festzusetzen sind, dass man ein $p_F = 0,05$ (=5%) mit einer relativen Abweichung von höchstens $\varepsilon = 30\%$ schätzen kann.

Das weitere Verfahren sollte dann so ablaufen, wie in § 42 RSAV beschrieben, mit einer (allerdings entscheidenden) Ausnahme: Ein Verwaltungshandeln unterbleibt. Es gibt keine Einforderung des auf die jeweilige Beispielskasse hochgerechneten Korrekturbetrages. Bei Veröffentlichung der Ergebnisse aus der Vorstufe bleiben die Beispielskassen anonym und es werden für sie im ersten Prüfzyklus neue Stichproben gezogen. Das BVA analysiert die im Rahmen der Vorbereitungsstufe von den Prüfdiensten erhobenen Daten zur Gewinnung von Planungsgrundlagen nach einem Konzept, wie in Abschnitt 7.5 des vorliegenden Gutachtens beschrieben.

(13) Für die Prüfungen wird unter Bezugnahme auf § 42 Abs. 5 RSAV ein zweistufiges Verfahren vorgeschlagen, wobei der Korrekturbetrag nur hochzurechnen ist, wenn die auf der ersten Stufe ermittelte Fehlerquote einen vorzugebenden Schwellenwert übersteigt. Die Planungsdaten zur Stichprobenplanung der Stichproben für die ersten Stufe können dem vorangegangenen Prüfzyklus und bei Planung des ersten Prüfzyklus der Pilotphase entnommen werden. Details sind in den Abschnitten 7.3 bis 7.5 dargestellt.

(14) Von dem Korrekturbetrag als Erhebungsziel abgesehen, sind für die Grundgesamtheiten, d.h. für die Versicherten der zu prüfenden Krankenkassen, eine Vielzahl von Merkmalen im Datensatz dokumentiert. Das Bundesversicherungsamt kann die Verteilungen dieser Merkmale ermitteln und Anteils- bzw. Mittelwerte berechnen. Daher ist die Prüfung einer Stichprobe nach § 42 RSAV auf Qualität konzeptionell einfach. Man wählt einige geeignete Anteils- und Mittelwerte aus, schätzt diese designadäquat aus der Stichprobe und vergleicht die Schätzungen mit den entsprechenden Kennziffern der Grundgesamtheit. Wenn die relativen Abweichungen im Absolutbetrag eine festzulegende Schwelle überschreiten, wird die Stichprobe abgelehnt, andernfalls akzeptiert. Bei der Auswahl der für die Prüfung heranzuziehenden Kennziffern ist zu beachten, dass herangezogene Anteilswerte in der vermuteten Größenordnung des Anteils p_F der Versicherten mit Fehlern in den übermittelten Daten liegen und Merkmale, deren Mittelwerte für die Qualitätsprüfung herangezogen werden sollen, eine ähnlich schiefe Verteilung aufweisen, wie diejenige, die für den Korrekturbetrag anzunehmen ist. In dieser Hinsicht eignen sich Anteilswerte von ausgewählten Versichertengruppen oder Hierarchischen Morbiditätsgruppen und die zugehörigen Zuweisungsbeträge besonders gut. In Abschnitt 6 werden hierfür Beispiele diskutiert.

(15) Was die laufende Überprüfung, Fortschreibung und Weiterentwicklung der Stichprobenverfahren nach § 42 RSAV betrifft, wird die Verfolgung dreier unterschiedlicher Ansätze vorgeschlagen. Der erste liegt auf der Hand und dient der Verbesserung der aus der Vorbereitungsstufe gewonnenen Planungsgrundlagen auf der Basis einer wesentlich größeren und von Prüfungszyklus zu Prüfungszyklus wachsenden empirischen Grundlage. Der zweite Ansatz ist Möglichkeiten der Prädiktion des Fehleranteils und des Korrekturbetrages aus Versichertenmerkmalen gewidmet. Von Regressionsverfahren, die den mit dem Morbi-RSA befassten Stellen bei den Krankenkassen, den Prüfdiensten und anderen Akteuren des Gesundheitswesens, die sich für die Verteilung des Geldes aus dem Gesundheitsfonds interessieren, mittlerweile ja bestens bekannt sind, gibt es

auch Versionen für Daten aus komplexen Stichproben, deren Design u. a. Schichtung enthalten kann. Zur Prädiktion des Fehleranteils kommt allerdings die im Umkreis des Morbi-RSA weniger bekannte logistische Regression zur Anwendung. Es sollte bei Verfolgung dieses Ansatzes untersucht werden, ob mit Hilfe solcher Regressionsmodelle der erforderliche Stichprobenumfang für zukünftige Prüfzyklen über mögliche Schichtungseffekte hinaus weiter reduziert werden kann. Der dritte Ansatz schließlich befasst sich mit der Prüfgerechtigkeit und sieht Untersuchungen zur Variation des kassenspezifischen Fehleranteils und des kassenspezifischen Anteils der Summe der Korrekturbeträge an der ursprünglichen Gesamtzuweisung vor, und zwar regional und temporal betrachtet, sowie auch nach Kassenart und Kassengröße.

1 Hintergrund und Aufgabenstellung

1.1 Hintergrund

Nach dem neu in die RSAV eingefügten § 42 sind die die Datenmeldungen für den Morbi-RSA mindestens alle zwei Jahre auf ihre Richtigkeit zu prüfen. Grundlage sind die Datenmeldungen für den korrigierten Jahresausgleich. Die Prüfungen wurden dabei entsprechend der prospektiven Systematik des Morbi-RSA zweigeteilt. Sie umfassen

- a) nach § 42 Abs. 1 S. 1 Nr. 1 RSAV die **Versichertentage** unter Angabe von Geburtsjahr und Geschlecht, sowie die Versichertentage mit Erwerbsminderungsrente, Krankengeld-Anspruch und DMP-Einschreibung (bezogen auf das Ausgleichsjahr) und
- b) nach § 42 Abs. 1 S. 1 Nr. 2 RSAV die **Morbiditätsdaten**, sowie die Versichertentage mit Erwerbsminderungsrente, Auslandsaufenthalt und Kostenerstattung (im Wesentlichen mit Bezug auf das Jahr vor dem Ausgleichsjahr).

Grob gesprochen wird mit der Prüfung unter a) die allgemeine und spezielle Anspruchsberechtigung im Ausgleichsjahr und mit derjenigen unter b) die Korrektheit der Zuweisung zu den Risikogruppen des versichertenbezogenen Klassifikationssystems im Jahr vor dem Ausgleichsjahr geprüft.

In beiden Fällen handelt es sich um Stichprobenprüfungen. Das BVA soll dazu aus der Menge der gemeldeten Pseudonyme jeweils eine Stichprobe ziehen und die Pseudonyme der Stichprobenversicherten den Prüfdiensten einschließlich der zu prüfenden Angaben übermitteln. Die Prüfdienste sollen dann bei den Krankenkassen überprüfen, ob für diese Pseudonyme die vorliegenden Angaben in den Satzarten belegt sind. Zur Reduzierung des notwendigen Prüfumfangs werden die Stichproben für die Prüfungen nach b) auf die Grundgesamtheit der Versicherten beschränkt, die einen prüfungsrelevanten Zuschlag erhalten haben.¹ Nach Abschluss der jeweiligen Prüfung teilen die Prüfdienste dem BVA die Ergebnisse mit.

Das BVA gruppiert die Versicherten dann neu auf Basis der korrigierten Angaben und berechnet für jede Krankenkasse den Betrag, der sich aufgrund der Korrekturen ergibt. Dieser Betrag wird auf den Gesamtbestand der Krankenkasse hochgerechnet und der sich ergebende Korrekturbetrag von der Krankenkasse eingefordert. Dabei werden nur die Änderungen der geprüften Krankenkasse berücksichtigt, nicht aber die Auswirkungen auf Zuweisungen an eine andere Krankenkasse.

¹ Nach Angaben des BVA betrifft das GKV-weit einen Versichertenanteil von ca. 35 %.

1.2 Aufgabenstellung

Die Festlegung der Stichprobenmethodik und des Hochrechnungsverfahrens soll nach § 42 Abs. 6 RSAV für die erstmalige Anwendung des Verfahrens auf Grundlage eines wissenschaftlichen Gutachtens erfolgen, das - getrennt für die Prüfungen nach § 42 Abs. 1 S. 1 Nr. 1 und Nr. 2 RSAV - Vorschläge enthalten soll für

- die Bestimmung der Stichprobenmethodik und des Stichprobenumfangs,
- die Sicherstellung der Stichprobengüte,
- die Bestimmung des Hochrechnungsverfahrens nach § 42 Abs. 5 RSAV, sowie für
- eine Weiterentwicklung dieser Verfahren in nachfolgenden Prüfzyklen.

1.3 Zu prüfende Angaben (Satzarten)

Nach Angaben des BVAs sind die zu prüfenden Angaben in folgender Weise den Satzarten des Datensatzes für den Morbi-RSA zugeordnet (vgl. BVA 2013)

1.3.1 Prüfung nach § 42 Abs. 1 S. 1 Nr. 1 RSAV – Versichertenzeiten

Satzart 110 – Korrektur SA100 EM

- Versichertentage
 - o unter Angabe von Geburtsjahr und Geschlecht
- Versichertentage mit Bezug einer Erwerbsminderungsrente (für K-EMG-Zuordnung)
- Versichertentage mit Krankengeldanspruch
- Versichertentage mit DMP-Einschreibung

1.3.2 Prüfung nach § 42 Abs. 1 S. 1 Nr. 2 RSAV – Morbiditätsdaten

Satzart 110 – Korrektur SA100 KM

- Versichertentage mit Bezug einer Erwerbsminderungsrente (für EMG-Zuordnung)
- Versichertentage mit Auslandsaufenthalt
- Versichertentage mit Kostenerstattung

Satzart 100 Korrekturmeldung

- Angabe zur extrakorporalen Blutreinigung

Satzart 400 Korrekturmeldung

- Arzneimittelkennzeichen und -verordnungen
 - o Verordnungsdatum
 - o Pharmazentralnummer
 - o Anzahl Einheiten

Satzart 500 Korrekturmeldung

- Stationäre Haupt- und Nebendiagnosen
 - o Entlassungsmonat
 - o Diagnose
 - o Art der Diagnose

Satzart 600 Korrekturmeldung

- Ambulante Diagnosen
 - o Leistungsquartal
 - o Diagnose
 - o Qualifizierungsmerkmal
 - o Abrechnungsweg

1.4 Verwendung der zu prüfenden Angaben im Morbi-RSA

Bei den Datenmeldungen im Morbi-RSA ist zwischen zwei Verwendungsarten zu unterscheiden:

Die Zuordnung zu den Risikomerkmale (Gruppierung) erfolgt anhand der Morbiditätsdaten, der Angabe von Geburtsjahr und Geschlecht, sowie der Zahl der Versichertentage mit Erwerbsminderung, Auslandsaufenthalt und Kostenerstattung. Dabei werden grundsätzlich die Informationen aus dem Jahr vor dem Ausgleichsjahr² herangezogen. Für die Zuordnung zu den AGG, AusAGG, K-AGG und K-EMG werden hiervon abweichend das Alter und Geschlecht aus dem Ausgleichsjahr zugrunde gelegt (um Neugeborene zu berücksichtigen). Zudem ist für die Zuordnung zu den K-AGG bzw. K-EMG die Zahl der Erwerbsminderungstage aus dem Ausgleichsjahr relevant, da das Krankengeldverfahren nicht prospektiv ist.

Nachdem die Gruppierung erfolgt ist, werden die Versichertentage, sowie die Versichertentage mit Krankengeldanspruch und mit DMP-Einschreibung aus dem Ausgleichsjahr herangezogen, um zu ermitteln, für wie viele Tage die Zu- und Abschläge für die Risikomerkmale, denen der Versicherte zugeordnet wurde, zugewiesen werden bzw. für wieviele Tage die Zuweisung der Grundpauschale und der DMP-Pauschale erfolgt.

Tabelle 1-1 gibt einen detaillierten Überblick über die Verwendung der Angaben bzw. Satzarten im Morbi-RSA, welche den Prüfungen nach § 42 RSAV unterliegen.

² Anzumerken ist, dass das Verfahren erstmals für das Ausgleichsjahr 2009 zur Anwendung kommen soll.

Tabelle 1-1 Verwendung der zu prüfenden Angaben im Morbi-RSA

	Satzart	SA110 (Korrektur SA100 EM)	SA110 (Korrektur SA100 KM)	SA100 - 600 KM
	Berichtsjahr	Ausgleichsjahr (Bsp.: 2009)	Ausgleichsjahr - 1 (Bsp.: 2008)	Ausgleichsjahr - 1 (Bsp.: 2008)
	Meldsjahr	Berichtsjahr + 2 (Bsp.: 2011)	Berichtsjahr + 3 (Bsp.: 2011)	Berichtsjahr + 2 (Bsp.: 2010)
Angabe	Meldung nach RSAV	§ 30 Abs. 4 S. 2 HS 2 RSAV	§ 30 Abs. 4 S. 2 HS 2 RSAV	§ 30 Abs. 4 S. 2 HS 1 RSAV
Versichertentage	§ 30 Abs. 1 S. 1 Nr. 1	Multiplikation mit Grundpauschale & Zu-/Abschlägen (ohne Krankengeld)	nicht relevant	
VT mit Krankengeldanspruch	§ 30 Abs. 1 S. 1 Nr. 11	Multiplikation mit Zuschlägen für Krankengeld	nicht relevant	
VT mit DMP-Einschreibung	§ 38 Abs. 2	Multiplikation mit DMP-Pauschale pro Tag	nicht relevant	
Geburtsjahr, Geschlecht	§ 30 Abs. 1 S. 1 Nr. 1	Zuordnung: welche AGG bzw. AusAGG Zuordnung: welche K-AGG bzw. K-EMG	Zuordnung: welche EMG Zuordnung: relevant für alters-/geschlechtsabhängige HMG	
VT mit Erwerbsminderung	§ 30 Abs. 1 S. 1 Nr. 2	Zuordnung: wenn >= 183 dann K-EMG, sonst K-AGG (nur relevant, wenn gleichzeitig KG-Anspruch)	Zuordnung: wenn >= 183 dann EMG, sonst keine EMG	
VT mit Auslandsaufenthalt	§ 30 Abs. 1 S. 1 Nr. 8	nicht relevant	Zuordnung: wenn >= 183 dann AusAGG, sonst keine AusAGG	
VT mit Kostenersatzung	§ 30 Abs. 1 S. 1 Nr. 9	nicht relevant	Zuordnung: wenn >= 183 dann KEG, sonst keine KEG	
Extrakorp. Blutreinigung	§ 30 Abs. 1 S. 1 Nr. 3			Zuordnung: HMG
Arzneimittelverordnungen	§ 30 Abs. 1 S. 1 Nr. 4			Zuordnung: HMG
Krankenhausdiagnosen	§ 30 Abs. 1 S. 1 Nr. 5			Zuordnung: HMG
Ambulante Diagnosen	§ 30 Abs. 1 S. 1 Nr. 6			Zuordnung: HMG

Quelle: BVA (2013)

1.5 Konsequenz der Prüfungen nach § 42 RSAV

Wie das BVA in seiner Darlegung der inhaltlichen Anforderungen an das Gutachten ausführt, lassen sich die möglichen Konsequenzen der Prüfungen nach § 42 RSAV auf die Zuweisungen im Morbi-RSA unmittelbar aus Tabelle 1-1 ableiten (BVA 2013).

So führt beispielsweise eine Kürzung der Versichertentage im Ausgleichsjahr zu einer proportionalen Kürzung der Zuweisungen aufgrund der Grundpauschale und der Zu- und Abschläge für einen Versicherten. Eine Kürzung der Versichertentage mit Erwerbsminderung hat dagegen nur dann eine Auswirkung, wenn die Zahl der Tage vor der Prüfung über der 183-Tage-Schwelle lag und anschließend darunter.

In diesem Fall entfällt der EMG-Zuschlag komplett, wenn es sich um die Meldung für das Jahr vor dem Ausgleichsjahr handelt. Handelt es sich um die Meldung für das Ausgleichsjahr, wird aus einem ggf. vorhandenen K-EMG-Zuschlag ein K-AGG-Zuschlag.

Die folgenden Tabellen fassen die Konsequenzen zusammen, die sich aus den Prüfungen nach § 42 RSAV für die Zuweisungen ergeben können.

Tabelle 1-2 Konsequenz der Prüfungen nach § 42 Abs. 1 S. 1 Nr. 1 RSAV

	Satzart	SA110 (Korrektur SA100 EM)
	Berichtsjahr Meldejahr	Ausgleichsjahr (Bsp.: 2009) Berichtsjahr + 2 (Bsp.: 2011)
Angabe	Meldung nach RSAV	§ 30 Abs. 4 S. 2 HS 2 RSAV
Versichertentage	§ 30 Abs. 1 S. 1 Nr. 1	Kürzung der Zuweisungen aufgrund der Grundpauschale und der Zu-/Abschläge (ohne KG)
VT mit KG-Anspruch	§ 30 Abs. 1 S. 1 Nr. 11	Kürzung der Zuweisungen aufgrund der Krankengeld-Zuschläge
VT mit DMP-Einschreibung	§ 38 Abs. 2	Kürzung der Zuweisungen aufgrund der DMP-Pauschale
Geburtsjahr, Geschlecht	§ 30 Abs. 1 S. 1 Nr. 1	Verlust aller Zuschläge, wenn Plausi-Prüfung nicht mehr erfolgreich (Abgleich mit Geburtsjahr & Geschlecht im Vorjahr)
VT mit Erwerbsminderung	§ 30 Abs. 1 S. 1 Nr. 2	wenn < 183, dann K-AGG- statt K-EMG-Zuschlag (nur wenn gleichzeitig KG-Anspruch)

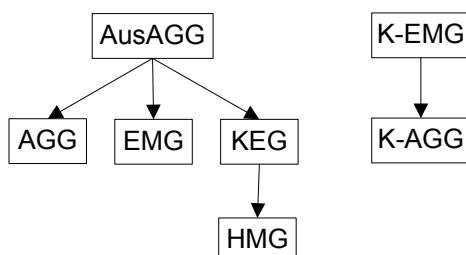
Quelle: BVA (2013)

Tabelle 1-3 Konsequenz der Prüfungen nach § 42 Abs. 1 S. 1 Nr. 2 RSAV

	Satzart	SA110 (Korrektur SA100 KM)	SA100 - 600 Korr.
	Berichtsjahr	Ausgleichsjahr – 1 (Bsp.: 2008)	Ausgleichsjahr – 1 (Bsp.: 2008)
	Meldejahr	Berichtsjahr + 3 (Bsp.: 2011)	Berichtsjahr + 2 (Bsp.: 2010)
Angabe	Meldung nach RSAV	§ 30 Abs. 4 S. 2 HS 2 RSAV	§ 30 Abs. 4 S. 2 HS 1 RSAV
VT mit Erwerbsminderung	§ 30 Abs.1 S.1 Nr.2	< 183, dann Verlust EMG- Zuschlag	
VT mit Auslandsaufenthalt	§ 30 Abs.1 S.1 Nr.8	< 183, dann Verlust Au- sAGG-Zuschlag, stattdessen AGG-Zuschlag und ggf. EMG-, HMG/KEG- Zuschläge	
VT mit Kostenerstattung	§ 30 Abs.1 S.1 Nr.9	< 183, dann Verlust KEG- Zuschlag. ggf. stattdessen HMG-Zuschläge	
Extrakorp. Blutreinigung	§ 30 Abs.1 S.1 Nr.3		ggf. Verlust HMG- Zuschlag, ggf. stattdessen untergeordn. HMG- Zuschlag*
Arzneimittelverordnungen	§ 30 Abs.1 S.1 Nr.4		
Krankenhausdiagnosen	§ 30 Abs.1 S.1 Nr.5		
Ambulante Diagnosen	§ 30 Abs.1 S.1 Nr.6		

Quelle: BVA (2013)

Wie das BVA (2013) ausführt kann der Wegfall der Zuordnung zu einem Risikomerkmal aufgrund der Hierarchisierung der Morbiditätsgruppen, aber auch anderer Risikomerkmale untereinander, dazu führen, dass nun stattdessen die Zuordnung zu einem untergeordneten Risikomerkmal erfolgt, die vorher nicht erfolgte. Dies kann die negative Auswirkung auf die Zuweisungen etwas dämpfen. Die folgende Übersicht zeigt die Hierarchiebeziehungen der Risikomerkmale untereinander.



2 Mögliche Stichprobenverfahren und Struktur des Gutachtens

2.1 Grundbegriffe der Stichprobentheorie

Bei der Planung einer Stichprobe richtet sich das erkenntnisleitende Interesse auf eine Grundgesamtheit, die endlich ist und aus N wohlunterschiedenen, gleichartigen Objekten besteht, die auch als Elemente, Untersuchungseinheiten oder Merkmalsträger bezeichnet werden. Von diesen müssen begrifflich die Auswahlseinheiten unterschieden werden, obwohl sie häufig mit den Untersuchungseinheiten übereinstimmen. So können beispielsweise die Untersuchungseinheiten einer Befragung die Schüler einer Region sein, aber die Auswahlseinheiten aus Schulklassen bestehen (was die Erhebungsorganisation erheblich vereinfacht). Für die vorliegende Planungsaufgaben besteht die Grundgesamtheit immer aus den Versicherten einer bestimmten, am Risikostrukturausgleich beteiligten Krankenkasse in einem vorgegebenen Jahr, wobei eine Person als Versicherte angesehen wird, wenn sie mindestens einen Tag des betrachteten Jahres bei dieser Kasse versichert war.

Es geht in einer statistischen Untersuchung der Grundgesamtheit dann um die Ermittlung einer Kennziffer, die unter Bezug auf ein festes Merkmal berechnet wird. Als Kennziffern kommen Anteilswerte, Mittelwerte oder Merkmalssummen infrage (z. B. der Frauenanteil, das Durchschnittsalter oder die Summe der Zuweisungen in einer Versichertengruppe), aber auch Standardabweichungen, Quantile usw.

Man kann die interessierende Kennziffer unter Umständen aus einer Vollerhebung berechnen. In den meisten Fällen ist es jedoch aus mehreren Gründen zweckmäßig, sich auf eine Teilgesamtheit zu beschränken, die als Stichprobe bezeichnet wird. Die aus der Stichprobe berechnete Kennziffer wird dann als Schätzung der unbekannten Kennziffer der Grundgesamtheit bezeichnet. Mit dem Verfahren der Schätzung ist dann die Hochrechnung (auf die Grundgesamtheit) verbunden, wobei das besondere Interesse dem Schätzfehler (Hochrechnungsfehler) gilt, d. h. der Frage, wie groß die Differenz zwischen der aus der Stichprobe bzw. der Grundgesamtheit berechneten Kennziffer ist.

Da die Verteilung des zu untersuchenden Merkmals zur Optimierung des Stichprobenplanes herangezogen wird, bezieht sich ein konkreter Stichprobenplan stets auf eine einzelne Kennziffer (bzw. auf ein einzelnes Merkmal, das als Erhebungsmerkmal bezeichnet wird). Sollen aus Kostengründen mehrere Kennziffern aus der gleichen Stichprobe geschätzt werden, so handelt es sich um eine Mehrzweckstichprobe. Ein Stichprobenplan für eine Mehrzweckstichprobe stellt immer einen Kompromiss der Pläne für die verschiedenen einzelnen Kennziffern dar.

Für die Planung der Stichproben nach § 42 RSAV ist das primäre Erhebungsmerkmal der versichertenbezogene Korrekturbetrag und die Kennziffer ist die Merkmalssumme. Daneben sind aber auch verschiedene andere Merkmale zur Charakterisierung der auftretenden Fehler von Interesse.

2.2 Taxonomie der Stichprobenverfahren

Nach Schäfer (2004) umfasst jedes Stichprobenverfahren die Auswahl der Elemente und die Schätzung der Populationswerte, d. h. das Hochrechnungskonzept. Daher kann eine Taxonomie der Stichprobenverfahren sich sowohl an Kriterien orientieren, welche die Auswahlprozedur betreffen, als auch an solchen, welche das Schätzverfahren zum Gegenstand haben. Üblich ist es, die Einteilung vorrangig mit Bezug auf das Auswahlverfahren vorzunehmen.

Von grundlegender Bedeutung ist dabei die Frage, ob die Wahrscheinlichkeiten bekannt sind, mit dem jedes Element der Grundgesamtheit für die Stichprobe ausgewählt wird, oder nicht. Kann diese Frage bejaht werden und sind diese Wahrscheinlichkeiten alle von Null verschieden, so sprechen wir von einer Wahrscheinlichkeits- bzw. Zufallsstichprobe. (Schäfer 2004, vgl. Abb. 2-1).

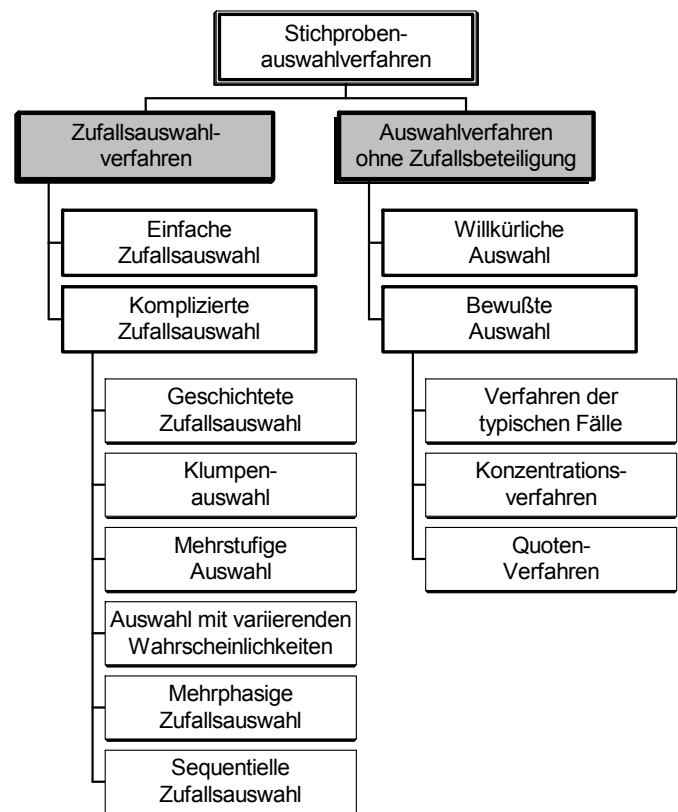


Abb. 2-1 Möglichkeiten der Stichprobenauswahl

Quelle: Schäfer (2004)

2.3 Die einfache Zufallsauswahl

Die einfache Zufallsauswahl, bei der jede mögliche Auswahl vom Umfang n die gleiche konstante Wahrscheinlichkeit trägt, ist unkompliziert hochzurechnen und benötigt nur wenige Informationen über die Grundgesamtheit. Allerdings besteht für die im Kontext zu planenden Stichproben die Gefahr, dass gering besetzte, aber für das Korrekturverfahren relevante Versichertengruppen (z. B. sehr teure Versicherte) bei realistischen Auswahlätzen in der Stichprobe nicht hinreichend vertreten sind. Außerdem weist die einfache Zufallsstichprobe, wenn die Grundgesamtheit bezüglich des zu erhebenden Merkmals (hier: der versichertenbezogene Korrekturbetrag) heterogen zusammengesetzt ist, einen höheren Schätzfehler auf als eine geschichtete Zufallsauswahl, die im nächsten Abschnitt näher beschrieben wird.

Die Antwort auf Frage, ob die geschichtete der einfachen Zufallsstichprobe vorzuziehen ist, hängt wesentlich davon ab, ob Schichtungsmerkmale gefunden werden, von denen plausibel ist, dass sie eine hinreichend hohe Korrelation zum Korrekturbetrag aufweisen.

2.4 Die geschichtete Zufallsauswahl³

Wenn die Grundgesamtheit in Bezug auf das Erhebungsmerkmal vergleichsweise inhomogen zusammengesetzt ist, lässt sich die Genauigkeit der Schätzungen unter Umständen dadurch erhöhen, dass die Grundgesamtheit unter Verwendung eines oder mehrerer Schichtungsmerkmale in Schichten (Strata) zerlegt wird, die in sich möglichst homogen und untereinander möglichst heterogen bezüglich des Erhebungsmerkmals sind. Wird dann aus jeder Schicht eine einfache Zufallsauswahl getroffen, so entsteht eine geschichtete (stratifizierte) Zufallsstichprobe. Der Gewinn an Genauigkeit, der sog. Schichtungseffekt hängt dabei davon ab, wie gut die Homogenitätsstruktur der Grundgesamtheit hinsichtlich des Erhebungsmerkmals (mit unbekannter Verteilung) durch das Schichtungsmerkmal (mit bekannter Verteilung) abgebildet werden kann. Neben dem Streben, die Varianz der Schätzung bei der Hochrechnung möglichst klein zu halten, gibt es noch andere Gründe für eine Schichtung. Diese können mit der Erhebungsorganisation zusammenhängen (z. B. die Schichtung nach Ländern in der amtlichen Statistik) oder mit dem Wunsch, die Genauigkeit in den durch die Schichten gebildeten Teilgesamtheiten separat kontrollieren zu können (dann wird gleichsam jede Schicht als eigene Grundgesamtheit aufgefasst).

Grundsätzlich sind bei Anwendung einer geschichteten Stichprobe vier Verfahrensschritte zu durchlaufen:

1. Zunächst ist die Grundgesamtheit vollständig und überlappungsfrei in Teilpopulationen aufzuteilen, welche die Schichten bilden
2. Aus jeder Schicht ist unabhängig voneinander eine einfache Zufallsauswahl zu treffen.
3. Aus den stratumsspezifischen Stichproben wird pro Schicht ein Mittelwert (oder ggf. eine andere Kennziffer) separat geschätzt. Die separaten Schätzungen werden im Rahmen der Hochrechnung geeignet gewichtet zu einer Schätzung des Populationsmittelwertes kombiniert.
4. Analog müssen auch die Varianzen zunächst separat für jede Schicht ermittelt werden. Aus diesen wird dann mit einem weiteren Satz von Gewichten eine Schätzung der Varianz der kombinierten Mittelwertschätzung wiederum als gewogener Mittelwert gewonnen.

Fest zu halten bleibt, dass die Planung einer geschichteten Stichprobe nur auf der Grundlage von Vorwissen über die Grundgesamtheit erfolgen kann.

2.5 Sequentielle Zufallsauswahlverfahren

Wenn der Stichprobenumfang n so klein wie möglich gehalten werden soll (wie z. B. in einer die Prüfstücke zerstörenden Qualitätsprüfung oder wegen begrenzter Prüfkapazitäten) kann auch eine sequentielle Vorgehensweise ins Kalkül gezogen werden, bei der

³ Die Darstellung der geschichteten Zufallsstichprobe ist weitgehend Schäfer (2004) entnommen.

nach jeder Aufnahme eines Elements in die Stichprobe nach vorgegebenen Kriterien geprüft wird, ob weiter gezogen werden soll oder die Stichprobe bereits komplett ist.

Sequentialverfahren folgen eigenen Gesetzmäßigkeiten und werden in der allgemeinen Stichprobentheorie meist nicht mitbehandelt. Der Unterschied zwischen einer einfachen Zufallsstichprobe und der sequentiellen Zufallsauswahl lässt sich im Kontext der Prüfung nach § 42 RSAV gut veranschaulichen, wenn man sich auf die Fragestellung bezieht, den Anteil der Versicherten mit Fehlern in den an das BVA übermittelten Daten (im Folgenden auch kurz als „Fehlerquote“ bezeichnet) in einer großen zu prüfenden Kasse zu schätzen (in der Theorie wird eine unendliche Grundgesamtheit vorausgesetzt). Im Rahmen einer einfachen Zufallsstichprobe werden per Zufall nacheinander n Pseudonyme (ohne Zurücklegen) gezogen. Findet der Prüfdienst dann bei $k \leq n$ der Versicherten Fehler in den übermittelten Daten, so wird der unbekannte Anteilswert aus der Stichprobe in naheliegender Weise wie folgt geschätzt:

$$(1) \quad \hat{p} = \frac{k}{n}$$

Bei der sequentiellen Vorgehensweise wird nicht der Stichprobenumfang n vorgegeben, sondern die Zahl k der Pseudonyme mit Fehlern in den Daten und es werden solange per Zufall Pseudonyme (ohne Zurücklegen) gezogen und die zugehörigen Datensätze überprüft, bis die Zahl der Versicherten mit Fehlern in den übermittelten Daten gerade k betrifft. Der resultierende Stichprobenumfang n ist jetzt das Ergebnis des Verfahrens (weshalb es in diesem Fall auch als „inverse sampling“ bezeichnet wird).

Der Anteilswert wird dann aus einer Formel (1) ganz ähnlichen Gleichung geschätzt:⁴

$$(2) \quad \hat{p} = \frac{k-1}{n-1}$$

Bei der einfachen Stichprobe wird der feste Stichprobenumfang n und im Sequentialverfahren die Fehlerzahl k planerisch so vorgegeben, dass der (jeweils verfahrensspezifisch operationalisierte) Hochrechnungsfehler unterhalb einer vorgegebenen Grenze bleibt.

Wenn beide Stichprobenverfahren anwendbar sind, führt das Sequentialverfahren vielfach zu einem reduzierten Stichprobenumfang und erweist sich als das ökonomisch günstigere Verfahren. Dennoch ist die sequentielle Vorgehensweise weniger verbreitet als Stichprobenverfahren mit festem Stichprobenumfang. Gründe hierfür sind einerseits darin zu suchen, dass sich nicht alle Fragestellungen der Stichprobentheorie für ein Sequentialverfahren eignen und andererseits darin, dass letztere konzeptionell, mathematisch und von der Umsetzung her einfacher sind, als erstere (vgl. Gosh und Sen 1991).

⁴ Die „-1“ in Formel (2) ist der Tatsache geschuldet, dass der zuletzt untersuchte Datensatz verfahrensimmanent immer ein fehlerhafter ist.

2.6 Für Stichproben nach § 42 RSAV ungeeignete Auswahlverfahren

Verfahren, bei denen die Auswahl der Stichprobenelemente nicht auf der Basis eines realen oder simulierten Zufallsexperimentes erfolgt, scheiden aus, weil der Schätzfehler bei der Hochrechnung nicht kontrolliert werden kann.

Klumpenauswahlverfahren gehen in der Regel mit einem Verlust an Genauigkeit einher, den man in Kauf nimmt, um Ziehungskosten zu sparen. Sie kommen für die hier zu besprechenden Stichproben nicht infrage, weil die Gruppierung der Pseudonyme zu Klumpen (Clustern) zum Zwecke der Ziehung keine Kostenvorteile bietet.

Wenn Zufallsauswahlen in einem Stichprobendesign auf verschiedenen Stufen vorgesehen sind, so handelt es sich um ein mehrstufiges Auswahlverfahren.⁵ Diese enthalten in der Praxis häufig auf mindesten einer Stufe Klumpen als Auswahlseinheiten und werden eingesetzt, um die Ziehungsorganisation zu erleichtern und dadurch Ziehungskosten einzusparen. Als vorteilhaft erweist sich ein mehrstufiger Auswahlplan vielfach, wenn die Grundgesamtheit hierarchisch gegliedert ist. Aber auch in Fällen, in denen keine geeignete Auswahlgrundlage für eine einfache Zufallsauswahl verfügbar ist, kann sich ein mehrstufiger Auswahlplan als zweckmäßig erweisen. Da die Grundgesamtheit sich im Kontext jedoch aus den Versicherten einer bestimmten Krankenkasse zusammensetzt, deren Stammdaten dem BVA pseudonymisiert (maschinenlesbar) zum Zwecke der Stichprobenziehung zur Verfügung stehen, bietet ein mehrstufiges gegenüber einem einstufigen Auswahlverfahren keine Vorteile.

Eine Auswahl mit variierenden Wahrscheinlichkeiten ist ein spezielles (ggf. mehrstufiges) Klumpenauswahlverfahren, das zweckmäßig ist, wenn die Klumpengröße sehr stark variiert. Eine solche Auswahl kommt hier nicht in Betracht.

Mehrphasige Auswahlverfahren können sich als zweckmäßig erweisen im Rahmen einer sog. gebundenen Hochrechnung, wenn das zu erhebende Merkmal korreliert ist mit einem anderen (Hilfsmerkmal), dessen Ausprägungen für die Untersuchungseinheiten man zwar nicht kennt, aber dessen Erhebung mit wesentlich weniger Aufwand verbunden ist. Gebundene Hochrechnungen können für die Weiterentwicklung der Verfahren eine Rolle spielen (vgl. Abschnitt 7.6), aber mehrphasige Auswahlen sind im Kontext nicht erforderlich, weil die Ausprägungen sämtlicher potentieller Hilfsmerkmale in der Grundgesamtheit der zu prüfenden Kasse ja bekannt sind.

⁵ Es gibt Mehrstufigkeit auch bei sequentiellen Stichprobenverfahren. Diese sind hier aber nicht gemeint.

2.7 Struktur des Gutachtens

Die Planung einer Stichprobe setzt ein gewisses Maß an Informationen über die Grundgesamtheit voraus. Zwar sind die Anforderungen bezüglich der Vorkenntnisse zwischen den potentiell möglichen Auswahlverfahren unterschiedlich (am größten sind sie für die geschichtete Zufallsauswahl), ganz ohne Vorkenntnisse kommt man bei Planung einer Stichprobe aus einer endlichen Grundgesamtheit jedoch nicht aus.

Die Aufgabenstellung nach § 42 Abs. 6 RSAV richtet sich aber nicht auf eine bestimmte zu prüfende Kasse, sondern schließt sämtliche am RSA beteiligten Krankenkassen ein. Die Aufgaben, die von dem Gutachter nach § 42 Abs. 6 RSAV zu lösen sind, können vor diesem Hintergrund nur exemplarisch für eine Reihe – nach Morbidität, Größe, Region und Kassenart gut ausgesuchter Beispielskassen – behandelt werden. Das BVA hat hierfür fünf Krankenkassen ausgewählt, die nach Auffassung des Amtes in ihrer Gesamtheit das Spektrum der 147 im Korrekturverfahren für das Ausgleichsjahr 2009 am Morbi-RSA beteiligten Krankenkassen der GKV gut abbilden. Aus dem Versichertenbestand jeder dieser Krankenkassen wurden vom BVA zum Zwecke der Anonymisierung zwei Stichproben gezogen, die ca. 80% der jeweiligen Grundgesamtheit umfassen und in denen ausgewählte Versichertenmerkmale dokumentiert sind.⁶ Die jeweils ersten der beiden Stichproben stellen die Grundgesamtheiten für die nach § 42 Abs. 1 S. 1 Nr. 1 RSAV⁷ und die jeweils zweiten die Grundgesamtheiten für die nach § 42 Abs. 1 S. 1 Nr. 2 RSAV⁸ zu ziehenden Stichproben dar (für sie wurde der Versichertenbestand der fünf Ausgangskassen auf die Versicherten mit einem prüfungsrelevanten Zuschlag beschränkt). Darüber hinaus wurde vom BVA mit gleichem Merkmalsatz je ein Datensatz für die Versicherten der GKV insgesamt (genauer: der o. g. 147 Krankenkassen) zur Verfügung gestellt. Die resultierenden zwölf Datenbestände wurden vom BVA nach der Zuweisungssumme je Versichertem aufsteigend sortiert und dann in einem zweiten Schritt für die Nutzung durch den Gutachter in zwanzig Schichten aggregiert, wobei die Schichtgrenzen so gewählt wurden, dass die Zuweisungssumme über alle Versicherten der Schicht in allen Schichten gleich groß ist (s. Abschnitte 8.1 und 8.2 im Anhang).

Diese Daten bilden die empirische Grundlage für das vorliegende Gutachten, welches seiner Struktur nach nur ein Rahmengutachten sein kann, das die Vorgehensweisen für die Stichproben nach Nr. 1 und Nr. 2 exemplarisch anhand der zehn (gestützten) Beispielskassen erarbeitet und beschreibt und daraus geordnete Verfahrensschritte und Empfehlungen ableitet.

⁶ Der zeitliche Bezug der Versichertenmerkmale wurde dabei so gewählt, dass mit den Daten die Prüfung in Bezug auf das Ausgleichsjahr 2009 simuliert werden kann.

⁷ Im Weiteren abgekürzt als „Stichprobe nach Nr.1“ bezeichnet.

⁸ Im Weiteren abgekürzt als „Stichprobe nach Nr.2“ bezeichnet.

3 Planung und Hochrechnung einer einfachen Zufallsstichprobe

3.1 Bezeichnungen

Um die Grundgesamtheit von der Stichprobe zu unterscheiden, verwendet man üblicherweise große Buchstaben mit Bezug auf die Grundgesamtheit und kleine mit Bezug auf die Stichprobe. Die für die Fragestellung wichtigen Kennziffern sind in der folgenden Tabelle zusammengestellt.

Tabelle 3-1 Wichtige Kennziffern der Grundgesamtheit und analoge Stichprobengrößen

Kenziffer	In der Grundgesamtheit (Umfang: N)	Stichprobe (Umfang: n)
Merkmalssumme	$\sum_{i=1}^N Y_i$	$\sum_{i=1}^n y_i$
Mittelwert	$\bar{Y} = \frac{1}{N} \sum_{i=1}^N Y_i$	$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$
Anteilswert	$P = \frac{A}{N}$	$p = \frac{a}{n}$
Verhältniszahl	$V = \frac{\sum Y_i}{\sum X_i} = \frac{\bar{Y}}{\bar{X}}$	$v = \frac{\sum y_i}{\sum x_i} = \frac{\bar{y}}{\bar{x}}$
Varianz	$\sigma_Y^2 = \sigma_{YY} = \frac{1}{N} \sum_{i=1}^N (Y_i - \bar{Y})^2$	$\hat{\sigma}_y^2 = \hat{\sigma}_{yy} = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2$
Korrigierte Varianz	$S_Y^2 = S_{YY} = \frac{1}{N-1} \sum_{i=1}^N (Y_i - \bar{Y})^2$	$s_y^2 = s_{yy} = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2$
Korrigierte Kovarianz	$S_{XY} = \frac{1}{N-1} \sum_{i=1}^N (X_i - \bar{X})(Y_i - \bar{Y})$	$s_{xy} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$

Quelle: Schäfer (2004)

Das Verhältnis vom Stichprobenumfang n und Umfang der Grundgesamtheit N, d. h.

$$(3) \quad f = \frac{n}{N}$$

wird als Auswahlatz und die Größe 1-f als Endlichkeitskorrektur bezeichnet.

3.2 Mittlerer Korrekturbetrag und Summe der Korrekturbeträge

Das Merkmal Y der Versicherten in der Grundgesamtheit, das den Korrekturbetrag beschreibt, ist nur dann größer als Null, wenn bei dem Merkmalsträger, d. h. dem zugehörigen Versicherten Fehler in den an das BVA übermittelten Daten vorliegen. Auf diesen Sachverhalt gehen wir weiter unter noch näher ein. Das Stichprobenmittel der versichertenbezogenen Korrekturbeträge

$$(4) \quad \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$$

stellt eine konsistente⁹ und erwartungstreue¹⁰ Schätzung des mittleren Korrekturbetrages dar, deren Varianz (der sog. *Standardfehler*) sich folgendermaßen berechnet

$$(5) \quad \text{Var}(\bar{y}) = \left(1 - \frac{n}{N}\right) \frac{S_Y^2}{n} = (1 - f) \frac{S_Y^2}{n}$$

(vgl. Cochran 1972).

Da die korrigierte Varianz S_Y^2 nicht bekannt ist, liegt es nahe, hierfür die korrigierte Stichprobenvarianz einzusetzen, um eine Schätzung der Varianz zu erhalten:

$$(6) \quad \hat{\text{Var}}(\bar{y}) = (1 - f) \frac{s_y^2}{n}$$

Auch diese Schätzung ist konsistent und erwartungstreu (vgl. Cochran 1972).

Nun ist aber nicht der Mittelwert, sondern die Summe der Korrekturbeträge die im Kontext entscheidende Zielgröße der Stichprobenplanung. Das bedeutet aber keine zusätzliche Schwierigkeit, denn es gilt offenbar

$$(7) \quad \sum_{i=1}^N Y_i = N \cdot \bar{Y}$$

Als erwartungstreue Schätzung der Summe der Korrekturbeträge erhalten wir daher

$$(8) \quad N \cdot \bar{y} = \frac{N}{n} \sum_{i=1}^n y_i$$

Die Schätzung ist auch konsistent. Der Faktor N/n (Kehrwert des Auswahlssatzes) in (8) wird als *Hochrechnungsfaktor* bezeichnet. Die Varianz der Stichprobenfunktion $N\bar{y}$ ergibt sich unmittelbar aus (6) in der Form

$$(9) \quad \text{Var}\left(\frac{N}{n} \sum_{i=1}^n y_i\right) = N^2 (1 - f) \frac{S_Y^2}{n}$$

und diese wird konsistent und erwartungstreu geschätzt, wenn man die korrigierte Stichprobenvarianz anstelle der korrigierten Populationsvarianz einsetzt:

$$(10) \quad \hat{\text{Var}}\left(\frac{N}{n} \sum_{i=1}^n y_i\right) = N^2 (1 - f) \frac{s_Y^2}{n}$$

3.3 Zur Verteilung des Korrekturbetrages

Es gibt kaum Vorwissen über die Verteilung des Korrekturbetrages. Wenn man davon

⁹ Ein Schätzwert einer Kennziffer heißt *konsistent*, wenn er für $n=N$ mit der Kennziffer übereinstimmt.

¹⁰ Die Schätzung $\hat{\theta}$ einer Kennziffer θ heißt *erwartungstreu*, wenn $E(\hat{\theta}) = \theta$ gilt. Ist die Schätzung nicht erwartungstreu, so wird $E(\hat{\theta}) - \theta$ als *Stichprobenverzerrung (sampling bias)* bezeichnet.

ausgeht, dass der Anteil der Versicherten mit Fehlern in den übermittelten Daten¹¹ klein ist und in einer Größenordnung von unter 10% angenommen werden kann, so lässt sich aber schlussfolgern, dass der Korrekturbetrag als Versichertenmerkmal eine extrem schiefe Verteilung aufweist, da dann mindestens 90% der Merkmalsausprägungen Null sind. Damit ist der Median der Verteilung a priori ebenfalls Null und der Mittelwert weist eine aufgeblähte Varianz auf, die jedenfalls wesentlich größer ist, als man z. B. bei Messwerten, die einer Normalverteilung genügen, erfahrungsgemäß erwarten kann.

Diese Überlegungen weisen auf ein Grundproblem des Versuchs hin, den mittleren Korrekturbetrag (oder, was aufs Gleiche hinauskommt, die Summe der Korrekturbeträge) auf Stichprobenbasis zu ermitteln. Der Stichprobenumfang, der benötigt wird, um den mittleren Korrekturbetrag mit vorgegebener Präzision zu schätzen, ist, wie im nächsten Abschnitt gezeigt wird, näherungsweise proportional zum quadrierten Variationskoeffizienten¹². Da die aufgeblähte Varianz mit einem aufgeblähten Variationskoeffizienten einhergeht, werden überdurchschnittlich große Stichprobenumfänge benötigt, um eine ausreichende Genauigkeit zu erreichen.

Wie die folgende Überlegung zeigt, kommt es dabei vor allem darauf an, den Anteil der Versicherten mit Fehlern in den übermittelten Daten genau zu schätzen.

Ist K deren Anzahl (in der Grundgesamtheit) und seien die Versicherten insgesamt so nummeriert, dass diejenigen mit Datenfehlern die ersten K Plätze einnehmen, dann ist

$$(11) \quad \bar{Y}_F = \frac{1}{K} \sum_{i=1}^K Y_i$$

ein mittlerer Korrekturbetrag, der ohne Einbeziehung der vielen Versicherten mit einem Null-Betrag berechnet wurde, also frei von den oben geschilderten Defiziten ist.

Der Anteil der Versicherten mit Fehlern in den Daten berechnet sich in der Grundgesamtheit wie folgt:

$$(12) \quad p_F = \frac{K}{N}$$

Es besteht nun offenbar folgende Beziehung zwischen dem mittleren Korrekturbetrag und den Größen aus (11) und (12):

$$(13) \quad \bar{Y} = p_F \cdot \bar{Y}_F,$$

so dass Fehler bei der Schätzung von p_F unmittelbar zu Fehlern in der Schätzung von $\bar{Y}$ führen.

¹¹ Diese Versichertengruppe sei im Weiteren unter dem Kürzel VG_F angesprochen.

¹² Der Variationskoeffizient V ist definiert als Standardabweichung/Mittelwert.

Eine einfache Überlegung zeigt, dass man den Variationskoeffizienten V als Funktion der Fehlerquote p_F nach unten abschätzen, und in einem zweiten Schritt unter Verwendung eines weiteren Parameters sogar berechnen kann.

Wir nehmen zunächst an, der Korrekturbetrag y sei für die Versicherten aus VG_F konstant $= B$.¹³ Dann hat ein Anteil von $1 - p_F$ Versicherten die Merkmalsausprägung $y=0$ und der restliche Anteil (p_F) die Merkmalsausprägung $y=B$.

Bei einer solch einfach strukturierten Verteilung sind Mittelwert und Standardabweichung leicht zu berechnen. Es gilt dann:

$$(14) \quad \bar{y} = p_F \cdot B$$

(vgl. auch Formel (13)) und

$$(15) \quad s_y = \sqrt{B^2 \cdot p_F - B^2 \cdot p_F^2} = B \sqrt{p_F(1-p_F)}$$

Aus (14) und (15) ergibt sich nun:

$$(16) \quad V = \frac{s_y}{\bar{y}} = \frac{B \sqrt{p_F(1-p_F)}}{p_F B} = \sqrt{\frac{1-p_F}{p_F}}$$

Natürlich ist der Korrekturbetrag in der Versichertengruppe VG_F hinweg in Wirklichkeit nicht konstant, sondern variabel. Aber durch die zusätzliche Variabilität kann V nicht kleiner, sondern nur größer werden. Mit Formel (16) haben wir also in Abhängigkeit von der Fehlerquote p_F eine untere Grenze für den tatsächlichen Variationskoeffizienten gefunden.

Wenn wir die Voraussetzung eines konstanten Korrekturbetrages für die Versicherten aus VG_F fallen lassen und den Mittelwert in dieser Gruppe nach wie vor mit B bezeichnen, so lässt sich die Abschätzung nach unten zu einer Gleichung verschärfen, wenn wir einen weiteren Parameter q einführen, der den Variationskoeffizienten des Korrekturbetrages misst, wenn wir diesen nur für Versicherte aus VG_F berechnen, so dass also für die Standardabweichung S_F in VG_F gilt:

$$(17) \quad S_F = q \cdot B$$

Wir erhalten dann, mit Hilfe der weiter unten erläuterten Streuungszerlegung (44), für die Varianz des Korrekturbetrages (berechnet über alle Versicherten der untersuchten Krankenkasse):

$$(18) \quad s_y^2 = (p_F q^2 + p_F(1-p_F)) \cdot B^2$$

Und daraus weiter:

¹³ Das entspricht der Zuweisung als Versichertenmerkmal bezogen auf eine bestimmte HMG, wobei in diesem Fall B gerade der Regressionskoeffizient der zugehörigen Indikatorvariable ist.

$$(19) \quad V = \frac{s_y}{\bar{y}} = \frac{B\sqrt{p_F q^2 + p_F(1-p_F)}}{p_F B} = \frac{\sqrt{p_F q^2 + p_F(1-p_F)}}{p_F} = \sqrt{\frac{q^2 + 1 - p_F}{p_F}}$$

Auch in (19) tritt B, jetzt der mittlere Korrekturbetrag in VG_F , nicht mehr auf. Allerdings finden wir in (19) zum Unterschied von (16) höchst plausibel im Zähler einen weiteren Summenterm, welcher der Variabilität des Korrekturbetrages innerhalb der Versichertengruppe VG_F Rechnung trägt. In der folgenden Tabelle sind die Werte des Variationskoeffizienten V für verschiedene Fehlerquoten p_F und verschiedene Werte von q zusammengestellt, wobei die Zeile für das in der Realität kaum mögliche $q = 0$ nur eingefügt wurde, um auch über ein paar tabellierte Werte der Formel (16) zu verfügen.

Tabelle 3-2 Abhängigkeit des Variationskoeffizienten V des Korrekturbetrages von der Fehlerquote und dem Variationskoeffizienten q innerhalb von VG_F

q / p_F	0,05	0,04	0,03	0,02	0,01	0,005
0,0	4,4	4,9	5,7	7,0	9,9	14,1
0,2	4,4	5,0	5,8	7,1	10,1	14,4
0,4	4,7	5,3	6,1	7,5	10,7	15,2
0,6	5,1	5,7	6,7	8,2	11,6	16,5
0,8	5,6	6,3	7,3	9,0	12,8	18,1
1,0	6,2	7,0	8,1	9,9	14,1	20,0

3.4 Benötigter Stichprobenumfang zur Erreichung einer vorgegebenen Präzision der Schätzung des mittleren Korrekturbetrages

Das asymptotische $(1-\alpha)$ -Konfidenzintervall für den mittleren Korrekturbetrag hat die Form

$$(20) \quad KI_{1-\alpha}(\bar{Y}) = [\bar{y} - u_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{N-n}{Nn}} S, \bar{y} + u_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{N-n}{Nn}} S]$$

mit der unbekannten Standardabweichung und dem $(1-\frac{\alpha}{2})$ -Quantil $u = u_{1-\frac{\alpha}{2}}$ der Standardnormalverteilung, das für $\alpha=5\%$ $u=1,96$ beträgt.

Gibt man jetzt eine relative Abweichung ε vor, um welche die Schätzung mit 95% Wahrscheinlichkeit maximal nur vom Mittelwert $\bar{Y}$ in der Grundgesamtheit abweichen darf, so führt (20) zu

$$(21) \quad u \cdot \sqrt{\frac{N-n}{Nn}} S = \varepsilon \bar{Y}$$

Aufgelöst nach n und unter Verwendung des Variationskoeffizienten erhalten wir

$$(22) \quad n = \frac{\frac{u^2}{\varepsilon^2} V^2}{1 + \frac{1}{N} \frac{u^2}{\varepsilon^2} V^2} = \frac{V^2}{\frac{1}{N} V^2 + \frac{\varepsilon^2}{u^2}}$$

(vgl. Sukhatme und Sukhatme, 1970, Abschnitt 1.12).

Für kleine Größen $\delta > 0$ gilt die Beziehung („Rechnen mit kleinen Größen“)

$$(23) \quad (1 + \delta)^{-1} \approx 1 - \delta,$$

die unmittelbar aus einer Potenzreihenentwicklung folgt.

Damit lässt sich die erste der beiden Gleichungen aus (22) in folgender Weise umformen

$$(24) \quad n = \frac{u^2}{\varepsilon^2} V^2 \cdot \left(1 + \frac{1}{N} \frac{u^2}{\varepsilon^2} V^2\right)^{-1} \approx \frac{u^2}{\varepsilon^2} V^2 \cdot \left(1 - \frac{1}{N} \frac{u^2}{\varepsilon^2} V^2\right) \approx \frac{u^2}{\varepsilon^2} V^2,$$

da der zweite Term in der Klammer infolge des Faktors $1/N$ bei üblichen Setzungen von ε klein gegen 1 ist. Der für eine gegebene Präzision erforderliche Stichprobenumfang ist also proportional dem Quadrat des Variationskoeffizienten des Korrekturbetrages.

Setzt man nun noch (19) in die Formel (24) ein und berücksichtigt, dass für kleine p_F die Größe $1 - p_F$ nahe bei Eins liegt, so erhält man näherungsweise den zur Beschränkung des maximalen relativen Fehler unter ε benötigten Stichprobenumfang n (ohne Berücksichtigung der Endlichkeitskorrektur, d. h. für nicht zu kleine Kassen):

$$(25) \quad n \approx \frac{u^2 (q^2 + 1)}{\varepsilon^2 p_F}$$

Für $\varepsilon = 0,02$ (20%), $q = 0,6$ und $p_F = 0,02$, ergibt sich $n \approx 6.500$ ($p_F = 0,01$: $n \approx 13.000$).

Die exakte Formel erhält man durch Einsetzen von (19) in (22). Nach einigen einfachen Umformungen erhält man:

$$(26) \quad n = \left(\frac{1}{N} + \left(\frac{\varepsilon}{u} \right)^2 \frac{p_F}{q^2 + 1 - p_F} \right)^{-1}$$

Die Formeln (22) bis (26) gelten auch für die Summe des Korrekturbedarfs, weil Mittelwert und Standardabweichung mit dem gleichen Faktor multipliziert werden.

Da über die Verteilung des Korrekturbedarfs nichts bekannt ist, kann eine Abschätzung des Variationskoeffizienten nur unter gewissen Annahmen über V bzw. q erfolgen.

In einer Art Best-Case-Szenario wurde zunächst angenommen, der Korrekturbedarf sei direkt proportional der Zuweisungssumme. Dann kann der Variationskoeffizient der Zuweisung für die Berechnung des Stichprobenumfangs herangezogen werden. Dieser liegt in den Beispielskassen nahe bei 1. Genauer beträgt er 1,25 in den Kassen A bis C, 1,29 in der Kasse D und 0,97 in der Kasse E der Nr. 1-Stichprobe. In den Beispielskassen der Nr. 2-Stichprobe ist er naturgemäß kleiner und beträgt 0,93 in den Kassen F und G, 1,01 in der Kasse H, 0,95 in der Kasse I und 0,86 in der Kasse J.

Um realistischere Schätzungen der Variationskoeffizienten zu erhalten, wurde in einem zweiten Schritt die Verteilung des Korrekturbedarfs in jeder der zehn Beispielskassen in

der im Folgenden beschriebenen Weise simuliert¹⁴:

1. Für jede der 20 Schichten wurde zunächst per Zufallszahlengenerator eine Zahl zwischen 3 und 7 erzeugt, die dem Prozentsatz der Versicherten, bei denen Datenfehler vorliegen, entsprechen soll.
2. Mit den so bestimmten Prozentsätzen wurden Versicherte per Zufall mit einer Indikatorvariable markiert. War der Prozentsatz beispielsweise 4, so finden wir nach diesem Schritt in der betreffenden Schicht 4% markierte, die zufällig über die Schicht verteilt sind.
3. Für jeden markierten Versicherten wurde per Zufall eine Zahl erzeugt, die in den fünf Kassen für die Nr. 1-Prüfung zwischen 5 und 100 und in den fünf Kassen für die Nr. 2-Prüfung zwischen 5 und 50 liegt. Diese Zahl soll dem Prozentsatz der versichertenbezogenen Zuweisung entsprechen, der als Korrekturbetrag gewertet wird. Hat ein markierter Versicherter z. B. den Prozentsatz 40 per Zufall zugeordnet bekommen, so wurden 40% der auf ihn entfallenden Zuweisung als Korrekturbetrag dokumentiert. Damit wurde im Datensatz der Beispielskassen jeweils eine neue Variable „Korrekturbedarf“ erzeugt, die für alle nicht markierten Versicherten auf Null gesetzt wurde.

Im Ergebnis dieser Simulation („Simulation I“) finden wir sehr große Variationskoeffizienten, die in Tabelle 3-3 zusammengestellt sind. Gegenüber dem Best Case-Szenario sind die Variationskoeffizienten (je nach Beispielskasse) um einen Faktor 6,6 bis 7,9 gewachsen. Weil diese Steigerung unerwartet groß ausgefallen ist, wurde der Punkt 3 insoweit abgeändert, als der Prozentsatz für die Nr. 1-Prüfung auf das Intervall zwischen 5 und 20 und für die die Nr. 2-Prüfung auf das Intervall zwischen 5 und 10 reduziert wurde.

Mit diesen veränderten Parametern wurde die Simulation mehrfach wiederholt („Simulationen II bis IV“). Über die mittleren Variationskoeffizienten informiert Tabelle 3-4. Es fällt auf, dass der Variationskoeffizient in allen Beispielskassen gegenüber der Simulation I in erheblich geringerem Umfang gefallen ist, als der Parameter der Simulation, der (versichertenbezogen) den Anteil des Korrekturbetrages an der Zuweisung regelt. Offenbar dominiert die Annahme über den Anteil der Versicherten mit Fehlern in den an das BVA übermittelten Daten das Geschehen.

¹⁴ Die erforderlichen Berechnungen wurden vom BVA durchgeführt.

Ein Vergleich von Tabelle 3-3 und Tabelle 3-4 mit Tabelle 3-2 macht deutlich, dass die aus der Simulation hervorgegangenen Variationskoeffizienten im Hinblick auf realistische Werte für die Fehlerquoten und den Variationskoeffizienten q des Korrekturbetrages innerhalb VG_F in plausibler Größenordnung liegen.

Tabelle 3-3 Variationskoeffizient des Korrekturbetrages aus der Simulation I in den zehn Beispielskassen

Kasse	A	B	C	D	E	F	G	H	I	J
V	8,3	8,3	8,2	8,7	6,6	6,8	6,8	7,0	6,6	6,8

Tabelle 3-4 Mittlerer Variationskoeffizient des Korrekturbetrages aus Simulationen II bis IV in den zehn Beispielskassen

Kasse	A	B	C	D	E	F	G	H	I	J
V	7,8	7,7	7,8	7,5	7,1	6,5	6,5	6,7	6,1	5,9

Im Folgenden wurde der für die Erreichung einer gewünschten Genauigkeit der Hochrechnung erforderliche Stichprobenumfang für ein als realistisch erscheinendes $q=0,6$ gemäß Formel (26) für verschiedene (klein gewählte) Fehlerquoten tabelliert (vgl. Tabelle 3-5 und Tabelle 3-6). Ein Ablesebeispiel mag den Gebrauch der beiden Tabellen erläutern: Wenn angenommen werden kann, dass die Fehlerquote in der Kasse C $p_F = 0,02$ beträgt und will man den mittleren Korrekturbetrag (KB) bzw. die Summe der Korrekturbeträge mit einem relativen Fehler schätzen, der mit 95%-Wahrscheinlichkeit die Schranke $\varepsilon = 20\%$ nicht überschreitet, so ist ($q=0,6$ vorausgesetzt) für eine einfache Zufallsstichprobe der Umfang $n = 6.360$ erforderlich.

In welcher Höhe der zu tolerierende Fehler (also ε) gewählt und ob er für alle zu prüfenden Kassen gleich gesetzt werden soll (was nahe liegt), ist eine Bewertungsfrage, die sich letztlich einer wissenschaftlichen Analyse entzieht. So ergab sich aus den Simulationen II bis IV für die Kasse A eine Summe der Korrekturbeträge im Mittel in Höhe von 91,9 Mio. Euro und für die Kasse E in Höhe von 50,1 Tsd. Euro. Setzt man nun beispielsweise $\varepsilon = 20\%$, so entspricht das einer möglichen Abweichung der Schätzung aus der Stichprobe in Höhe von rund ± 18 Mio. Euro bei Kasse A und rund ± 10 Tsd. Euro bei Kasse E. Die Spanne zwischen beiden Werten illustriert das o. g. Bewertungsproblem.

Tabelle 3-5 Erforderliche Stichprobenumfänge in den fünf Beispielskassen für die Nr. 1-Stichproben zur Schätzung des KBs in Abhängigkeit von der Fehlerquote p_F und der gewünschten (relativen) Genauigkeit ε ($q=0,6$)

Kasse A	N =	7.058.189				
$p_F \setminus \varepsilon$	5%	10%	15%	20%	25%	30%
0,01	201.516	51.481	22.974	12.941	8.288	5.757
0,02	101.471	25.644	11.421	6.429	4.116	2.859
0,05	40.030	10.050	4.470	2.515	1.610	1.118
Kasse B	N =	1.495.360				
$p_F \setminus \varepsilon$	5%	10%	15%	20%	25%	30%
0,01	182.168	50.121	22.699	12.853	8.252	5.740
0,02	96.320	25.302	11.352	6.407	4.107	2.854
0,05	39.203	9.997	4.460	2.512	1.609	1.117
Kasse C	N =	552.350				
$p_F \setminus \varepsilon$	5%	10%	15%	20%	25%	30%
0,01	150.804	47.409	22.125	12.668	8.175	5.703
0,02	86.777	24.592	11.207	6.360	4.088	2.845
0,05	37.524	9.885	4.437	2.505	1.606	1.116
Kasse D	N =	52.546				
$p_F \setminus \varepsilon$	5%	10%	15%	20%	25%	30%
0,01	41.926	26.100	16.021	10.399	7.166	5.193
0,02	34.790	17.276	9.394	5.732	3.819	2.712
0,05	22.794	8.447	4.122	2.401	1.562	1.095
Kasse E	N =	6.407				
$p_F \setminus \varepsilon$	5%	10%	15%	20%	25%	30%
0,01	6.215	5.702	5.013	4.288	3.615	3.034
0,02	6.032	5.130	4.107	3.210	2.507	1.977
0,05	5.527	3.915	2.634	1.807	1.287	952

Tabelle 3-6 Erforderliche Stichprobenumfänge in den fünf Beispielskassen für die Nr. 2-Stichproben zur Schätzung des KBs in Abhängigkeit von der Fehlerquote p_F und der gewünschten (relativen) Genauigkeit ε ($q=0,6$)

Kasse F	N =	2.804.204				
$p_F \setminus \varepsilon$	5%	10%	15%	20%	25%	30%
0,01	193.151	50.918	22.861	12.905	8.273	5.750
0,02	99.305	25.504	11.393	6.420	4.112	2.857
0,05	39.689	10.029	4.466	2.514	1.609	1.118
Kasse G	N =	783.593				
$p_F \setminus \varepsilon$	5%	10%	15%	20%	25%	30%
0,01	164.019	48.641	22.390	12.754	8.211	5.720
0,02	90.996	24.919	11.274	6.382	4.097	2.849
0,05	38.291	9.937	4.448	2.508	1.607	1.117
Kasse H	N =	126.110				
$p_F \setminus \varepsilon$	5%	10%	15%	20%	25%	30%
0,01	78.430	36.748	19.487	11.756	7.785	5.510
0,02	56.680	21.375	10.488	6.122	3.988	2.796
0,05	30.517	9.321	4.320	2.467	1.590	1.108
Kasse I	N =	21.742				
$p_F \setminus \varepsilon$	5%	10%	15%	20%	25%	30%
0,01	19.679	15.319	11.188	8.122	6.006	4.555
0,02	17.951	11.786	7.495	4.965	3.462	2.527
0,05	14.118	6.880	3.710	2.255	1.499	1.064
Kasse J	N =	1.197				
$p_F \setminus \varepsilon$	5%	10%	15%	20%	25%	30%
0,01	1.190	1.170	1.138	1.096	1.046	991
0,02	1.183	1.144	1.084	1.009	927	844
0,05	1.162	1.070	944	811	687	578

3.5 Anteil der Versicherten mit Fehlern in den übermittelten Daten

Infolge mangelnder Vorkenntnisse und den damit verbundenen Schwierigkeiten, Annahmen über den Variationskoeffizienten des Korrekturbetrages zu treffen, besteht mit Blick auf Gleichungen (13) und (26) ein alternativer Ansatz darin, den Stichprobenumfang so festzulegen, dass man die Fehlerquote p_F aus der zu ziehenden Stichprobe mit vorgegebener Genauigkeit schätzen kann. Zwar hängt der Stichprobenumfang von der unbekannten Größe von p_F ab, aber die Annahme, dass p_F verhältnismäßig klein ist und etwa in einem Bereich zwischen 0,005 (0,5%) und 0,05 (5%) liegt, mag für jemanden, der die Datenverarbeitungsprozesse in den Krankenkassen kennt, plausibler klingen, als jede Annahme über den Variationskoeffizienten V (oder q) des Korrekturbetrages.

Werden k Versicherte mit Fehlern in den an das BVA übermittelten Daten in einer einfachen Zufallsstichprobe vom Umfang n gefunden, so liegt es nahe, p_F folgendermaßen zu schätzen (und zwar erwartungstreu, vgl. Cochran 1972):

$$(27) \quad \hat{p}_F = \frac{k}{n}$$

Die Varianz von $\hat{p}_F$ berechnet sich mit $q_F = 1 - p_F$ wie folgt:

$$(28) \quad \text{Var}(\hat{p}_F) = (1-f) \frac{N}{(N-1)} \frac{p_F q_F}{n} \quad (\text{wobei } (1-f) \frac{N}{(N-1)} = \frac{N-n}{N-1} \text{ gilt})$$

Sie wird erwartungstreu geschätzt durch

$$(29) \quad \hat{\text{Var}}(\hat{p}_F) = (1-f) \frac{n}{(n-1)} \frac{\hat{p}_F \hat{q}_F}{n} = (1-f) \frac{\hat{p}_F \hat{q}_F}{(n-1)} \quad (\text{vgl. Cochran 1972}).$$

Das asymptotische Konfidenzintervall hat die folgende Form:

$$(30) \quad \text{KI}_{1-\alpha}(p_F) = [\hat{p}_F - u_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{N-n}{N-1} \frac{p_F q_F}{n}}, \hat{p}_F + u_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{N-n}{N-1} \frac{p_F q_F}{n}}]$$

Soll nun der Stichprobenumfang so bestimmt werden, dass der relative Fehler bei der Schätzung von p_F mit 95% Wahrscheinlichkeit die Größe ε nicht übersteigt, so ist εp_F der Grenze des Konfidenzintervalls (30) gleichzusetzen und man erhält:

$$(31) \quad n = \frac{\frac{u^2}{\varepsilon^2} \frac{q_F}{p_F}}{1 + \frac{1}{N} \left(\frac{u^2}{\varepsilon^2} \frac{q_F}{p_F} - 1 \right)}, \quad \text{mit } q_F = 1 - p_F \text{ und } u = 1,96 \text{ (wie oben)}$$

Man beachte, dass (31) anders als das zugehörige Konfidenzintervall (30) nicht symmetrisch in p_F und q_F ist, da die Vorgabe der relativen Fehlers mit einem konstanten ε in asymmetrischer Weise an p_F gebunden ist. In (31) wächst vielmehr n monoton mit fallendem p_F , so dass die Schätzung kleiner Anteilswerte große Stichprobenumfänge erfordert (siehe die beiden nachfolgenden Tabellen).

Tabelle 3-7 Erforderliche Stichprobenumfänge in den fünf Beispielskassen für die Nr. 1-Stichproben in Abhängigkeit vom zu schätzenden Fehleranteil p_F und der gewünschten (relativen) Genauigkeit ε

Kasse A	N =	7.058.189				
$p_F \setminus \varepsilon$	10%	15%	20%	25%	30%	40%
0,01	37.827	16.862	9.495	6.080	4.223	2.376
0,02	18.773	8.356	4.703	3.010	2.091	1.176
0,05	7.291	3.242	1.824	1.168	811	456
Kasse B	N =	1.495.360				
$p_F \setminus \varepsilon$	10%	15%	20%	25%	30%	40%
0,01	37.087	16.714	9.448	6.060	4.214	2.373
0,02	18.589	8.319	4.691	3.006	2.089	1.176
0,05	7.263	3.237	1.822	1.167	811	456
Kasse C	N =	552.350				
$p_F \setminus \varepsilon$	10%	15%	20%	25%	30%	40%
0,01	35.581	16.401	9.347	6.019	4.194	2.367
0,02	18.203	8.241	4.666	2.995	2.084	1.174
0,05	7.204	3.225	1.819	1.165	810	456
Kasse D	N =	52.546				
$p_F \setminus \varepsilon$	10%	15%	20%	25%	30%	40%
0,01	22.063	12.789	8.051	5.453	3.911	2.274
0,02	13.859	7.217	4.319	2.848	2.011	1.151
0,05	6.409	3.055	1.763	1.142	799	452
Kasse E	N =	6.407				
$p_F \setminus \varepsilon$	10%	15%	20%	25%	30%	40%
0,01	5.483	4.646	3.828	3.121	2.547	1.734
0,02	4.780	3.629	2.713	2.049	1.577	994
0,05	3.412	2.154	1.420	988	720	426

Tabelle 3-8 Erforderliche Stichprobenumfänge in den fünf Beispielskassen für die Nr. 2-Stichproben in Abhängigkeit vom zu schätzenden Fehleranteil p_F und der gewünschten (relativen) Genauigkeit ε

Kasse F	N =	2.804.204				
$p_F \setminus \varepsilon$	10%	15%	20%	25%	30%	40%
0,01	37.522	16.801	9.475	6.072	4.219	2.375
0,02	18.698	8.341	4.698	3.008	2.090	1.176
0,05	7.280	3.240	1.824	1.167	811	456
Kasse G	N =	783.593				
$p_F \setminus \varepsilon$	10%	15%	20%	25%	30%	40%
0,01	36.270	16.546	9.394	6.038	4.203	2.370
0,02	18.382	8.277	4.678	3.000	2.086	1.175
0,05	7.231	3.231	1.820	1.166	810	456
Kasse H	N =	126.110				
$p_F \setminus \varepsilon$	10%	15%	20%	25%	30%	40%
0,01	29.219	14.905	8.841	5.805	4.089	2.333
0,02	16.379	7.845	4.537	2.941	2.057	1.166
0,05	6.900	3.163	1.799	1.157	806	455
Kasse I	N =	21.742				
$p_F \setminus \varepsilon$	10%	15%	20%	25%	30%	40%
0,01	13.834	9.510	6.615	4.754	3.538	2.143
0,02	10.089	6.041	3.869	2.645	1.908	1.116
0,05	5.465	2.823	1.683	1.108	782	447
Kasse J	N =	1.761				
$p_F \setminus \varepsilon$	10%	15%	20%	25%	30%	40%
0,01	1.683	1.595	1.486	1.366	1.243	1.012
0,02	1.610	1.455	1.282	1.111	956	706
0,05	1.419	1.142	896	702	555	362

Um einen Eindruck davon zu erhalten, mit welchem Fehler die Schätzung des mittleren Korrekturbetrages (und anderer Größen) versehen ist, wenn man sich bei der Stichprobenplanung auf den Anteil p_F der Versicherten mit Fehlern in den übermittelten Daten bezieht, sind aus jeder der zehn Beispielskassen unabhängig voneinander drei Stichproben gezogen worden, deren Umfang so festgelegt wurde, dass man $p_F = 0,05$ mit einer relativen Genauigkeit von $\varepsilon = 10\%$ schätzen kann. Die Ziehung dieser (einfachen) Zufallsstichproben erfolgte (jeweils) in dem gleichen Programmlauf, in dem auch die oben bereits angesprochenen Simulationen II bis IV realisiert wurden. So bezieht sich jede der Stichprobe auf eine andere Simulation, wenn wir Stichprobenergebnisse mit Ergebnissen aus dem simulierten Teil der Grundgesamtheiten vergleichen.

3.5.1 Relative Fehler bei der Schätzung ausgewählter Kennziffern: Nr. 1-Stichproben in der Simulation

Zur Beurteilung der Qualität der Nr. 1-Stichproben erscheint es als zweckmäßig, die folgenden Kennziffern heranzuziehen:

- K_1 : p_F
- K_2 : Mittelwert des Korrekturbetrages
- K_3 : Mittelwert der Versichertentage im Ausgleichsjahr
- K_4 : Mittelwert der Versichertentage mit Krankengeldanspruch im Ausgleichsjahr
- K_5 : Mittelwert der Versichertentage mit DMP-Einschreibung im Ausgleichsjahr
- K_6 : Mittelwert der Versichertentage mit Erwerbsminderung im Ausgleichsjahr
- K_7 : Mittelwert der Zuweisung

In Tabelle 3-9 sind relative Abweichungen um mehr als $\pm 10\%$ rot markiert worden. Solche Abweichungen finden wir in der ersten Stichprobe nur in den kleineren Beispielskassen D und E. In der Kasse D finden wir sogar eine relative Abweichung von p_F in einer Höhe von 10,1%, d. h. eine Überschreitung der für die Planung des Stichprobenumfangs gesetzten Grenze von 10%, was nach dem planerischen Ansatz nur in 5% aller denkbaren Stichproben passieren kann.

Die vergleichsweise großen Abweichungen bei der Schätzung des Mittelwertes der Versichertentage mit Erwerbsminderung im Ausgleichsjahr in den Beispielskassen D und E müssen vor dem Hintergrund beurteilt werden, dass der Anteil der Versicherten mit Erwerbsminderung im Ausgleichsjahr in der Kasse D nur 1,71% und in der Kasse E nur 0,33% beträgt (GKV-weit: 2,5%, vgl. Tabelle 8-11 im Anhang). Für solche niedrigen Anteilswerte war die Stichprobe nicht geplant.

Bei der Schätzung des Korrekturbetrages finden wir in Kasse D (zweite Stichprobe) und in Kasse C (dritte und vierte Stichprobe) relative Fehler von über 10%, was nach einem Vergleich der Werte in der 10%-Spalte der Tabelle 3-5 für $p_F=0,05$ mit den aus Gleichung (24) aber zu erwarten war, wenn man sich erinnert, dass der Variationskoeffizient des Korrekturbetrages in den Simulationen ja noch deutlich oberhalb $V=5$ lag (vgl. auch Tabelle 8-13 im Anhang).

Tabelle 3-9 Relative Abweichung der aus der Stichprobe von den aus der Grundgesamtheit berechneten ausgewählten Kennziffern (Nr. 1- Stichprobe)

Simulation II und zug. Stichprobe							
	Kennziffer						
Kasse	K ₁	K ₂	K ₃	K ₄	K ₅	K ₆	K ₇
A	1,9%	-8,8%	-0,16%	0,9%	-4,5%	2,9%	-2,3%
B	-0,4%	-7,5%	-0,18%	-1,9%	1,1%	-1,6%	-1,9%
C	-2,8%	0,6%	-0,28%	-0,1%	1,7%	8,3%	2,9%
D	10,1%	12,9%	-0,19%	-1,4%	0,5%	12,8%	-0,3%
E	2,5%	9,0%	-0,12%	1,4%	0,0%	-19,3%	-0,4%
Simulation III und zug. Stichprobe							
	Kennziffer						
Kasse	K ₁	K ₂	K ₃	K ₄	K ₅	K ₆	K ₇
A	-4,4%	-4,7%	0,01%	0,6%	-1,4%	-7,0%	-1,7%
B	4,2%	6,5%	0,29%	-5,2%	1,7%	6,2%	-1,1%
C	-3,9%	-17,6%	-0,16%	-1,4%	-4,4%	1,8%	-2,1%
D	0,4%	-9,6%	0,17%	-1,4%	0,9%	3,0%	-2,1%
E	-7,7%	-3,3%	-0,27%	0,5%	0,0%	-1,6%	-0,2%
Simulation IV und zug. Stichprobe							
	Kennziffer						
Kasse	K ₁	K ₂	K ₃	K ₄	K ₅	K ₆	K ₇
A	-1,2%	2,1%	-0,28%	-1,3%	-1,6%	12,6%	1,8%
B	-5,3%	-6,9%	-0,33%	-0,5%	0,0%	-0,9%	2,1%
C	10,8%	11,2%	0,01%	0,2%	-9,8%	-11,8%	0,0%
D	-2,0%	4,6%	0,18%	-3,6%	2,4%	-14,5%	3,7%
E	8,7%	6,4%	-0,41%	0,5%	0,0%	21,2%	0,4%

*) Rot eingefärbt sind relative Abweichungen, die im Absolutbetrag größer sind als 10%.

3.5.2 Relative Fehler bei der Schätzung ausgewählter Kennziffern: Nr. 2-Stichproben in der Simulation

Für die Nr. 2-Stichproben ziehen wir die folgenden Kennziffern heran:

- K₁ : p_F
- K₂ : Mittelwert des Korrekturbetrages
- K₃ : Mittelwert der Versichertentage im Ausgleichsjahr
- K₄ : Mittelwert der Versichertentage mit Auslandsaufenthalt im Vorjahr
- K₅ : Mittelwert der Versichertentage mit Erwerbsminderung im Vorjahr
- K₆ : Mittelwert der Versichertentage mit Kostenerstattung im Vorjahr
- K₇ : Mittelwert der Zuweisung
- K₈ : Mittlere Zahl der belegten HMGs
- K₉ : Mittlere Zahl der relevanten stationären Hautdiagnosen
- K₁₀ : Mittlere Zahl der relevanten stationären Nebendiagnosen
- K₁₁ : Mittlere Zahl der gesicherten ambulanten Diagnosen
- K₁₂ : Mittlere Zahl der Arzneimittelverordnungen mit zuordenbarem ATC-Code

Wie ein Blick auf Tabelle 3-10 lehrt, ist der Mittelwert der Versichertentage mit Kosten-erstattung im Vorjahr besonders schwer zu schätzen. Allerdings ist der Anteil der Versicherten mit Kostenerstattung auch äußerst gering (0,77% / 0,01% / 0,13% / 0,02% / 5,51% in den Beispielskassen F bis J; GKV-weit liegt er bei 0,31%, vgl. Tabelle 8-12 im Anhang). Die -100% relative Abweichung in der ersten und dritten Stichprobe in den Kassen G und I sind folgendermaßen entstanden: In der Grundgesamtheit finden wir sehr kleine Mittelwerte, in der Stichprobe sind sie Null, was daran liegt, dass die wenigen Versicherten mit Kostenerstattung im Vorjahr, die es in den beiden Kassen gibt, in der Stichprobe nicht aufgenommen wurden.

In der zweiten und dritten Stichprobe ist der Mittelwert der Versichertentage mit Auslandsaufenthalt im Vorjahr in den Kassen nur sehr ungenau geschätzt worden. Aber der Anteil der betroffenen Versicherten ist auch in diesem Fall sehr klein (F: 0,43% G: 0,11%, GKV-weit: 0,94%, vgl. Tabelle 8-12 im Anhang).

Tabelle 3-10 Relative Abweichung der aus der Stichprobe von den aus der Grundgesamtheit berechneten ausgewählten Kennziffern (Nr. 2- Stichprobe)

Simulation II und zug. Stichprobe												
	Kennziffer											
Kasse	K ₁	K ₂	K ₃	K ₄	K ₅	K ₆	K ₇	K ₈	K ₉	K ₁₀	K ₁₁	K ₁₂
F	-4,6%	-4,7%	0,09%	2,3%	6,3%	20,4%	0,0%	0,2%	-3,3%	0,9%	-0,3%	-1,4%
G	-4,0%	-2,4%	-0,19%	10,5%	4,5%	-100,0%	0,2%	0,3%	-0,6%	0,3%	0,2%	0,5%
H	-3,0%	-12,5%	-0,04%	-0,6%	-1,5%	-32,2%	0,6%	-0,1%	2,3%	-2,0%	-0,9%	0,0%
I	3,5%	7,8%	-0,07%	-1,7%	2,3%	-100,0%	0,4%	1,2%	3,7%	5,3%	0,3%	1,0%
J	-5,5%	-15,3%	-0,14%	6,7%	-2,8%	-2,0%	-0,6%	-0,6%	4,2%	3,1%	-1,0%	0,4%
Simulation III und zug. Stichprobe												
	Kennziffer											
Kasse	K ₁	K ₂	K ₃	K ₄	K ₅	K ₆	K ₇	K ₈	K ₉	K ₁₀	K ₁₁	K ₁₂
F	-2,9%	-8,6%	0,13%	32,3%	-4,2%	-9,3%	-1,5%	-0,4%	-0,3%	5,1%	-1,6%	-2,4%
G	5,3%	9,3%	-0,09%	-28,3%	-6,8%	166,0%	-1,7%	-0,2%	-8,1%	-1,8%	0,7%	-1,2%
H	-9,5%	-3,7%	0,05%	7,2%	-3,0%	8,3%	-0,1%	-0,1%	-4,0%	-0,6%	0,7%	-0,5%
I	-1,5%	-2,9%	0,05%	-4,6%	-1,1%	59,1%	0,6%	0,0%	1,9%	5,1%	-0,1%	1,3%
J	-1,1%	1,5%	-0,11%	-3,0%	1,6%	-1,3%	0,8%	1,0%	6,2%	2,1%	-0,4%	0,3%
Simulation IV und zug. Stichprobe												
	Kennziffer											
Kasse	K ₁	K ₂	K ₃	K ₄	K ₅	K ₆	K ₇	K ₈	K ₉	K ₁₀	K ₁₁	K ₁₂
F	5,5%	8,1%	-0,09%	-11,0%	-2,2%	-4,4%	0,7%	1,2%	1,2%	1,0%	0,4%	0,0%
G	8,9%	4,3%	-0,07%	28,7%	1,8%	-100,0%	2,1%	1,2%	5,1%	2,0%	1,1%	2,1%
H	-10,7%	-11,3%	0,15%	-3,1%	2,8%	-33,0%	2,2%	0,5%	-3,0%	-6,7%	1,1%	0,3%
I	-1,3%	0,9%	-0,18%	9,8%	-0,1%	-100,0%	2,4%	1,3%	3,7%	5,3%	1,3%	1,8%
J	-1,9%	1,1%	0,06%	-22,4%	-28,1%	3,3%	-0,2%	0,1%	6,7%	4,0%	-0,9%	-1,6%

*) Rot eingefärbt sind relative Abweichungen, die im Absolutbetrag größer sind als 10%.

3.6 Erwarteter Stichprobenumfang bei sequentieller Schätzung der Fehlerquote

Im Rahmen des „inverse sampling“ wird die Fehlerquote aus der Gleichung (2) geschätzt (vgl. Abschnitt 2.5). Das Stichprobenverfahren ist abgeschlossen, wenn eine vorgegebene Anzahl k_0 von fehlerhaften versichertenbezogenen Datensätzen gefunden

wurde. Zur Festlegung von k_0 benötigen wir die Varianz der Schätzung

$$\hat{p}_F = \frac{k-1}{n-1}$$

aus (2). Diese berechnet sich folgendermaßen (vgl. Bauer et al. 1986):

$$(32) \quad S_{\hat{p}_F}^2 = \frac{\hat{p}_F(1-\hat{p}_F)}{n-2} = \frac{(k-1)(n-k)}{(n-1)^2(n-2)} \quad (k > 2)$$

Soll nun so k_0 wie im vorigen Abschnitt bestimmt werden, so ist ε_{p_F} dem Produkt aus u und der Wurzel aus (32) gleichzusetzen (mit u = Quantil der Standardnormalverteilung vgl. Abschnitt 3.4) und wir erhalten:

$$(33) \quad u \frac{S_{\hat{p}_F}}{\hat{p}_F} = u \sqrt{\frac{(1-\hat{p}_F)}{k_0-1-\hat{p}_F}} = u \sqrt{\frac{(n-k_0)}{(k_0-1)(n-2)}} = \varepsilon \quad (k > 2)$$

Für kleine Fehlerquoten vereinfacht sich dieser Ausdruck zu

$$(34) \quad u \frac{S_{\hat{p}_F}}{\hat{p}_F} \approx u \sqrt{\frac{1}{k_0-1}} = \varepsilon \quad (k > 2) \quad (\text{vgl. Bauer et al. 1986}).$$

Aus (34) erhält man k_0 :

$$(35) \quad k_0 = \left(\frac{u}{\varepsilon}\right)^2 + 1$$

Der resultierende Stichprobenumfang n ist eine Zufallsvariable. Erwartungswert und Varianz von n berechnen sich folgendermaßen:

$$(36) \quad E(n) = \frac{k_0}{p_F}$$

$$(37) \quad \text{Var}(n) = \frac{k_0(1-p_F)}{p_F^2}$$

(vgl. Bauer et al. 1986).

Wählt man nun z. B. $\varepsilon = 40\%$ so ergibt sich aus (34) $k_0 = 25$ und aus (36) $E(n) = 2.500$ für $p_F = 0,01$ oder $E(n) = 1.250$ für $p_F = 0,02$ oder $E(n) = 500$ für $p_F = 0,05$.

Der Vergleich dieser bei sequentieller Vorgehensweise zu erwartenden Stichprobenumfänge mit den in Tabelle 3-7 und Tabelle 3-8 für die gleichen Eingangsparameter ausgewiesenen zeigt für die hier betrachteten endlichen Grundgesamtheiten infolge der Endlichkeitskorrektur kein Vorteil der sequentiellen Vorgehensweise gegenüber der einfachen Zufallsstichprobe mit festem Stichprobenumfang.

4 Planung und Hochrechnung einer geschichteten Zufallsstichprobe

4.1 Bezeichnungen

In der h-ten Schicht betrachten wir als schichtspezifische Größen jeweils:

N_h	Schichtumfang (Zahl der Untersuchungseinheiten)
$W_h = \frac{N_h}{N}$	Schichtgewicht
n_h	Stichprobenumfang
$f_h = \frac{n_h}{N_h}$	Auswahlsatz
$\bar{Y}_h = \frac{1}{N_h} \sum_{i=1}^{N_h} Y_i$	Schichtmittelwert
$\bar{y}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} y_i$	Stichprobenmittelwert
$S_h^2 = \frac{1}{N_h - 1} \sum_{i=1}^{N_h} (Y_{hi} - \bar{Y}_h)^2$	Schichtvarianz
$s_h^2 = \frac{1}{n_h - 1} \sum_{i=1}^{n_h} (y_{hi} - \bar{y}_h)^2$	Stichprobenvarianz

Ist die Grundgesamtheit in L Schichten zerlegt, so gilt offenbar:

$$(38) \quad \sum_{h=1}^L N_h = N, \quad \sum_{h=1}^L n_h = n, \quad \sum_{h=1}^L W_h \bar{Y}_h = \bar{Y}$$

4.2 Schätzung des Mittelwertes und der Varianz

Da jeder schichtspezifische Stichprobenmittelwert den Mittelwert der zugehörigen Schicht konsistent und unverzerrt schätzt, ergibt sich aus (38) unmittelbar, dass

$$(39) \quad \bar{y}_{st} = \sum_{h=1}^L W_h \bar{y}_h$$

eine konsistente und unverzerrte Schätzung des Gesamtmittels ist, wobei der Index „st“ auf die geschichtete (stratifizierte) Stichprobe verweisen soll.

Weil die Auswahlen in verschiedenen Schichten unabhängig voneinander vorgenommen wurden, gilt für die Varianz dieses Schätzers die Gleichung

$$(40) \quad \text{Var}(\bar{y}_{st}) = \sum_{h=1}^L W_h^2 \text{Var}(\bar{y}_h),$$

aus der dann unmittelbar die Varianzformel der geschichteten Stichprobe folgt:

$$(41) \quad \text{Var}(\bar{y}_{st}) = \sum_{h=1}^L (1 - f_h) W_h^2 \frac{S_h^2}{n_h} = \sum_{h=1}^L \left(\frac{1}{n_h} - \frac{1}{N_h} \right) W_h^2 S_h^2$$

Geschätzt wird die Varianz wieder in konsistenter und unverzerrter Weise, indem in (41) anstelle der Schichtvarianzen der Grundgesamtheit die zugehörigen schichtspezifischen Stichprobenvarianzen eingesetzt werden (vgl. Cochran 1972).

4.3 Potentielle Schichtungsvariablen

Obwohl mehrdimensionale Schichtungen (kreuzklassifiziert nach zwei oder mehreren Schichtungsvariablen) grundsätzlich möglich und auch schon untersucht worden sind (vgl. Drexl 1982), findet man sie in der Praxis außerordentlich selten, und für die erstmalige Planung der Stichproben nach § 42 RSAV kommen sie mangels Vorwissens über die Verteilung des Korrekturbetrages nicht in Betracht. Es ist schwer genug, eine einzelne plausible Schichtungsvariable zu finden. Hauptkandidat sowohl für die Nr. 1-, als auch für die Nr. 2-Stichprobe ist die versichertenbezogene Zuweisung, da man zwar a priori keine direkte Proportionalität aber doch eine namhafte positive Korrelation zwischen der Zuweisung und dem Korrekturbetrag erwarten kann.

Von den anderen Merkmalen, die Gegenstand der Prüfung sind, kommen die Versichertentage im Ausgleichsjahr wegen einer viel zu geringen Variabilität nicht infrage (im Gesamtkollektiv der GKV-Versicherten, deren Daten im Rahmen der Nr. 1-Stichprobe zu untersuchen sind, finden wir einen Variationskoeffizienten in Höhe von 0,16 und bezogen auf die Nr. 2-Stichprobe in Höhe von 0,14).

In Bezug auf andere relevante Merkmale findet man statistisch signifikante, vielfach hohe Korrelationen¹⁵ (und noch höhere Rangkorrelationen¹⁶), wenn man die Verteilung der Mittelwerte der in die Prüfverfahren involvierten Merkmale über die zwanzig vom BVA dokumentierten Schichten der jeweiligen Grundgesamtheit zueinander in Beziehung setzt (vgl. Tabelle 4-1 und Tabelle 4-2).

Für die Untersuchung des potentiellen Schichtungsgewinns wurde in dem vorliegenden Gutachten daher der versichertenbezogene Zuweisungsbetrag als Schichtungsvariable herangezogen.

Es ist darauf hinzuweisen, dass im Rahmen der ersten Umsetzung des Verfahrens nach § 42 RSAV eine empirisch fundierte Begründung für die Auswahl von Schichtungsvariablen erst aus den Ergebnissen der empfohlenen Pilotphase (vgl. Abschnitt 7.4) abgeleitet werden kann.

¹⁵ Korrelationskoeffizient nach Pearson

¹⁶ Korrelationskoeffizient nach Spearman

Tabelle 4-1 Korrelation der mittleren Zuweisung mit dem Mittelwert ausgewählter Variablen (als Merkmale der 20 Schichten) für die Versicherten der GKV-insgesamt, Nr. 1-Stichprobe*)

Mittelwert des Merkmals	Korrelation nach	
	Pearson	Spearman
Versichertentage (V-Tage)	0,171	0,356
Alter	0,433	0,860
V-Tage mit KG-Anspruch im Ausgleichsjahr	-0,484	-0,875
V-Tage mit DMP-Einschreibung im Ausgleichsjahr	0,494	0,917
V-Tage mit Erwerbsminderung im Ausgleichsjahr	0,845	0,995

*) Rot eingefärbt sind statistisch auf dem 5%-Niveau nicht signifikanten Werte

Tabelle 4-2 Korrelation der mittleren Zuweisung mit dem Mittelwert ausgewählter Variablen (als Merkmale der 20 Schichten) für die Versicherten der GKV-insgesamt, Nr. 2-Stichprobe*)

Mittelwert des Merkmals	Korrelation nach	
	Pearson	Spearman
Alter	0,209	0,508
Zahl der belegten HMGs	0,798	0,995
Zahl der V-Tage mit Auslandsaufenthalt	-0,422	-0,737
Zahl der V-Tage mit Erwerbsminderungsrente	0,742	0,991
Zahl der V-Tage mit Kostenerstattung	-0,333	-0,872
Zahl der unterschiedlichen relevanten ambulanten Diagnosen	0,809	1,000
Zahl der unterschiedlichen relevanten stationären Hauptdiagnosen	0,913	0,973
Zahl der unterschiedlichen relevanten stationären Nebendiagnosen	0,929	0,970
Zahl der relevanten Arzneimittelverordnungen mit unterschiedlichem ATC-Code	0,851	0,994

*) Rot eingefärbt sind statistisch auf dem 5%-Niveau nicht signifikanten Werte; mit „relevant“ ist gemeint, dass es sich um Diagnosen handelt, die zu dem im Morbi-RSA berücksichtigten Krankheiten gehören.

4.4 Wahl der Schichtenzahl und der Schichtgrenzen

Es gibt aufwendige Verfahren der nichtlinearen und dynamischen Optimierung, mit denen bei vorgegebener Zahl der Schichten und vorgegebenem Gesamtstichprobenumfang die Schichtgrenzen und die Stichprobenumfänge simultan bestimmt werden können (vgl. Drexl 1982). Diese haben sich wegen des hohen Programmier- und Rechenaufwandes in der Praxis jedoch nicht durchgesetzt. Vielmehr ist es üblich, sukzessiv erst die Schichtenzahl und die Schichtgrenzen und dann die Verteilung des Gesamtstichprobenumfangs auf die Schichten zu bestimmen.

Was die Zahl L der Schichten anbetrifft, so wird in der Regel iterativ vorgegangen. Dabei beginnt man mit $L = 2$ und erhöht die Anzahl der Schichten solange um eins, bis entweder der Schichtungseffekt hinreichend groß ist, oder eine vorgegebene Abbruchschranke für den Zugewinn an Genauigkeit durch weiter Erhöhung erreicht wird (Drexl 1982).

Als Schichtungseffekt bezeichnen wir dabei die relative Abweichung des Standardfehlers der untersuchten Kennziffer, bezogen auf die geschichtete Stichprobe, von dem

Standardfehler, der sich unter Bezug auf die einfache Zufallsstichprobe ergibt.

Cochrane kommt aufgrund theoretischer und empirischer Untersuchungen zu dem Schluss, „dass es nur in wenigen Fällen vorteilhaft sein wird, mehr als sechs Schichten zu bilden“ (Cochrane 1972).

Zur Bestimmung der Schichtgrenzen $x_1 < x_2 < \dots < x_L$ wurden verschiedene Verfahren diskutiert. Von Dalenius und Gurney (1951) stammt der Vorschlag, die Schichten so zu bilden, dass

$$(42) \quad W_h \cdot S_h = \text{konstant} \quad \text{für } h = 1, \dots, L$$

gilt. Eine andere Konstanzregel empfiehlt Ekman (1959):

$$(43) \quad W_h \cdot (x_h - x_{h-1}) = \text{konstant} \quad \text{für } h = 1, \dots, L$$

Weitere Schichtungsvorschläge basieren auf speziellen Verteilungen des Erhebungsmerkmals (für einen Überblick siehe Drexl 1982, nach dem auch die beiden o. g. Vorschläge zitiert wurden).

Man kann sich bei der Bildung der Schichten auch einfacher heuristischer Verfahren bedienen, wenn man beachtet, dass der Schichtungseffekt bei optimaler Aufteilung des Stichprobengesamtumfangs auf die Schichten (s. u.) umso größer ist, je stärker die Schichtmittelwerte oder je stärker die Schichtvarianzen streuen. Am meisten profitiert man von einer geschichteten Stichprobe mit optimaler Aufteilung, wenn beide Populationsparameter - sowohl die Schichtmittelwerte als auch die Schichtvarianzen - eine erhebliche Streuung aufweisen (vgl. Sukhatme und Sukhatme, 1970, Abschnitt 3.5).

Für das vorliegende Gutachten wurden die Schichten mit dem so geschärften Blick auf die bereits vom BVA in zwanzig Strata voraggregierten Daten durch Studium der Variation der Mittelwerte und Standardabweichungen des Zuweisungsbetrages über die Strata auf heuristische Weise festgelegt. Dabei wurde angestrebt, für die Kassen der Nr.1- und Nr. 2-Stichproben jeweils möglichst ähnliche Schichtungskonzepte zu verwenden.

Im Ergebnis fanden zwei verschiedene Schemata der Zusammenfassung der zwanzig ursprünglichen Schichten auf fünf neue Schichten Anwendung (vgl. Abbildung 4-1 und Abbildung 4-2). Dabei wurde das Zuordnungsschema I bei Aggregation der Schichten in der Kasse A der Nr. 1-Stichprobe und der Kasse F in der Nr. 2-Stichprobe eingesetzt. In allen anderen Kassen wurde das Zuordnungsschema II angewendet.

Abbildung 4-1 Zuordnungsschema I für die Zusammenfassung der 20 Schichten

	Alte Schichten																			
neu	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	■																			
2		■	■	■	■	■	■	■	■	■										
3										■	■	■	■	■	■	■				
4																■	■	■	■	
5																			■	■

Abbildung 4-2 Zuordnungsschema II für die Zusammenfassung der 20 Schichten

	Alte Schichten																			
neu	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	■																			
2		■	■	■	■	■	■	■	■	■										
3										■	■	■	■	■	■					
4																■	■	■	■	
5																			■	■

Bei der Aggregation von k alten zu einer neuen Schicht sind die im Folgenden zusammengestellten Regeln zu beachten.

Aggregationsregeln für die Zusammenfassung von k Schichten

- Absolute Zahlen (Zählzahlen) und Schichtsummen werden addiert.
- Mittelwerte werden mit den Pseudonymzahlen gewichtet gemittelt.

Dies sei am Beispiel für $k = 2$ in einer Formel dargestellt:

$$\bar{X}_{\text{Pool}} = \frac{n_1 \bar{X}_1 + n_2 \bar{X}_2}{n_1 + n_2}$$

- Am schwierigsten ist die Berechnung der gepoolten Varianz. Hier gilt die Streuungszerlegung, d. h. die Gesamtvarianz setzt sich additiv zusammen aus der Varianz innerhalb der Schichten plus der Varianz zwischen den Schichten. Auch hier hilft eine Formel für den Spezialfall $k = 2$ weiter:

$$s_{\text{Pool}}^2 = \frac{(n_1 - 1) \cdot s_1^2 + (n_2 - 1) \cdot s_2^2}{n_1 + n_2 - 2} + \frac{n_1 \cdot (\bar{X}_1 - \bar{X})^2 + n_2 \cdot (\bar{X}_2 - \bar{X})^2}{n_1 + n_2} \quad \left(\text{mit } \bar{X} = \frac{n_1 \bar{X}_1 + n_2 \bar{X}_2}{n_1 + n_2} \right)$$

$\uparrow$

Gesamte
Varianz

$\uparrow$

Varianz innerhalb
der Schichten

$\uparrow$

Varianz zwischen
den Schichten

- Für allgemeines k gelten folgende Formeln:

$$(44) \quad \bar{X}_{\text{Pool}} = \frac{\sum_{i=1}^k n_i \bar{X}_i}{\sum_{i=1}^k n_i} \quad s_{\text{Pool}}^2 = \frac{\sum_{i=1}^k (n_i - 1) s_i^2}{\sum_{i=1}^k n_i - k} + \frac{\sum_{i=1}^k n_i (\bar{X}_i - \bar{X}_{\text{Pool}})^2}{\sum_{i=1}^k n_i}$$

Unter Beachtung dieser Regeln wurden die zwanzig Schichten in jeder der zehn Bei-

spielskassen und den beiden GKV-Gesamtkollektiven nach der Aggregationsvorschrift gemäß den beiden oben dargestellten Zuordnungsschemata zu jeweils fünf neuen Schichten aggregiert.

Mittelwerte und Standardabweichungen des Zuweisungsbetrages weisen, wie angestrebt, einen erheblichen Anstieg über die neuen Schichten auf, wie Tabelle 4-3 und Tabelle 4-4 exemplarisch für das jeweilige GKV-Gesamt zeigen. Die Verhältnisse in den zehn Beispielskassen weichen nur unwesentlich von diesem Bild ab.

Tabelle 4-3 GKV-Gesamt (in Bezug auf Nr. 1- Stichprobe), neu geschichtet

Schicht (neu)	Schichten (alt)	Pseudonyme	Schichtgewicht W_h	Zu weisung	
				Mittelwert	Stnd.abw.
1	1	13.167.364	0,181	590 €	189 €
2	2 bis 10	47.021.882	0,647	1.488 €	623 €
3	11 bis 15	8.862.878	0,122	4.385 €	870 €
4	16 bis 18	2.814.567	0,039	8.284 €	1.451 €
5	19 und 20	812.708	0,011	19.126 €	10.722 €
Insgesamt	1 bis 20	72.679.399	1,000	2.139 €	2.780 €

Tabelle 4-4 GKV-Gesamt (in Bezug auf Nr. 2- Stichprobe), neu geschichtet

Schicht (neu)	Schichten (alt)	Pseudonyme	Schichtgewicht W_h	Zu weisung	
				Mittelwert	Stnd.abw.
1	1	3.810.934	0,140	1.436 €	408 €
2	2 bis 10	16.894.985	0,620	2.916 €	789 €
3	11 bis 15	4.508.496	0,165	6.070 €	978 €
4	16 bis 18	1.574.464	0,058	10.429 €	1.670 €
5	19 und 20	458.611	0,017	23.870 €	12.314 €
Insgesamt	1 bis 20	27.247.490	1,000	4.018 €	3.830 €

4.5 Aufteilung des Gesamtstichprobenumfangs auf die Schichten

Erfolgt die Aufteilung der Gesamtstichprobe auf die Schichten proportional zu der Aufteilung der Grundgesamtheit auf diese, d. h. gilt

$$(45) \quad \frac{n_h}{n} = \frac{N_h}{N} = W_h,$$

so spricht man von einer *proportional geschichteten Stichprobe*. Diese hat einen über die Schichten hinweg konstanten Auswahlssatz. In diesem Fall vereinfachen sich die Schätzformeln erheblich. So lässt sich $\bar{y}_{st}$ unmittelbar als Stichprobenmittel der „gepoolten“ Gesamtstichprobe berechnen, da bei proportionaler Aufteilung offenbar gilt:

$$(46) \quad \bar{y}_{st} = \sum_{h=1}^L W_h \bar{y}_h = \sum_{h=1}^L W_h \frac{1}{n_h} \sum_{i=1}^{n_h} y_{hi} = \sum_{h=1}^L W_h \frac{1}{n W_h} \sum_{i=1}^{n_h} y_{hi} = \frac{1}{n} \sum_{h=1}^L \sum_{i=1}^{n_h} y_{hi}$$

Diese Eigenschaft wird als *Selbstgewichtung* bezeichnet.

Wegen der Gültigkeit von

$$(47) \quad n_h = nW_h \quad \text{und} \quad \frac{n_h}{N_h} = \frac{n}{N}$$

vereinfacht sich die Varianz bei proportionaler Aufteilung zu

$$(48) \quad \text{Var}(\bar{y}_{st}) = (1-f) \sum_{h=1}^L W_h \frac{S_h^2}{n} = \left(\frac{1}{n} - \frac{1}{N}\right) \sum_{h=1}^L W_h S_h^2.$$

Soll nicht das Gesamtmittel, sondern die Merkmalssumme geschätzt werden, so berechnet sich diese nach Schichtung der Grundgesamtheit als mit den Schichtumfängen gewichtete Summe der Schichtmittelwerte und wird deswegen konsistent und erwartungstreu geschätzt durch $\sum_{h=1}^L N_h \bar{y}_h$. Die zugehörige Varianz berechnet sich aus

$$(49) \quad \text{Var}\left(\sum_{h=1}^L N_h \bar{y}_h\right) = \sum_{h=1}^L N_h^2 (1-f_h) \frac{S_h^2}{n_h}.$$

(vgl. Cochran 1972).

Die proportionale Schichtung hat viele Vorteile.¹⁷ Dennoch hat sich das Interesse der Stichprobentheoretiker naturgemäß auf diejenige Aufteilung gerichtet, mit der die Varianz minimiert, der Schichtungseffekt maximiert wird. Unter Zugrundelegung einer Kostenfunktion stellt sich in diesem Zusammenhang eine Optimierungsaufgabe mit Nebenbedingungen, die eine besonders einfache Lösung findet, wenn die stratumsspezifischen Stückkosten alle gleich sind:

$$(50) \quad n_h = n \frac{N_h S_h}{\sum_{h=1}^L N_h S_h}$$

Eine Aufteilung gemäß (50) wird nach ihren Entdeckern als *Neymann-Tschuprow-Allokation* oder auch kurz nur als *Neymann-Allokation* bezeichnet. Diese wird vielfach auch ohne Bezugnahme auf eine Kostenfunktion als optimal bezeichnet.

Ein Nachteil der Optimallösung besteht darin, dass (50) die unbekannten Schichtvarianzen enthält, so dass man diese aus Vorerhebungen schätzen oder von geeigneten Annahmen über ihre Relationen untereinander ausgehen muss. Darüber hinaus ist bei kleinen Grundgesamtheiten nicht auszuschließen, dass sich Auswahlätze größer als Eins ergeben, weil die gefundenen Stichprobenumfänge größer sind, als die zugehörigen Schichtumfänge. Für solche Strata ist dann eine Vollerhebung durchzuführen. Die überschüssig geplanten Stichprobenelemente verteilt man dann am besten auf die übrigen Schichten nach einem Schlüssel der dem Verhältnis der berechneten n_h entspricht.

¹⁷ Wenn die Mittelwerte zwischen den Schichten große Unterschiede aufweisen, die Varianzen aber nicht, so ist sie darüber hinaus dicht am Optimum (vgl. Cochran 1970).

4.6 Abschätzung des erreichbaren Schichtungseffekts

Wie bereits in Abschnitt 4.4 beschrieben, definieren wir den Schichtungseffekt η wie folgt:

$$(51) \quad \eta = \frac{\sqrt{\text{Var}(\bar{y}_{\text{st}}) - \text{Var}(\bar{y}_{\text{einf}})}}{\sqrt{\text{Var}(\bar{y}_{\text{einf}})}} = \sqrt{\frac{\text{Var}(\bar{y}_{\text{st}})}{\text{Var}(\bar{y}_{\text{einf}})}} - 1$$

Damit weichen wir von der üblichen Definition ab, die sich unmittelbar auf die Varianzen (und nicht auf die Standardabweichungen) bezieht (vgl. Sukhatme und Sukhatme 1970, Abschnitt 3.7). Der Vorteil der Definition (51) ist, dass mit dem Standardfehler, anders als mit der Varianz, die Präzision der Schätzung direkt gemessen wird.

Wenn man im für die Planung günstigsten (aber nicht realistischen) Fall unterstellt, dass der Korrekturbeitrag versichertenbezogen direkt (konstant) proportional zur Zuweisung ist, lassen sich für eine Stichprobe beliebigen Umfangs die Stichprobenmittelwerte bzw. –standardabweichungen unmittelbar aus den schichtspezifischen Mittelwerten und Standardabweichungen der Zuweisung ableiten und der Schichtungseffekt bei optimaler Allokation des Gesamtstichprobenumfangs berechnen.

Für die im Folgenden dargelegte Abschätzung des Schichtungseffektes wurden für die zehn Beispielsklassen die gleichen Stichprobengesamtumfänge verwendet, die schon für die Diskussion der einfachen Zufallsstichprobe zugrunde gelegt wurden (vgl. Abschnitt 3.5), und gemäß Formel (50) jeweils auf die fünf Schichten verteilt. Dabei ergaben sich in den Beispielsklassen D (in der Schicht 6), E (in den Schichten 5 und 6) und J (in den Schichten 4, 5 und 6) zunächst Auswahlsätze von über 100%, die dann in einem zweiten Schritt nach dem weiter oben beschriebenen Verfahren bereinigt wurden. Die resultierenden strataspezifischen Stichprobenumfänge sind in Tabelle 4-5 zusammengestellt.

Wir finden in diesem Best-Case-Szenario traumhafte Schichtungseffekte, die in den zehn Beispielsklassen von -71,8% bis zu -86,6% reichen (vgl. Tabelle 4-6). Für die Klasse A bedeutet das beispielsweise, dass der Standardfehler der Schätzung des Mittelwerts (und auch der Summe) der Korrekturbeiträge, berechnet aus der optimal geschichteten Stichprobe, um rund 74% kleiner wäre, als der Standardfehler der Schätzung aus einer einfachen Zufallsstichprobe betragen würde, wenn die Annahmen des Szenarios erfüllt wären. Die besonders hohen Schichtungseffekte in den Beispielsklassen E und J hängen sicher damit zusammen, dass es hier in einigen Schichten zu einer Vollerhebung gekommen ist (mit einem schichtspezifischen Standardfehler = 0).

Sofern wir die Werte der Tabelle 4-6 als obere Grenze für maximal erreichbare Schichtungseffekte interpretieren, so hilft die in Abschnitt 3.4 beschriebene Simulation I weiter, wenn wir eine untere Grenze für die potentiellen Schichtungseffekte suchen. Der Schichtungseffekt bezieht sich in diesem Fall auf den versichertenbezogen simulierten Korrekturbeitrag, wobei in den Mittelwert auch noch der simulierte Anteil der Versicherten mit Fehlern in den übermittelten Daten eingeht. Allerdings ändern sich auch die Ver-

teilungen der Gesamtstichprobenumfangs gemäß Neymann-Allokation auf die Schichten gegenüber den in Tabelle 4-5 ausgewiesenen (s. Tabelle 4-7).

Tabelle 4-5 Stichprobenumfänge nach Neymann-Allokation für die (neu) geschichteten Stichproben bei der Zuweisungssumme proportionalem Korrekturbetrag

Nr. 1-Stichproben					
Schicht / Kasse	A	B	C	D	E
1	330	322	571	276	413
2	4.024	4.565	2.660	4.061	1.585
3	1.624	843	1.368	961	938
4	794	491	898	515	377
5	519	1.043	1.707	596	99
Insgesamt	7.291	7.263	7.204	6.409	3.412
Nr. 2-Stichproben					
Schicht / Kasse	F	G	H	I	J
1	385	409	539	331	133
2	3.484	3.711	2.632	3.202	563
3	1.827	1.060	1.302	997	239
4	1.005	677	760	563	72
5	579	1.373	1.667	372	21
Insgesamt	7.280	7.231	6.900	5.465	1.028

Tabelle 4-6 Schichtungseffekte im Best-Case-Szenario für die Planung

Nr. 1-Stichproben					
Beispielkasse	A	B	C	D	E
Schichtungseffekt	-73,8%	-71,8%	-78,8%	-76,9%	-82,5%
Nr. 2-Stichproben					
Beispielkasse	F	G	H	I	J
Schichtungseffekt	-73,8%	-73,2%	-78,7%	-78,1%	-86,6%

Bei Verwendung des gemäß den Annahmen der Simulation I simulierten Korrekturbetrages geht der Schichtungseffekt gegenüber den in Tabelle 4-6 ausgewiesenen Werten erheblich zurück auf Werte zwischen -25,2% für die Beispielskasse F und -58,0% für die Beispielskasse J (s. Tabelle 4-8), wobei letzter wiederum deswegen vergleichsweise hoch ausgefallen ist, weil es zu Vollerhebungen in den Schichten 4 und 5 gekommen ist. In der Kasse E sind die Schichten 4 und 5 und in der Kasse D die Schicht 5 ebenfalls voll erhoben worden.

Es bleibt festzuhalten, dass auch das vermutete Worst-Case-Szenario noch mit erwähnenswerten Schichtungseffekten aufwarten kann, die um die -30% herum streuen.

Tabelle 4-7 Stichprobenumfänge nach Neymann-Allokation zur Erhebung des in der Simulation I erzeugten Korrekturbetrages

Nr. 1-Stichproben					
Schicht / Kasse	A	B	C	D	E
1	339	336	352	303	199
2	3.562	3.553	3.365	3.085	1.793
3	2.006	1.646	1.678	1.519	944
4	942	912	953	906	377
5	442	817	856	596	99
Insgesamt	7.291	7.263	7.204	6.409	3.412
Nr. 2-Stichproben					
Schicht / Kasse	F	G	H	I	J
1	340	332	324	253	41
2	3.374	3.295	3.126	2.626	651
3	2.078	1.732	1.686	1.348	239
4	1.112	1.121	1.058	866	72
5	376	751	705	372	21
Insgesamt	7.280	7.231	6.900	5.465	1.028

Tabelle 4-8 Schichtungseffekte bei einer Verteilung des Korrekturbetrages gemäß Simulation I

Nr. 1-Stichproben					
Kasse	A	B	C	D	E
Schichtungseffekt	-35,8%	-34,3%	-35,7%	-43,0%	-30,8%
Nr. 2-Stichproben					
Kasse	F	G	H	I	J
Schichtungseffekt	-25,2%	-24,7%	-27,6%	-29,7%	-58,0%

4.7 Nachträgliche Schichtung

Auch wenn nur eine einfache Zufallsstichprobe gezogen wurde, lässt sich durch nachträgliche Schichtung ein Schichtungseffekt realisieren, d. h. die Varianz der Schätzung verkleinern. Die nachträgliche Schichtung wird mit Vorteil angewendet, wenn – wie bei den erstmaligen Prüfungen nach § 42 RSAV – a priori kein gutes Schichtungsmerkmal bekannt ist. Aus den Daten, welche aus einer einfachen Zufallsstichprobe gewonnen wurden, kann beispielsweise die Korrelation des versichertenbezogenen Korrekturbetrages mit anderen, potentiell als Schichtungsmerkmale infrage kommenden Merkmalen geschätzt und auf der Basis solcher Korrelationen dann das am besten geeignete Schichtungsmerkmal ausgewählt werden. Man wird dann die Verteilung des auf Stichprobenbasis ausgewählten Merkmals in der Grundgesamtheit studieren und darauf aufbauend die Schichtgrenzen sowie die (auch in diesem Verfahren benötigten) Schichtgewichte W_h bestimmen. Im nächsten Schritt müssen die gezogenen Versicherten den Schichten zugeordnet werden, woraus sich (als zufällige Größen) die schichtspezifischen Stichprobenumfänge n_h ergeben, die dann in bekannten Formeln der geschichteten Stichprobe eingesetzt werden.

Bei diesem Verfahren wird der Populationsmittelwert nach wie vor erwartungstreu geschätzt (s. Sukhatme und Sukhatme, 1970, Abschnitt 3.8). Ferner sind bei nicht allzu kleinen strataspezifischen Stichprobenumfängen die Genauigkeitsverluste gegenüber einem Verfahren der proportional geschichteten Zufallsauswahl im allgemeinen sehr gering, wie die Untersuchung einer Näherungsformel der Varianz zeigt, die sich aus einer Taylorentwicklung ergibt (zu Details s. Sukhatme und Sukhatme, 1970, Abschnitt 3.8):

$$(52) \quad \text{Var}(\bar{y}_{st}) \cong \left(\frac{1}{n} - \frac{1}{N}\right) \sum_{h=1}^L W_h S_h^2 + \frac{1}{n^2} \sum_{h=1}^L (1 - W_h) S_h^2$$

Der erste Term ist die Varianz der Mittelwertschätzung einer proportional geschichteten Stichprobe. Der zweite Term stellt eine Korrektur hinsichtlich der nachträglichen Schichtung dar, die mit wachsendem Stichprobenumfang n bei festgehaltenen Schichtgewichten und -varianzen gegen Null geht.

4.8 Vorteile einer Schichtung mit Neymann-Allokation bei Erfassung teurer HMGs

Hierarchische Morbiditätsgruppen (HMGs) mit hohen Zuschlägen gehen, wenn bei der Übermittlung der Daten an das BVA Fehler aufgetreten sind, in der Nr. 2-Prüfung potenziell mit hohen Korrekturbeträgen einher. Gleichzeitig werden sie den Versicherten verhältnismäßig selten zugewiesen und laufen Gefahr, in Stichproben – gemessen an ihrer Bedeutung – untererfasst zu werden

Eine Übersicht über die zehn HMGs, die im Jahresausgleich 2009 die Rangliste der HMGs nach Zuschlagshöhe angeführt haben, enthält Tabelle 4-9, und in Tabelle 4-10 sind die zehn seltensten HMGs des Jahresausgleiches 2009 aufgelistet. Die HMGs 004 (Myeloische Leukämie), 043 (Hämophilie) und 107 (Mukoviszidose) sind in beiden Top-ten-Listen vertreten.

Tabelle 4-9 Die zehn teuersten HMGs im Klassifikationsmodell 2009, nach absteigender Zuschlagshöhe geordnet

HMG	Bezeichnung	Pseudonymanzahl	Zuschlag
HMG043	Hämophilie	11.732	55.562,14 €
HMG130	Dialysestatus	75.926	36.050,12 €
HMG107	Mukoviszidose	7.386	17.300,21 €
HMG004	Myeloische Leukämie	23.420	16.531,20 €
HMG008	Lungenmetastasen und Metastasen der Verdauungsorgane	108.929	14.232,22 €
HMG009	Metastasen sonstiger Lokalisation, Kaposi-Sarkom	65.222	10.587,70 €
HMG001	HIV/AIDS	59.861	10.532,23 €
HMG006	Multiples Myelom, akute lymphatische Leukämie	38.827	9.657,99 €
HMG174	Transplantation eines wichtigen Organs, Status des Empfängers	38.033	8.091,92 €
HMG067	Quadruplegie, andere ausgeprägte Lähmungen	124.907	8.080,96 €

Aufgrund der ständigen Weiterentwicklung und Ausdifferenzierung des Klassifikationsmodells sind in späteren Ausgleichsjahren noch deutlich mehr Hochkostengruppen mit teilweise auch erheblich höheren Zuschlägen zu verzeichnen. Im Klassifikationsmodell 2011 beispielsweise gibt es nach Angaben des BVA bereits zwölf HMGs mit Zuschlägen über 10.000 Euro und der höchste liegt bei ca. 272.000 Euro. Die in diesem Abschnitt untersuchte Problematik einer Prüfung von Versicherten mit hohen Kosten wird sich da-

her in späteren Prüfzyklen noch verschärfen und als Thema für die Stichprobenziehung an Relevanz zunehmen.

Tabelle 4-10 Die zehn seltensten HMGs im Klassifikationsmodell 2009, nach aufsteigender Häufigkeit geordnet*)

HMG	Bezeichnung	Pseudonymanzahl	Zuschlag
HMG102	Aphasie	2.311	747,20 €
HMG159	Pathologische Fraktur des Humerus, der Tibia oder Fibula	4.037	573,66 €
HMG158	Hüftluxation	4.043	1.439,25 €
HMG028	Leberversagen, akute Lebererkrankung	6.632	3.740,61 €
HMG107	Mukoviszidose	7.386	17.300,21 €
HMG087	Schwere angeborene Herzfehler	11.522	3.278,74 €
HMG043	Hämophilie	11.732	55.562,14 €
HMG003	Nicht virale Infektionen des Zentral- Nervensystems	14.182	2.674,60 €
HMG070	Muskeldystrophie	21.645	3.905,18 €
HMG004	Myeloische Leukämie	23.420	16.531,20 €

*) Gelb eingefärbt sind solche HMGs, die auch in der Liste der zehn teuersten auftauchen (Tabelle 4-9).

Im Folgenden untersuchen wir die Häufigkeit des Auftretens der teuren HMGs in den für die Nr. 2-Prüfung ausgewählten Beispielskassen F bis J sowie in den daraus gezogenen Stichproben der letzten Ziehung. Die seltenen HMGs, die in der Topten-Liste der teuren nicht enthalten sind, weisen im Vergleich zu diesen deutlich geringere Zuschläge auf, die sich maximal in der Größenordnung der durchschnittlichen Gesamtzuweisung der Versicherten einer Kasse bewegen, und müssen daher hier nicht gesondert untersucht werden.

Tabelle 4-11 Häufigkeit der zehn teuersten HMGs (in absteigender Zuschlagshöhe) und ausgewählte Charakteristika in den Beispielskassen F bis J

Kasse	F	G	H	I	J
Pseudonyme	2.804.204	783.593	126.110	21.742	1.197
Mittlere Zuweisung	4.003	4.723	3.096	4.114	2.894
Frauenanteil	63,0%	57,6%	52,1%	48,4%	63,0%
Mittl. Alter	59,1	65,1	45,1	59,5	43,2
HMG043	1.343	256	70	5	0
HMG130	6.342	3.636	159	73	0
HMG107	858	121	45	4	0
HMG004	2.625	730	69	17	4
HMG008	11.883	3.566	321	91	5
HMG009	7.420	2.214	166	63	2
HMG001	6.267	2.296	271	36	3
HMG006	4.357	1.169	111	33	3
HMG174	4.299	1.203	86	31	2
HMG067	13.051	4.106	343	94	1

Ein Vergleich der absoluten Häufigkeiten der teuren HMGs in den Grundgesamtheiten (Tabelle 4-11) und in den daraus gezogenen einfachen Stichproben, geplant zu Schätzung einer Fehlerquote in Höhe von $p_F=0,05$ mit einem maximalen relativen Fehler in Höhe $\varepsilon = 10\%$, vgl. Abschnitt 3.5 (Tabelle 4-12) zeigt, dass die Zahl der in den Grundgesamtheiten schwach besetzten HMGs 004, 043 und 107 (die auch in der Topten-Liste

der seltenen HMGs vertreten sind) in den Stichproben, den jeweiligen Auswahl­sätzen folgend, nur noch in verschwin­den geringer Anzahl vertreten sind. Am vollständigsten erfasst werden sie noch in der Stichprobe aus der Beispielskasse J, die so klein ist, dass die Planvorgaben zu einem Auswahl­satz von fast 86% geführt haben und die Stichprobe einer Vollerhebung nahe kommt.

Tabelle 4-12 Häufigkeit der zehn teuersten HMGs (in absteigender Zuschlagshöhe) und ausgewählte Charakteristika in den einfachen Zufallsstichproben aus den Beispielskassen F bis J (geplant für $p_F=0,05$ und $\varepsilon=10\%$)

Kasse	F	G	H	I	J
Pseudonyme	7.280	7.231	6.900	5.465	1.028
Mittlere Zuweisung	4.031	4.822	3.164	4.214	2.887
Frauenanteil	62,2%	57,7%	52,0%	47,7%	63,0%
Mittl. Alter	59,4	65,5	44,9	59,9	43,2
HMG043	1	5	13	3	0
HMG130	18	46	7	25	0
HMG107	3	2	2	0	0
HMG004	8	9	6	0	4
HMG008	26	27	11	20	5
HMG009	23	21	7	18	1
HMG001	21	22	18	7	3
HMG006	17	11	4	13	2
HMG174	10	8	6	9	2
HMG067	26	30	17	26	1

Wenn wir jetzt aber die Stichproben nach Zuweisungshöhe schichten, wie in Abschnitt 4.4 beschrieben, so zeigt eine Übersicht der Schichtgrenzen (Tabelle 4-13), dass die Versicherten mit teuren HMGs sich in den oberen Schichten 3, 4 und 5 sammeln.

Tabelle 4-13 Schichtgrenzen der fünf Schichten bei Schichtung der Stichproben aus den Beispielskassen F bis J nach Zuweisungshöhe

Schicht / Kasse	F	G	H	I	J
1	1.897 €	2.053 €	1.545 €	1.841 €	1.592 €
2	4.621 €	5.669 €	3.404 €	4.895 €	2.847 €
3	9.374 €	9.427 €	6.409 €	8.448 €	5.077 €
4	20.713 €	16.421 €	12.025 €	14.408 €	10.753 €
5	113.123 €	106.447 €	68.489 €	66.070 €	32.396 €

Da die Varianz der Zuweisungssumme mit der Nummer der Schicht steigt (vgl. Tabelle 4-4) und die schichtspezifischen Auswahl­sätze bei der Neymann-Allokation den Schichtvarianzen proportional sind, steigen die Auswahl­sätze mit wachsender Nummer der Schicht an und begünstigen so hohe Ausschöpfungsquoten in Bezug auf Versicherte mit teuren HMGs.¹⁸ In der kleinsten Kasse J sind diese für alle HMGs, die der Grundgesamtheit aufgetreten sind, sogar 100% (Tabelle 4-15). In den großen Beispielskassen mit vielen Versicherten sind die Auswahlquoten naturgemäß kleiner als in den kleinen

¹⁸ Dies gilt nicht für eine proportionale Aufteilung, weil diese mit einem über die Schichten hinweg konstanten Auswahl­satz verbunden ist.

Kassen. Sie liegen aber immer noch um Faktoren über dem Auswahlssatz der einfachen Zufallsstichprobe, wie ein Vergleich der Werte im Inneren von Tabelle 4-15 mit denjenigen zeigt, die in der letzten Zeile der Tabelle ausgewiesen sind.

Tabelle 4-14 Häufigkeit der zehn teuersten HMGs (in absteigender Zuschlagshöhe) und ausgewählte Charakteristika in den nach der Zuweisungssumme geschichteten Stichproben mit Neymann-Allokation aus den Beispielskassen F bis J (geplant für $p_F=0,05$ und $\varepsilon=10\%$)

Kasse	F	G	H	I	J
Pseudonyme	7.280	7.230	6.900	5.465	1.028
Mittlere Zuweisung	3.983	4.709	3.087	4.102	2.897
Frauenanteil	63,4%	58,5%	52,9%	48,2%	64,3%
Mittl. Alter	59,1	65,0	44,9	59,3	43,8
HMG043	47	20	60	5	0
HMG130	196	345	135	72	0
HMG107	16	9	35	4	0
HMG004	64	49	60	16	4
HMG008	135	250	210	70	5
HMG009	74	108	106	45	2
HMG001	57	93	149	25	3
HMG006	71	57	57	24	3
HMG174	69	75	62	23	2
HMG067	128	148	155	60	1

Tabelle 4-15 Ausschöpfungsquoten der zehn teuersten HMGs in den nach der Zuweisungssumme geschichteten Stichproben mit Neymann-Allokation aus den Beispielskassen F bis J (geplant für $p_F=0,05$ und $\varepsilon=10\%$)

Kasse	F	G	H	I	J
HMG043	3,5%	7,8%	85,7%	100,0%	--
HMG130	3,1%	9,5%	84,9%	98,6%	--
HMG107	1,9%	7,4%	77,8%	100,0%	--
HMG004	2,4%	6,7%	87,0%	94,1%	100,0%
HMG008	1,1%	7,0%	65,4%	76,9%	100,0%
HMG009	1,0%	4,9%	63,9%	71,4%	100,0%
HMG001	0,9%	4,1%	55,0%	69,4%	100,0%
HMG006	1,6%	4,9%	51,4%	72,7%	100,0%
HMG174	1,6%	6,2%	72,1%	74,2%	100,0%
HMG067	1,0%	3,6%	45,2%	63,8%	100,0%
Auswahlsatz der einfachen Zufallsstichp.	0,3%	0,9%	5,5%	25,1%	85,9%

Der infolge der Schichtung mit Neymann-Allokation realisierte Effekt lässt sich gut durch das Studium der relativen Abweichungen der in den stichprobenbezogenen Tabellen ausgewiesenen Charakteristika von den entsprechenden Werten in den Grundgesamtheiten (Tabelle 4-11) quantifizieren. Wie nach der Theorie nicht anders zu erwarten war, sind die relativen Abweichungen der Schätzungen des mittleren Zuweisungsbetrages aus den geschichteten Stichproben (mit Neymann-Allokation) dem Absolutbetrag nach geringer, in den Beispielskassen G bis I sogar um einen Faktor zwischen 7 und 8 geringer, als diejenigen der Schätzungen aus den einfachen Zufallsstichproben mit gleichem

Umfang (vgl. Tabelle 4-16). Der Effekt hängt wesentlich daran, dass die Zuweisung gerade das Merkmal ist, nachdem geschichtet wurde. Die anderen ausgewiesenen Charakteristika (Frauenanteil, Durchschnittsalter) werden dann auch aus der geschichteten Stichprobe in einer Mehrheit der Beispielskassen etwas ungenauer geschätzt, als aus der einfachen Zufallsstichprobe (vgl. Tabelle 4-16).

Es bleibt festzuhalten, dass an einer Schichtung mit Neymann-Allokation kein Weg vorbeiführt, wenn – was nicht unplausibel ist – sich die Zuweisungssumme als eine gute Schichtungsvariable erweist und man Schichtungseffekte zur Reduzierung des Stichprobenumfangs nutzen will.

Darüber hinaus kann man Tabelle 4-13 in Verbindung mit Tabelle 4-15 und Tabelle 4-16 (jeweils Spalte für die Kasse F im Vergleich zur Spalte G) entnehmen, dass das zweite Schema (Abbildung 4-2) zur Aggregation der zwanzig ursprünglichen Schichten auch für die großen Kassen dem ersten Schema (Abbildung 4-1) vorgezogen werden sollte.

Tabelle 4-16 Relative Abweichungen ausgewählter, aus den Stichproben geschätzter Charakteristika der Beispielskassen von den entsprechenden Werten der Grundgesamtheiten

Kasse	F	G	H	I	J
Einfache Zufallsstichproben					
Mittlere Zuweisung	0,7%	2,1%	2,2%	2,4%	-0,2%
Frauenanteil	-1,2%	0,2%	-0,2%	-1,3%	0,1%
Mittl. Alter	0,5%	0,6%	-0,3%	0,7%	0,0%
Geschichtete Zufallsstichproben mit Neyman-Allokation					
Mittlere Zuweisung	-0,5%	-0,3%	-0,3%	-0,3%	0,1%
Frauenanteil	0,7%	1,5%	1,6%	-0,5%	2,1%
Mittl. Alter	-0,1%	-0,2%	-0,5%	-0,3%	1,5%

5 Partielle Sequentialverfahren: Zweistufige Stichprobenziehung

Ein reines Sequentialverfahren, wie es in Abschnitt 2.5 exemplarisch beschrieben wurde, eignet sich für die Prüfungen nach § 42 RSAV aus organisatorischen Gründen nicht. Das Prüfverfahren sieht nämlich eine institutionelle Trennung vor zwischen den mit der Erhebung der Daten vor Ort befassten Institution (den Prüfdiensten) auf der einen und der Institution, welche die Stichprobenziehungen vornimmt, sowie Daten bereitstellt und die erhobenen Daten auswertet (dem Bundesversicherungsamt), auf der anderen Seite. Daher würde ein reines Sequentialverfahren an der Vielzahl der zwischen dem BVA und den Prüfdiensten erforderlichen Kommunikationsakte scheitern.

Das ursprünglich auf Neymann (1934) zurückgehende Konzept der Vorbereitungsstichprobe (preliminary sample), das in eine sequentielle Vorgehensweise eingebettet ist, hat den beschriebenen Nachteil nicht, da es nur eine zweite Stufe vorsieht und dann endet. Der Zweck der Vorbereitungsstichprobe liegt in der Beschaffung der erforderlichen Planungsdaten zur Festlegung des endgültigen Stichprobenumfangs.

Es gibt zweistufige Sequentialverfahren für viele Fragestellungen, insbesondere auch für die Schätzung von Anteils- und Mittelwerten.

5.1 Steins Zweistichprobenplan zur Schätzung des Mittelwertes einer Normalverteilung mit vorgegebener Präzision

Stein (1945) konnte u. a. zeigen, dass es unter Zuhilfenahme einer Vorbereitungsstichprobe gelingt, für den unbekannten Mittelwert μ einer Normalverteilung (mit ebenfalls unbekannter Varianz σ^2) ein $(1-\alpha)$ -Konfidenzintervall zu konstruieren, dessen Breite eine vorgegebene Breite von $2d$ nicht überschreitet. Die Vorbereitungsstichprobe habe den (moderaten) Umfang n_0 und möge das Stichprobenmittel $\bar{y}_0$ und die Stichprobenvarianz s_0^2 aufweisen. Nun setzte man

$$(53) \quad n = \max \left\{ n_0, \left\lceil \frac{s_0^2 \cdot t(n_0 - 1, 1 - \frac{\alpha}{2})}{d^2} \right\rceil + 1 \right\},$$

und $n_1 = n - n_0$ falls $n > n_0$ und ziehe eine weitere einfache Zufallsstichprobe vom

Umfang n .¹⁹ Mit allen n Beobachtungen berechne man nun das Stichprobenmittel $\bar{y}_n$. Dann hat das gesuchte Konfidenzintervall die Form

$$(54) \quad \text{KI}_{1-\alpha}(\bar{Y}) = \left[\bar{y}_n - \frac{s_0 \cdot t(n_0 - 1, 1 - \frac{\alpha}{2})}{d}, \bar{y}_n + \frac{s_0 \cdot t(n_0 - 1, 1 - \frac{\alpha}{2})}{d} \right]$$

(vgl. die Darstellung der Steinschen Ideen bei Chatterjee 1991).

Man beachte, dass in (54) der Mittelwert aus der Gesamtstichprobe, die Varianz aber

¹⁹ .Die eckige Klammer in (53) gibt die größte ganze Zahl an, die kleiner als der Term in der Klammer ist.

nur aus der Vorbereitungsstichprobe geschätzt wird.

Die Voraussetzung der Normalverteilung lässt sich abschwächen. So reicht es, eine beliebige, durch einen Lageparameter μ und einen Skalenparameter σ^2 parametrisierte Verteilung mit stetiger Dichte vorauszusetzen (Mukhopadhyay 1982). Aber auch eine solche Verteilung passt nicht zu dem versichertenbezogen Korrekturbetrag mit einer so großen relativen Häufigkeit auf der Null als Merkmalsausprägung.

Darüber hinaus stellt auch die Wahl von d ein Problem dar, wenn man keine Vorstellung davon hat, in welcher Größenordnung μ liegt. So kann jedes feste d bei einem hinreichend großen μ zu einer sehr kleinen relativen Abweichung ε , und umgekehrt bei einem sehr kleinen μ zu einem sehr großen ε führen.

Schließlich spricht auch gegen den Einsatz des Steinschen Zweistichprobenplanes für die Prüfungen nach § 42 RSAV, dass er für eine unendlich große Grundgesamtheit konzipiert worden ist, was bei kleinen Krankenkassen zu einem Problem werden kann, weil der resultierende Stichprobenumfang größer ist, als die Kasse Versicherte hat.

5.2 Der Zweistichprobenplan von Cox zur Schätzung eines Anteilswertes mit einem vorgegebenem Variationskoeffizienten

Da die Planung des Stichprobenumfangs, der benötigt wird, um einen Anteilswert p mit vorgegebener Präzision zu schätzen, von der Größe des Anteilswertes abhängt (s. Abschnitt 3.5), schlug Cox (1952) vor, sich mit Hilfe einer Vorbereitungsstichprobe Kenntnis über die Größe von p zu verschaffen. Dabei operationalisierte er den Begriff „Präzision“ etwas anders als es in Abschnitt 3.5 geschehen ist, nämlich durch den Variationskoeffizienten der Schätzung.

Dieser sei mit $\sqrt{c}$ vorgegeben. Sei n_0 der Umfang der Vorbereitungsstichprobe, die ohne Zurücklegen gezogen wird, und ist $\hat{p}_0$ die Schätzung von p aus der Vorbereitungsstichprobe (gemäß Formel (27)), so wird mit $\hat{q}_0 = 1 - \hat{p}_0$ der Gesamtstichprobenumfang folgendermaßen berechnet:

$$(55) \quad n = n(n_0, \hat{p}_0) = \frac{\hat{q}_0}{c\hat{p}_0} + \frac{3}{\hat{p}_0\hat{q}_0} + \frac{1}{cn_0\hat{p}_0}$$

(vgl. die Darstellung des Cox-Verfahrens bei Sinha 1991).

Ist $n \leq n_0$, so endet die Prozedur und die Schätzung erfolgt mit den Verfahren der einfachen Zufallsstichprobe (allerdings bei Cox ohne Endlichkeitskorrektur, da seine Grundgesamtheit unendlich ist). Wenn $n \geq n_0$ gilt, so wird eine weitere einfache Zufallsstichprobe ohne Zurücklegen mit dem Umfang $n_1 = n - n_0$ gezogen. Ist $\hat{p}_n$ die Schätzung von p aus der Gesamtstichprobe gemäß Formel (27), und $\hat{q}_n = 1 - \hat{p}_n$, so lautet die abschließende Schätzung von p in diesem Verfahren:

$$(56) \quad \hat{p} = \hat{p}_n - c \frac{\hat{p}_n}{\hat{q}_n},$$

wobei der zweite Term, wenn ein kleiner Anteilswert geschätzt werden soll, eine Art Korrekturterm ist, der klein gegen den ersten Term ausfällt, zumal auch c in aller Regel klein ist (z. B. 0,1). Daher besteht näherungsweise ein einfacher Zusammenhang zwischen c , dem Quadrat des Variationskoeffizienten und der in Abschnitt 3.5 verwendeten relativen Abweichung ε . Unter Vernachlässigung des Korrekturterms in (56) und der Endlichkeitskorrektur gilt nämlich

$$(57) \quad \sqrt{c} = \sqrt{\frac{q}{np}} \quad \text{sowie} \quad \varepsilon p = u_{1-\frac{\alpha}{2}} \sqrt{\frac{pq}{n}} \quad \text{bzw.} \quad \varepsilon = u_{1-\frac{\alpha}{2}} \sqrt{\frac{q}{np}}$$

Aus der ersten und der letzten Gleichung folgt, mit Rücksicht auf die beiden o. g. Vereinfachungen allerdings nur näherungsweise:

$$(58) \quad c \approx \left(\frac{\varepsilon}{u_{1-\frac{\alpha}{2}}} \right)^2$$

Die Wahl von n_0 , dem Stichprobenumfang der Vorbereitungsstichprobe unterliegt nach Cox (1952) der folgenden Restriktion:

$$(59) \quad n_0 \leq \frac{q}{cp}$$

Zur Illustration des Cox'schen Zweistichprobenverfahrens betrachten wir das Beispiel, einen Anteilswert in Höhe von $p=0,05$ (5%) mit einer relativen Genauigkeit von 30% zu schätzen (mit $\alpha=0,05$). Dann ist $\varepsilon = 0,30$ und nach Gleichung (58) gilt $c = 0,023$. Der Stichprobenumfang der Vorbereitungsstichprobe darf nach (59) den Wert von $n_0 = 811$ nicht überschreiten, und wählen wir beispielsweise $n_0 = 400$ so hängt der Gesamtstichprobenumfang von der Initialschätzung von p aus der Vorbereitungsstichprobe ab. Wir diskutieren zwei Fälle:

- p wird aus der Vorbereitungsstichprobe unterschätzt, z. B. $\hat{p}_0 = 0,046$. Dann erhalten wir einen Gesamtstichprobenumfang in Höhe von $n=958$.
- p wird aus der Vorbereitungsstichprobe überschätzt, z. B. $\hat{p}_0 = 0,054$. Dann erhalten wir einen Gesamtstichprobenumfang in Höhe von $n=706$.

Vergleichen wir diese Ergebnisse mit denjenigen, die in Tabelle 3-7 und Tabelle 3-8 ausgewiesen sind, wird deutlich, dass der Zweistichprobenplan von Cox für die hier diskutierte Problematik kaum Vorteile aufweist, es sei denn die Annahmen über p_F sind weit von der Wirklichkeit entfernt.

5.3 Anscombes zentraler Grenzwertsatz für Sequentialverfahren

Anscombe (1949, 1952 und 1953) hat sich ausführlich mit Sequentialverfahren beschäftigt und insbesondere die spezielle Situation großer Stichproben untersucht. Eines seiner vielzitierten Resultate ist die Verallgemeinerung des Zentralen Grenzwertsatzes auf Summen von n unabhängigen und identisch verteilten Zufallsvariablen, wenn das n nicht den Stichprobenumfang einer fest geplanten Zufallsstichprobe darstellt, sondern (vom Zufall abhängig) das resultierende n eines Sequentialverfahrens (vgl. Gosh 1991). Umgangssprachlich lässt sich eines der für die Anwendungen besonders relevanten Resultate von Anscombe dahingehend zusammenfassen, dass für hinreichend große resultierende Stichprobenumfänge die für eine Zufallsstichprobe mit festem Stichprobenumfang gültigen (asymptotischen) Verfahren auch auf Daten aus Sequentialverfahren angewendet werden können.

In der abschließenden Diskussion seiner o. g. Veröffentlichung von 1953 argumentiert Anscombe folgendermaßen: Die Anwendung von für festen Stichprobenumfang entwickelten Verfahren auf Daten aus Sequentialverfahren sei vollständig korrekt, wenn die bedingte Verteilung der Beobachtung, gegeben der Stichprobenumfang, sich nicht unterscheidet zwischen einem festen und dem aus einem Sequentialverfahren resultierenden zufallsabhängigen Stichprobenumfang. Diese Situation sei in der Praxis häufig gegeben (Anscombe 1953).

6 Überprüfung der Stichproben auf Qualität

Es ist in der Statistik nicht üblich, von der Qualität einer Zufallsstichprobe zu sprechen. Reine Zufallsstichproben haben per se Qualität und die erhobenen Daten vermitteln ein „repräsentatives“ Bild von der Grundgesamtheit. Natürlich kann der Stichprobenumfang zu klein sein und es treten hohe Schätzfehler auf. Dagegen versucht man sich durch Planung des Stichprobenumfangs zu schützen. Bei der Realisierung der Ziehung kann ein schlechter Zufallszahlengenerator verwendet werden, dann kann der Charakter einer reinen Zufallsstichprobe verloren gehen. Mit kleinen Wahrscheinlichkeiten kann die Stichprobe auch per Zufall atypische Häufungen in Teilen der Grundgesamtheit aufweisen.

Wenn wir im Folgenden von der Qualität der Stichprobe sprechen, so ist das in dem in diesem Sinne eingeschränkten Begriffsverständnis gemeint.

Von dem Korrekturbetrag als Erhebungsziel mal abgesehen, ist für die Grundgesamtheiten, d. h. für die Versicherten der zu prüfenden Krankenkassen, eine Vielzahl von Merkmalen im Datensatz dokumentiert. Das Bundesversicherungsamt kann die Verteilungen dieser Merkmale ermitteln und Anteils- bzw. Mittelwerte berechnen.

Daher ist die Prüfung einer Stichprobe nach § 42 RSAV auf Qualität konzeptionell einfach. Man wählt einige geeignete Anteils- und Mittelwerte aus, schätzt diese designadäquat aus der Stichprobe und vergleicht die Schätzungen mit den entsprechenden Kennziffern der Grundgesamtheit. Wenn die relativen Abweichungen im Absolutbetrag eine festzulegende Schwelle überschreiten, wird die Stichprobe abgelehnt, andernfalls akzeptiert.

Bleibt die Frage nach den geeigneten Kennziffern. Bei der Auswahl ist zu beachten, dass herangezogene Anteilswerte in der vermuteten Größenordnung des Anteils p_F der Versicherten mit Fehlern in den übermittelten Daten liegen und Merkmale, deren Mittelwerte für die Qualitätsprüfung herangezogen werden sollen, eine ähnlich schiefe Verteilung aufweisen, wie diejenige, die für den Korrekturbetrag anzunehmen ist.

Wenn wir für p_F eine Größenordnung von ca. 5% unterstellen und die Daten der Beispielskassen heranziehen, so eignet sich in dieser Hinsicht

- die Hierarchische Morbiditätsgruppe (HMG) 109 „Chronisch obstruktive Lungenerkrankung (COPD)“ für die Nr. 1- Stichproben, und
- die HMG 036 „Reflux“ für die Nr. 2-Stichproben

vergleichsweise gut, sofern man die Anteile der Versicherten betrachtet, die diesen HMGs im Jahr 2008 zugeordnet waren (vgl. Tabelle 6-1 und Tabelle 6-3). Die damit verbundene Zuweisung ist zwar für alle betroffenen Versicherten jeweils gleich, aber wenn man das Merkmal auf alle Versicherten erweitert und den nicht betroffenen Versicherten die HMG-bezogene Zuweisung 0 € zuordnet, so entsteht eine ähnlich schiefe Verteilung wie sie der Korrekturbetrag aufweist, so dass auch der zugehörige Mittelwert eine für die Qualitätsprüfung interessante Kennziffer darstellt (vgl. Tabelle 6-2 und

Tabelle 6-4). Für die Nr. 1-Stichprobe könnte, einer Anregung des GKV-Spitzenverbandes folgend, auch eine Teilgruppe der EMR-Rentner einer Altersklasse herangezogen werden, auf die ein der Fehlerquote entsprechender Anteil der Versichertenzeiten bzw. der Zuweisungen fällt.

In den nachfolgenden Tabellen sind für alle Beispielskassen auch die relativen Abweichungen der Schätzungen der betrachteten Kennziffern aus der mit der Simulation III einhergehenden Stichproben von den entsprechenden Werten der Grundgesamtheit ausgewiesen. Man sieht deutlich, dass sehr kleine Anteilswerte wie z. B. der Verstorbenenanteil oder der Anteil der Neugeborenen mit einem vergleichsweise hohen Fehler einhergehen. Das ist verständlich, denn bei der Planung des Stichprobenumfangs war ein Anteilswert in Höhe von 5% zugrunde gelegt worden (vgl. Abschnitt 3.5).

Anteilswerte, die nicht in der Größenordnung des vermuteten Anteils p_F liegen, eignen sich daher als Kennziffern im Rahmen der Überprüfung der Qualität der Stichprobe nicht. Denn wenn sie zu klein sind, werden sie zu häufig in ungerechtfertigter Weise zu einer Ablehnung der Stichprobe führen, und wenn sie zu groß sind, auch Stichproben minderer Qualität nicht entdecken können.

Der Anteil der Versicherten, die der HMG 109 zugeordnet waren, passt in der Größenordnung in den Beispielskassen der Nr. 1-Stichprobe (bei $p_F = 5\%$), und derjenige, die der HMG 036 zugeordnet waren, passt in den Beispielskassen der Nr. 2-Stichprobe. Beide Anteile werden aus der Stichprobe in vernünftiger Weise geschätzt. Die relativen Abweichungen sind moderat und das gilt auch hinsichtlich der Mittelwerte der jeweiligen erweiterten Zuweisungen (vgl. Tabelle 6-2 und Tabelle 6-4). Dies liegt in erster Linie daran, dass die Stichprobenumfänge der Simulation III-Stichproben mit der Vorgabe einer maximalen relativen Genauigkeit von $\varepsilon = 10\%$ (bei einem angenommenen $p_F = 5\%$) mit einer vergleichsweise geringen Fehlertoleranz geplant wurden.

Es lohnt sich, noch einen Blick auf die in den beiden Tabellen ebenfalls ausgewiesenen Variationskoeffizienten V zu werfen. So hat die erweiterte Zuweisung bezogen auf die HMG 109 in den Beispielskassen für die Nr. 1-Stichproben einen Variationskoeffizienten, der um die 5 herum schwankt und von 3,7 bis 7,5 reicht (s. Tabelle 6-2). Der Variationskoeffizient der erweiterten Zuweisung mit Bezug auf die HMG 036 liegt in den Beispielskassen für die Nr. 2-Stichproben nahe bei 5 und reicht von 4,3 bis 5,6 (s. Tabelle 6-4). Auch vor diesem Hintergrund dürfte es sich empfehlen, für den versichertenbezogenen Korrekturbetrag eher von der Annahme $V \geq 5$, als von der Annahme $V < 5$ auszugehen.

Man beachte die gute Übereinstimmung der nach Formel (16) berechneten Variationskoeffizienten mit den in Tabelle 6-2 und Tabelle 6-4 bzw. ausgewiesenen, wenn man die in den analogen Zellen von Tabelle 6-1 bzw. Tabelle 6-3 dokumentierten Werte für p_F einsetzt.

Tabelle 6-1 Ausgewählte Anteilswerte in den Nr. 1-Grundgesamtheiten der Beispielskassen und relative Abweichungen bei Schätzung aus der Simulation III-Stichprobe*)

Kasse	Anteil				
	Verstorbener	Neugeborener	der V. mit EMR	der V. mit HMG 036	der V.mit HMG 109
A	0,97%	0,73%	2,48%	1,92%	4,49%
B	2,33%	0,68%	5,38%	2,83%	7,05%
C	0,23%	1,16%	1,12%	0,94%	2,96%
D	1,19%	0,63%	1,71%	2,20%	5,67%
E	0,08%	1,79%	0,33%	0,64%	1,78%
GKV	1,05%	0,81%	2,50%	1,74%	4,51%
Kasse	Relative Abweichung der Stichprobenanteilswerte (Simulation III-Stichprobe)				
A	-10,6%	33,3%	-7,0%	-17,1%	5,3%
B	-6,0%	-27,4%	6,9%	-1,6%	-4,0%
C	34,8%	-13,0%	4,2%	3,3%	0,7%
D	-5,2%	-1,2%	2,9%	-6,9%	2,4%
E	12,7%	7,8%	-1,6%	-3,8%	-2,8%

*) Zeitlicher Bezug: Ausgleichsjahr; Rote Einfärbung: Relative Abweichung von (absolut) > 10%

Tabelle 6-2 Erweiterte Zuweisungen für zwei ausgewählte HMGs in den Nr. 1-Grundgesamtheiten der Beispielskassen und relative Abweichungen bei Schätzung aus der Simulation III-Stichprobe*)

Kasse	Erweiterte Zuweisung bezogen auf HMG 036			Erweiterte Zuweisung bezogen auf HMG109		
	Mittelwert	Std.abw.	Var.koeff.	Mittelwert	Std.abw.	Var.koeff.
A	14,05 €	101,31 €	7,2	44,13 €	205,48 €	4,7
B	20,47 €	121,54 €	5,9	68,28 €	251,75 €	3,7
C	6,77 €	70,39 €	10,4	28,55 €	165,98 €	5,8
D	16,05 €	108,12 €	6,7	55,45 €	228,83 €	4,1
E	4,70 €	58,97 €	12,5	17,73 €	132,30 €	7,5
GKV	12,69 €	96,25 €	7,6	43,91 €	204,72 €	4,7
Kasse	Relative Abweichung der Stichprobenwerte (Simulation III-Stichprobe)					
A	-16,9%	-8,8%	9,8%	5,3%	2,5%	9,8%
B	-4,1%	-2,4%	1,7%	-4,0%	-1,4%	1,7%
C	1,6%	0,3%	-1,3%	0,7%	-0,3%	-1,3%
D	-5,7%	-2,7%	3,2%	2,4%	1,1%	3,2%
E	-1,8%	-0,5%	1,4%	-2,8%	-1,6%	1,4%

*) Zeitlicher Bezug: Vorjahr; Rote Einfärbung: Relative Abweichung von (absolut) > 10%; Erweiterte Zuweisung: Mittelwert und Standardabweichung wurden berechnet unter Einschluss der Versicherten, welche der HMG im Vorjahr nicht zugeordnet waren.

Tabelle 6-3 Ausgewählte Anteilswerte in den Nr. 2-Grundgesamtheiten der Beispielskassen und relative Abweichungen bei Schätzung aus der Simulation III-Stichprobe*)

	Anteil				
Kasse	Verstorbener	Neugeborener	der V. mit EMR	der V. mit HMG 036	der V. mit HMG 109
F	2,23%	0,12%	6,08%	4,82%	11,32%
G	4,08%	0,10%	10,18%	5,37%	13,44%
H	0,76%	0,33%	4,30%	4,09%	12,92%
I	2,64%	0,13%	4,01%	5,38%	13,90%
J	0,58%	0,33%	1,50%	3,09%	10,36%
GKV	2,55%	0,15%	6,43%	4,65%	12,02%
Kasse	Relative Abweichung der Stichprobenanteilswerte (Simulation III-Stichprobe)				
F	-1,3%	-33,5%	-3,6%	-2,1%	-3,0%
G	-2,1%	-43,3%	-6,7%	0,0%	2,1%
H	6,7%	6,2%	-4,7%	5,2%	6,8%
I	2,1%	13,7%	-0,5%	7,9%	-3,2%
J	16,4%	16,4%	3,5%	3,9%	-3,3%

*) Zeitlicher Bezug: Vorjahr; Rote Einfärbung: Relative Abweichung von (absolut) > 10%

Tabelle 6-4 Erweiterte Zuweisungen für zwei ausgewählte HMGs in den Nr. 2-Grundgesamtheiten der Beispielskassen und relative Abweichungen bei Schätzung aus der Simulation III-Stichprobe*)

	Erweiterte Zuweisung bezogen auf HMG 036			Erweiterte Zuweisung bezogen auf HMG 109		
Kasse	Mittelwert	Std.abw.	Var.koeff.	Mittelwert	Std.abw.	Var.koeff.
F	35,33 €	158,28 €	4,5	111,14 €	314,48 €	2,8
G	38,88 €	165,35 €	4,3	130,10 €	335,74 €	2,6
H	29,43 €	144,49 €	4,9	124,32 €	328,66 €	2,6
I	39,38 €	166,66 €	4,2	136,16 €	343,00 €	2,5
J	22,73 €	127,96 €	5,6	102,67 €	304,13 €	3,0
GKV	33,85 €	154,89 €	4,6	117,14 €	321,28 €	2,7
Kasse	Relative Abweichung der Stichprobenwerte (Simulation III-Stichprobe)					
F	-2,3%	-1,1%	1,2%	-2,7%	-1,1%	1,6%
G	-0,4%	-0,3%	0,1%	2,0%	0,9%	-1,1%
H	3,8%	1,7%	-2,0%	6,7%	2,8%	-3,6%
I	8,2%	3,7%	-4,1%	-3,2%	-1,4%	1,9%
J	3,6%	1,7%	-1,9%	-3,7%	-1,7%	2,0%

*) Zeitlicher Bezug: Vorjahr; Erweiterte Zuweisung: Mittelwert und Standardabweichung wurden berechnet unter Einschluss der Versicherten, welche der HMG im Vorjahr nicht zugeordnet waren.

7 Resümee und Verfahrensvorschlag

7.1 Planungsrelevante Vorkenntnisse

Im Rahmen des Anhörungsverfahrens wurden dem Gutachter einerseits Ergebnisse der Prüfungen nach § 15a RSAV und andererseits Daten zu den Prüfkapazitäten der Prüfdienste bekannt gemacht.

Nach einer Zusammenstellung vom 25. März 2010 liegen in der Dokumentation der Ergebnisse der Prüfungen nach § 15a RSAV für das Jahr 2006 Daten für alle geprüften Krankenkassen vor (BVA 2013a). Diese sind gegliedert nach Versichertenzeiten insgesamt (nach Alter und Geschlecht) und den Versichertenzeiten mit DMP-Einschreibung. Versichertenzeiten mit Erwerbsminderungsrente bzw. mit Krankengeld-Anspruch (beide prüfungsrelevant nach § 42 Abs. 1 S. 1 Nr. 1 RSAV) wurden in den Prüfungen nach § 15a RSAV nicht überprüft. Nach Einschätzung des Leiters des bundesweiten Arbeitskreises „Versichertenzeiten Prüfung“ der Prüfdienste sind jedoch spürbare Änderungen der Fehlerquote infolge dieser Erweiterungen nicht zu erwarten (Kirsch 2013).

Die Ergebnisdokumentation unterscheidet Fehlerquoten in Bezug auf die überprüften Fälle und solche in Bezug auf die gemeldeten Versichertenzeiten. Für das Jahr 2006 wurden bei der Prüfung der Versichertenzeiten insgesamt 106.725 Fälle und 37.585.684 Versichertentage geprüft. Dabei ließ sich fall- (d. h. versicherten-)bezogen eine Fehlerquote von rund 0,81% ermitteln. Was die Versichertenzeiten mit DMP-Einschreibung betrifft, so wurden 5.147 Fälle und 1.198.079 zugehörige Versichertentage mit einer fallbezogenen Fehlerquote von rund 0,70% dokumentiert.

Eine Aufstellung der Personalressourcen und Prüfkapazitäten der Prüfdienste des Bundes und der Länder insgesamt (Polaszek 2013) weist bundesweit ca. 45 Prüfer nach, die jährlich für Prüfungen der Daten des RSA zur Verfügung stehen (nach Berechnungen aus dem Jahr 2003). Bei einer Anzahl von (im Mittel) 180 Prüftagen könnten im Durchschnitt 8.100 Arbeitstage für RSA-Prüfungen pro Jahr aufgewendet werden. Für die weiteren Berechnungen wurden für die Nr. 1-Prüfungen nach den Erfahrungen aus den § 15a RSAV-Prüfungen keine maschinellen Vorprüfungen vorgesehen, aber für die Nr. 2-Prüfungen unterstellt, dass 80% der Fälle in einer maschinellen Vorprüfung abschließend behandelt und nur 20% manuell-zerebral abgeklärt werden müssen.

Mit diesen Annahmen ergibt sich für unterstellte 130 zu prüfende Krankenkassen eine Prüfkapazität von etwas mehr als 1.600 Versicherten pro Jahr und Krankenkasse für die Nr. 1- und von rund 1.500 Versicherten für die Nr. 2-Prüfung.

7.2 Verfahrensimmanente Kernprobleme der Prüfungen nach § 42 RSAV

Ein entscheidendes Problem bei der Planung der Stichprobenziehung und –hochrechnung für den ersten Zyklus des Prüfverfahrens nach § 42 RSAV besteht darin, dass in Bezug auf die neuen, morbiditätsbezogenen Nr. 2-Prüfungen keine Vorkenntnisse über Häufigkeit und Art der Fehler in den Datenmeldungen an das BVA vorliegen.. Die Planung der Anlage der Stichprobe und des Stichprobenumfangs ist aber wesentlich auf

zumindest rudimentäre Vorkenntnisse angewiesen. Beruht die Planung ausschließlich auf Annahmen, so läuft man Gefahr, entweder Ressourcen zu verschwenden, weil die Stichprobe zu groß, oder man handelt sich enorme Fehlerbreiten ein, weil sie zu klein angelegt wurde. Es sind daher, was die Nr. 2-Prüfungen angeht, Vorbereitungsstichproben zur Gewinnung von Planungsgrundlagen erforderlich.

Vor diesem Hintergrund wird vorgeschlagen, sich vor Beginn des ersten Prüfzyklus im Rahmen einer vorgeschalteten Pilotphase durch Untersuchung der Verhältnisse in einigen gut ausgesuchten Beispielskassen auf der Basis je einer einfachen Zufallsstichprobe mit festem Stichprobenumfang die erforderlichen Planungsdaten zu beschaffen. Dieses Konzept wird im übernächsten Abschnitt ausführlich dargestellt.

Auf das zweite Kernproblem führt ein Satz aus den vom Bundesversicherungsamt formulierten inhaltlichen Anforderungen an das Gutachten: „Das Stichprobenverfahren soll den Stichprobenfehler so gering wie möglich halten. Es darf dabei aber die Gesamtkapazität der Prüfdienste nicht überschritten werden.“ (Bundesversicherungsamt 2013).

Infolge der extremen Schiefe der Verteilung des versichertenbezogenen Korrekturbetrages ist mit einem hohen Variationskoeffizienten V , und in der Folge mit hohen Stichprobenumfängen zu rechnen. Wie darüber hinaus die Beziehung (26) lehrt, muss der für eine gewünschte Genauigkeit der Schätzung des Korrekturbetrages erforderliche Stichprobenumfang um so größer gewählt werden, je kleiner die Fehlerquote, d. h. der Anteil der Versicherten mit Fehlern in den an das BVA übermittelten Daten ist.

In den § 15a RSAV-Prüfungen wurde eine sehr kleine Fehlerquote von unter 1% gefunden, die auf massive potenzielle Probleme bei der Lösung der Aufgabe, den Korrekturbetrag im Prüfverfahren nach § 42 Abs. 1 S. 1 Nr. 1 RSAV zu schätzen, schließen lässt. Wie hoch die Fehlerquote in den Nr. 2-Prüfungen ausfällt, wird aus den vorgeschlagenen Vorbereitungsstichproben hervorgehen. Dann könnte sich erweisen, dass die zitierte Forderung der Quadratur des Kreises gleicht, weil die Summe der Stichprobenumfänge die Gesamtkapazität der Prüfdienste für die Prüfungen nach § 42 RSAV übersteigt.

7.3 Zweistufiges Prüfverfahren mit Schwellenwert für die Fehlerquote

In § 42 Abs. 5 RSAV wird dem BVA die Möglichkeit eingeräumt, im Benehmen mit dem Spitzenverband Bund der Krankenkassen zu bestimmen, dass die Hochrechnung des Korrekturbetrages nur erfolgt, wenn die Fehlerquote eine vorgegebene Schwelle überschreitet.

Mit dieser Regelung wird die Verhältnismäßigkeit des Verfahrens in den Blick genommen. Sie ist auch sehr hilfreich, wenn es darum geht, den Stichprobenumfang zu reduzieren. Wie ein Vergleich von Tabelle 3-5 mit Tabelle 3-7 und Tabelle 3-6 mit Tabelle 3-8 zeigt, werden selbst bei gleichen Anforderungen für die Genauigkeit (gemessen an dem höchsten zugelassenen Fehler ε) für die Hochrechnung des Korrekturbetrages bei einer gegebenen Fehlerquote deutlich höhere Stichprobenumfänge erforderlich, als bei

Schätzung eben dieser Fehlerquote aus der Stichprobe. Dabei ist der zusätzliche Parameter q (der Variationskoeffizient des Korrekturbetrages in der Gruppe VG_F der Versicherten mit Fehlern in den an das BVA übermittelten Daten) mit 0,6 in Tabelle 3-5 und Tabelle 3-6 äußerst moderat angenommen worden. Wenn sich empirisch erweist, dass q nahe an Eins oder sogar größer als Eins ist, verschärft sich die Situation im Hinblick auf die Hochrechnung der Korrekturbeträge noch weiter.

Darüber hinaus wird man an die Genauigkeit der Hochrechnung des Korrekturbetrages höhere Anforderungen stellen, als an die Genauigkeit der Schätzung der Fehlerquote. So wurde beispielweise im Rahmen der § 15a RSAV-Prüfungen der Stichprobenumfang zur Schätzung einer Fehlerquote von 0,02 mit einem relativen Fehler von $\varepsilon=50\%$ geplant (Fischer 2013). Eine solche Fehlermarge für die Hochrechnung des Korrekturbetrages dürfte von den betroffenen Krankenkassen als inakzeptabel angesehen werden.

Es erscheint daher als äußerst zweckmäßig, die Prüfungen nach § 42 RSAV von vornherein zweistufig anzulegen. Im Folgenden wird ein solches Verfahren für eine bestimmte zu prüfende Krankenkasse beschrieben.

Stufe 1: Es wird eine Stichprobe zur Schätzung der Fehlerquote geplant, gezogen und die Daten der Stichprobenversicherten werden geprüft. Der Planwert der Fehlerquote, der für die Stichprobenumfangsbestimmung gemäß Formel (31) benötigt wird, wird kassenspezifisch aus dem vorangegangenen Prüfzyklus abgeleitet. Für den ersten Prüfzyklus wird der Planwert im Fall der Nr. 1-Prüfungen den Ergebnissen der § 15a RSAV-Prüfungen entnommen (wenn möglich kassenspezifisch²⁰, andernfalls als GKV-weiter Durchschnittswert). Der Planwert für die Nr. 2-Prüfungen wird, bezogen auf den ersten Prüfzyklus, der Pilotphase entnommen (vgl. hierzu den nächsten Abschnitt). Dafür gibt es zwei Konzepte. Bei Verfolgung des ersten wird die zu prüfende Krankenkasse derjenigen unter den Beispielskassen des Pilotverfahrens zugeordnet, zu der sie unter Zugrundelegung ausgewählter Kriterien „am besten“ passt.²¹ Falls es gegenüber dem Zuordnungskonzept Akzeptanzhemmnisse gibt, könnte man den Durchschnitt der Fehlerquote aus allen Beispielskassen heranziehen (zweites Konzept). Der zugelassene relative Fehler für die erste Stufe sollte im Hinblick auf die knappen Prüfkapazitäten nicht zu knapp bemessen werden (z. B. $\varepsilon = 30\%$ oder sogar $\varepsilon = 40\%$).

Stufe 2: Wenn die Prüfung der Daten aus der Stufe 1 eine Schätzung der Fehlerquote ergibt, der oberhalb des vorher festgelegten Schwellenwertes liegt (z. B. oberhalb von 2%), so wird mit dem ermittelten Wert der Fehlerquote (und einem q z. B. in einer Höhe von 0,6) der Stichprobenumfang unter Verwendung der Formel (26) für die Schätzung des mittleren Korrekturbetrages mit einem kleineren relativen Fehler als in der ersten

²⁰ Für ggf. inzwischen durch Fusionen entstandene neue Krankenkassen müssen die einzelnen Fehlerquoten der fusionierten Kassen mit den jeweiligen Versichertenzahlen gewichtet gemittelt werden.

²¹ Bezogen auf ausgewählte Versichertenmerkmale lässt sich der Abstand zwischen zwei Krankenkassen mit Hilfe des Kolmogorov-Abstandes für Verteilungen sogar messen, Die Merkmale müssen hierzu nur in gleicher Weise klassifiziert werden. Die optimal passende Beispielskasse könnte daher als die „nächstliegende“ operationalisiert werden.

Stufe geplant (z. B. $\varepsilon = 20\%$). Sodann wird die Stichprobe der ersten Stufe um so viele Stichprobenversicherte ergänzt, dass der für die Gesamtziehung geplante Stichprobenumfang erreicht wird.²² Die Aufstockungsziehung erfolgt ggf. so, dass die Gesamtstichprobe geschichtet ist mit Neymann-Allokation. Die Schätzung des mittleren Korrekturbetrages erfolgt trotz des sequentiellen Verfahrens designadäquat nach den im vorliegenden Gutachten zusammengestellten Formeln. Die Rechtfertigung dafür wird den Grenzwertsätzen und Empfehlungen von Anscombe (1953) entnommen, da auch moderate Stichprobenumfänge für die Prüfungen nach § 42 RSAV im Sinne asymptotischer Betrachtungen der theoretischen Statistik als groß gewertet werden können.²³

7.4 Verfahrensvorschlag zur Vorbereitung der erstmaligen Prüfung der Krankenkassen („Pilotphase“)

7.4.1 Nr. 1-Stichproben: Auswertung der Ergebnisse der Prüfungen nach § 15a RSAV

Das BVA sollte die Ergebnisse der Prüfungen nach § 15a RSAV hinsichtlich der Fehlerquoten und (wenn möglich) der Korrekturbeträge und potentieller Schichtungsgewinne sowie geeigneter Variablen und Versichertengruppen für die Überprüfung der Qualität der Stichproben (gegliedert nach allgemeinen Versichertenzeiten und Versichertenzeiten mit DMP) analysieren, um auf diese Weise Planungsdaten für die Planung des ersten Prüfzyklus zu gewinnen.

7.4.2 Nr. 2-Stichproben: Vorbereitungsstichproben im Rahmen simulierter Prüfverfahren für ausgewählte Beispielskassen

Vor der erstmaligen Realisierung der Prüfungen nach § 42 Abs.1 Nr. 2 RSAV wird eine Vorstufe vorgeschlagen, die in folgender Weise ablaufen soll:

1. Das Bundesversicherungsamt wählt (ähnlich, wie schon zur Unterstützung des vorliegenden Gutachtens) eine Anzahl von nach Morbidität, Größe, Region und Kassenart unterschiedlicher Beispielskassen aus, die in ihrer Gesamtheit das Spektrum der im Korrekturverfahren für das Ausgleichsjahr 2009 am Morbi-RSA beteiligten Krankenkassen der GKV gut abbilden. Im Rahmen des Anhörungsverfahrens wurde hierzu der folgende Vorschlag diskutiert: Es werden acht Krankenkassen in das Pilotverfahren einbezogen, drei BKKn, zwei EKKn, zwei AOKn (Ost und West) und eine IKK (in Ost und West tätig).
2. Für die ausgewählten Kassen wird unter Berücksichtigung der Größe der Grundgesamtheiten - je ein Stichprobenumfang gemäß Formel (31) mit $p_F = 0,05$ und $\varepsilon =$

²² Falls in den Auswertungen der Daten vorangegangener Prüfzyklen bzw. (für den ersten Prüfzyklus) der § 15a RSAV-Prüfungen (Nr. 1-Stichproben) oder der Pilotphase (Nr. 2-Stichproben) geeignete Schichtvariable identifiziert und die zugehörigen Schichtgewinne quantifiziert werden konnten, wird der errechnete Gesamtstichprobenumfang entsprechend der potentiellen Schichtgewinne reduziert.

²³ So strebt beispielsweise die $t(n)$ -Verteilung für $n \rightarrow \infty$ gegen die Standardnormalverteilung, wobei $n=100$ im Sinne der Asymptotik insofern schon eine große Zahl ist, als die Dichten beider Verteilungen ab $n=100$ mit bloßem Auge kaum noch voneinander unterschieden werden können.

30% berechnet und die Stichproben jeweils als einfache Zufallsstichproben gezogen.

3. Das BVA übermittelt, wie vorgesehen, den Prüfdiensten die gezogenen Pseudonyme der Stichprobenversicherten einschließlich der zu prüfenden Angaben. Die Prüfdienste überprüfen bei den Beispielskassen, ob für diese Pseudonyme die vorliegenden Angaben in den Satzarten belegt sind und melden dem BVA das Ergebnis der Überprüfungen in gegliederter Form zurück. Insbesondere sollte in dieser Rückmeldung versichertenbezogen dokumentiert werden, bei welcher Satz- bzw. Belegart Fehler aufgetreten sind.
4. Das BVA gruppiert die Versicherten dann neu auf Basis der korrigierten Angaben und berechnet für jede Beispielskasse das jeweilige p_F sowie für jeden Stichprobenversicherten den Betrag, der sich aufgrund der Korrekturen ergibt.
5. Ein Verwaltungsvollzug unterbleibt, d. h. es findet keine Einforderung des auf Kassenebene hochgerechneten Korrekturbetrages statt.
6. Bei Veröffentlichung der Ergebnisse aus der Vorstufe bleiben die Beispielskassen anonym und es werden für sie im ersten Prüfzyklus neue Stichproben gezogen.

7.5 Planung und Durchführung des ersten Prüfzyklus

Nach Abschluss der Pilotphase verfügt das BVA über die empirischen Daten für eine sorgfältige Stichprobenplanung. Die Auswertungen sollten dabei möglichst für jede der beteiligten Beispielskassen und jeweils die Nr. 1- und die Nr. 2-Grundgesamtheiten folgende Punkte umfassen:

- Statistische Analysen zur Identifizierung geeigneter Schichtungsvariable und des maximal möglichen Schichtungseffektes. Hierzu werden Korrelationen zwischen dem versichertenbezogenen Korrekturbetrag²⁴ und potentiell geeigneten Schichtungsvariablen (z. B. dem Zuweisungsbetrag) in der jeweiligen Stichprobe berechnet.²⁵ Dieser Auswertungsschritt hat explorativen Charakter. Wenn eine vielversprechende Schichtungsvariable gefunden wurde, ist eine nachträgliche Schichtung durchzuführen, um den zugehörigen Schichtungseffekt nach Maßgabe von Formel (51) zu berechnen.
- Gegliederte Auswertung des in den Stichproben gefundenen Fehleranteils p_F und des versichertenbezogenen Korrekturbetrages nach Satz- bzw. Belegart sowie in den fünf wie im vorliegenden Gutachten gebildeten Schichten der Zuweisungssumme.

²⁴ Für Versicherte ohne Korrektur in den übermittelten Datensätzen wird die Merkmalsausprägung 0 € vergeben.

²⁵ Der Stichprobenkorrelationskoeffizient stellt unter milden Voraussetzungen eine konsistente Schätzung des Korrelationskoeffizienten in der Grundgesamtheit dar (vgl. z. B. Degen und Lorscheid 2002, Kapitel 19).

- Analyse der Verteilung des versichertenbezogenen Korrekturbetrages (Mittelwert, Standardabweichung, Variationskoeffizient, Variationskoeffizient q innerhalb der Versichertengruppe VG_F , Histogramm)
- Ermittlung geeigneter HMGs oder Versichertengruppen zur Überprüfung der Qualität der Stichprobe gemäß den Ausführungen des Abschnitts 6 unter Zugrundelegung des gefundenen Fehleranteils p_F und des Variationskoeffizienten V des Korrekturbetrages.

Es ist möglich, dass die statistischen Analysen aus der Pilotphase zu Ergebnissen führen, die zwischen den ausgewählten Beispielskassen variieren. Die Erfahrungen mit Auswertungen der Daten der gesetzlichen Krankenversicherung haben aber gezeigt, dass es eine hohe strukturelle Stabilität (regional, temporal und über verschiedenen Kassenarten hinweg) von vielen Verhältniszahlen und anderen komplexen Kennziffern gibt. Von daher besteht Hoffnung, dass die Variabilität der Ergebnisse der Datenauswertung über die Beispielskassen hinweg nicht allzu groß ausfallen wird. Eher sind Unterschiede zwischen den Nr. 1- und Nr. 2-Stichproben zu erwarten.

Wenn sich aus den Daten der Pilotphase keine schwerwiegenden Argumente dagegen ableiten lassen, sollte die Schichtung der Nr. 1- bzw. Nr. 2-Grundgesamtheiten kassenübergreifend jeweils in gleicher Weise erfolgen (und maximal sechs Schichten vorsehen), um Verfahrensunterschiede zwischen den Krankenkassen zu vermeiden.

Wenn die Schichten festgelegt sind, kann das BVA nun die konkrete Stichprobenplanung, -ziehung und -hochrechnung für jede der im Ausgleichsjahr 2009 am RSA beteiligten Krankenkassen im Rahmen des in Abschnitt 7.3 beschriebenen, zweistufigen Verfahrens vornehmen. Auch hier sind mehrere Schritte erforderlich:

1. Einteilung der Nr. 1.- und Nr. 2 Grundgesamtheiten für jede Kasse nach dem entwickelten Schichtungskonzept für die Stichprobenziehung der Stufe 2.
2. Festlegung der Schwellenwerte für die Nr. 1- und die Nr. 2- Stichproben und der Fehlertoleranzgrenzen für die beiden Stufen des Verfahrens in Abstimmung mit den anderen am Verfahren Beteiligten (Krankenkassen, Prüfdienste).
3. Berechnung der Stichprobenumfänge für die Stufe 1 (beide Grundgesamtheiten) nach Maßgabe der Formel (31) unter Zugrundelegung der aus dem Pilotverfahren geschätzten Fehlerquoten. Ziehung der Stichprobe für die Stufe 1 als einfache Zufallsstichprobe seitens des Bundesversicherungsamtes und Übermittlung der Daten der Stichprobenversicherung an den zuständigen Prüfdienst.
4. Der Prüfdienst prüft die versichertenbezogenen Datensätze und das BVA berechnet aus dem Ergebnis die Fehlerquote. Ist diese nicht größer als der vorgegebene Schwellenwert, endet die Prüfung und es gibt keine zweite Stufe. Insbesondere wird der Korrekturbetrag nicht hochgerechnet, aber vom BVA zwecks fortzuschreibender Planung dokumentiert. Liegt die berechnete Fehlerquote oberhalb des Schwellenwertes geht das Verfahren in die Stufe 2.

5. Nach Eintritt in die zweite Stufe erfolgt die Berechnung des Gesamtstichprobenumfangs (für beide Stufen zusammen) nach Maßgabe der Formel (26) unter Verwendung der abschließend ermittelten Fehlerquote und dem vom BVA aus den Daten der ersten Stufe ermittelten Variationskoeffizienten q für die Versicherten aus VG_F .
6. Reduzierung des berechneten Stichprobenumfangs nach Maßgabe der in der Pilotphase gefundenen Schichtungseffekte.
7. Ziehung der Ergänzungsstichprobe so, dass die Gesamtstichprobe ggf. eine geschichtete Stichprobe mit Neymann-Allokation darstellt²⁶ und Überprüfung der Qualität der Gesamtstichprobe gemäß den Ausführungen des Abschnitts 6 und in Bezug auf die aus den Daten der Pilotphase ermittelten HMGs und Versichertengruppen. Ggf. ist die Stichprobenziehung für die Gesamtstichprobe (nunmehr ohne Stufen) zu wiederholen.
8. Falls es sich organisatorisch einrichten lässt, alle Stufe 1-Prüfungen abzuschließen, bevor die erste Stufe 2-Prüfung beginnt, empfiehlt es sich, die Zahl der in der Stufe 2 insgesamt zu überprüfenden Datensätze über alle Kassen zu summieren und mit der Restkapazität der Prüfdienste abzugleichen. Sollte diese überschritten sein, muss entweder die Fehlertoleranzgröße ε für die zweite Stufe vergrößert oder das Prüfverfahren unter Berücksichtigung der Ergebnisse der Pilotphase, was Fehlerart und -häufigkeit betrifft, in den Grundgesamtheiten oder im zu prüfenden Datensatz eingeschränkt werden.
9. Übermittlung der in der Ergänzungsstichprobe gezogenen Pseudonyme an die Prüfdienste. Durchführung der Prüfungen (gilt für beide Grundgesamtheiten).
10. Übernahme der Ergebnisse der Prüfung und designadäquate Hochrechnung der Summe der Korrekturbeträge je Kasse auf die Nr. 1- und Nr. 2-Grundgesamtheiten nach Maßgabe der Formeln in den Abschnitten 3 und 4 und Einforderung der auf Kassenebene hochgerechneten Korrekturbeträge von den Krankenkassen.

Sollte keine geeignete Schichtungsvariable gefunden werden, entfallen die mit einer Schichtung verbundenen Verfahrensschritte und es werden auch auf der zweiten Stufe einfache Zufallstichproben gezogen.

Die Prüfungen nach Nr. 2 sind unabhängig von denen nach Nr. 1 und können vom Prüfdienst für jede konkrete zu prüfende Kasse, unbeschadet einer möglichen Koppelung

²⁶ Die Ergänzung der Stichprobenversicherten der ersten Stufe, die aus einer einfachen Zufallsstichprobe gezogen wurden, in einer Art und Weise, dass die Gesamtstichprobe zu einer geschichteten Stichprobe mit Neymann-Allokation wird, ist nicht ganz einfach. Das zweistufige Verfahren ist mit einer prospektiven Schichtung wenig gut verträglich und im schlimmsten Fall können bei sehr kleinen Kassen einige Datensätze aus der ersten Stufe in die Gesamtstichprobe nicht übernommen werden. Wenn das Prüfverfahren, abhängig vom Ergebnis der ersten, in die zweite Stufe eintritt, berechnet das BVA mit Input-Daten aus der zweiten Stufe die schichtspezifischen Stichprobenumfänge einer geschichteten Stichprobe mit Neymann-Allokation. In der Regel sind diese größer, als die schichtspezifischen Besetzungszahlen aus der einfachen Zufallsstichprobe (mit wesentlich kleinerem Umfang) der ersten Stufe und können schichtspezifisch durch Ziehen ohne Zurücklegen ergänzt werden. Sind sie kleiner (nur bei sehr kleinen Kassen) müssen die Überhänge per Zufall herausgenommen werden.

der Arbeitsschritte beim BVA, grundsätzlich in einem zeitlichen Abstand von diesen vorgenommen werden.

7.6 Laufende Überprüfung und Weiterentwicklung der Verfahren in nachfolgenden Prüfzyklen

Für die Überprüfung und Weiterentwicklung des Verfahrens kommen drei verschiedene Ansätze infrage, die parallel angewendet werden können.

a) Verbesserung der Planungsgrundlagen

Die Stichprobenverfahren des ersten Prüfzyklus basieren, wenn die Pilotphase – wie in Abschnitt 7.4 beschrieben – realisiert wird, auf den Ergebnissen der Auswertungen der in der Pilotphase erhobenen Daten, wie in Abschnitt 7.5 dargestellt. Wenn der erste und weitere Prüfzyklen abgeschlossen worden sind, stehen die im Rahmen dieser Prüfungen erhobenen Daten dem Bundesversicherungsamt ebenfalls zur Verfügung und bilden eine wesentlich breitere empirische Basis, die auch für die Nr. 2-Stichproben alle Kassen (und nicht nur die acht in der Vorstufe geprüften Beispielskassen) umfasst. Daher ist es im Sinne einer kontinuierlichen Fortschreibung und Verbesserung der Planungsgrundlagen zweckmäßig, die im ersten Teil des Abschnitts 7.5 beschriebenen Analysen unter Bezug auf den weit größeren Datenbestand zu wiederholen und dabei u. a. die Stabilität der Ergebnisse zu untersuchen. Die fortgeschriebenen Planungsdaten hinsichtlich des Fehleranteils, der Verteilung des Korrekturbetrages, des Schichtungskonzeptes und die für Zwecke der Prüfung der Stichprobenqualität auszuwählenden HMGs werden dann für den nächst folgenden Prüfzyklus herangezogen.

b) Untersuchungen zu Möglichkeiten der Prädiktion des Fehleranteils und des Korrekturbetrages aus Versichertenmerkmalen

Logistische und normale Regressionsanalysen sind in den statistischen Programmpaketen wie SAS, SPSS oder R auch in sog. Survey-Versionen verfügbar und in diesen Versionen geeignet, auch Daten aus komplexen Stichproben auszuwerten. Zur Weiterentwicklung der Stichprobenverfahren nach § 42 RSAV dürfte es vor diesem Hintergrund sinnvoll sein, basierend auf den Daten einiger gut ausgewählter Kassen (jeweils separat) Prädiktionsmodelle für den Fehleranteil auf der einen und den Korrekturbetrag auf der anderen Seite zu suchen, zu eichen und die mit den Modellen vorhergesagten Werte mit den aus der Stichprobe hochgerechneten zu vergleichen.

c) Untersuchungen zur Variation des Fehleranteils und des kassenspezifischen Anteils der Summe der Korrekturbeträge an der ursprünglichen Gesamtzuweisung

Die Variation der auf Basis der Prüfergebnisse ermittelten kassenspezifischen Werte – regional und temporal sowie auch nach Kassenart und Kassengröße – sollte in regelmäßigen Abständen untersucht werden, um sicher zu gehen, dass das Verfahren nicht mit unplausiblen Unterschieden einhergeht.

8 Anhang

8.1 Beispielskassen für die Stichprobenplanung nach § 42 Abs. 1 S. 1 Nr. 1 RSAV

Tabelle 8-1 Beispielskasse A nach ausgewählten Kennziffern

Schicht	Pseudonyme		Alter		Zahl der HMGs		Zuweisung	
	Anzahl	Anteil	Mittel	StdAbw.	Mittel	StdAbw.	Mittel	StdAbw.
1	1.250.241	35,2%	23,2	14,1	0,05	0,43	628 €	186 €
2	867.468	35,2%	21,6	14,3	0,01	0,19	905 €	38 €
3	734.234	75,9%	37,0	15,4	0,03	0,26	1.069 €	59 €
4	634.937	90,8%	40,6	11,6	0,04	0,30	1.237 €	44 €
5	545.455	75,2%	53,3	12,1	0,18	0,50	1.439 €	88 €
6	456.209	58,4%	55,0	18,0	0,49	0,62	1.721 €	88 €
7	390.905	61,7%	53,5	22,8	0,83	0,54	2.009 €	61 €
8	347.637	63,8%	55,4	19,1	1,02	0,61	2.259 €	95 €
9	300.830	63,3%	59,2	18,0	1,39	0,70	2.610 €	105 €
10	261.993	64,3%	62,2	18,1	1,67	0,81	2.997 €	108 €
11	229.622	62,0%	64,0	18,2	1,99	0,94	3.419 €	130 €
12	202.104	62,6%	64,1	19,1	2,25	0,98	3.885 €	138 €
13	177.743	60,3%	66,3	17,5	2,71	1,07	4.417 €	168 €
14	155.168	60,3%	68,5	15,6	3,20	1,13	5.060 €	208 €
15	134.141	60,2%	67,3	16,7	3,59	1,37	5.853 €	247 €
16	114.367	59,7%	69,1	15,4	4,25	1,54	6.865 €	348 €
17	95.128	59,3%	66,3	17,3	4,52	2,10	8.254 €	459 €
18	76.577	58,9%	65,2	18,1	5,05	2,53	10.253 €	759 €
19	56.093	55,6%	64,9	16,9	5,51	3,06	13.997 €	1.647 €
20	27.337	53,5%	62,2	18,0	6,01	3,38	28.721 €	13.404 €
Kasse A	7.058.189	58,7%	44,3	23,2	0,87	1,52	2.225 €	2.786 €

Tabelle 8-2 Beispielskasse B nach ausgewählten Kennziffern

Schicht	Pseudonyme		Alter		Zahl der HMGs		Zuweisung	
	Anzahl	Anteil	Mittel	StdAbw.	Mittel	StdAbw.	Mittel	StdAbw.
1	333.531	32,6%	24,4	15,6	0,08	0,60	663 €	208 €
2	215.815	59,2%	35,1	17,4	0,06	0,47	1.024 €	83 €
3	156.487	65,9%	52,3	15,2	0,28	0,73	1.413 €	158 €
4	116.734	58,0%	57,6	21,8	0,71	0,79	1.894 €	122 €
5	97.038	53,7%	60,3	21,0	1,12	0,80	2.278 €	134 €
6	81.077	59,7%	65,0	17,8	1,60	0,84	2.727 €	131 €
7	70.333	55,5%	67,8	16,7	1,92	0,93	3.143 €	127 €
8	61.476	56,9%	67,9	17,4	2,26	1,04	3.596 €	124 €
9	54.771	58,4%	68,9	17,1	2,56	1,09	4.036 €	135 €
10	49.013	57,7%	70,0	16,2	2,95	1,13	4.510 €	139 €
11	44.031	56,6%	71,0	14,6	3,34	1,15	5.021 €	156 €
12	39.547	56,5%	70,9	15,0	3,73	1,23	5.590 €	175 €
13	35.479	56,8%	71,3	14,3	4,14	1,34	6.231 €	201 €
14	31.619	56,5%	71,7	14,0	4,61	1,47	6.992 €	240 €
15	27.968	55,1%	71,1	14,3	4,94	1,79	7.905 €	286 €
16	24.495	54,7%	69,3	16,3	5,34	2,10	9.025 €	365 €
17	20.906	54,0%	69,7	15,8	6,01	2,30	10.574 €	546 €
18	17.048	51,1%	68,7	16,1	6,42	2,88	12.967 €	896 €
19	12.512	48,7%	67,9	15,3	6,93	3,30	17.669 €	2.147 €
20	5.480	45,9%	67,0	15,5	7,39	3,37	40.342 €	12.012 €
Kasse B	1.495.360	52,5%	50,9	24,8	1,46	2,04	2.957 €	3.699 €

Tabelle 8-3 Beispielskasse C nach ausgewählten Kennziffern

Schicht	Pseudonyme		Alter		Zahl der HMGs		Zuweisung	
	Anzahl	Anteil	Mittel	StdAbw.	Mittel	StdAbw.	Mittel	StdAbw.
1	86.679	24,1%	25,3	11,4	0,05	0,28	453 €	196 €
2	54.745	37,6%	24,1	12,0	0,01	0,14	718 €	28 €
3	46.632	33,5%	21,8	15,1	0,01	0,15	843 €	21 €
4	44.135	46,5%	18,6	10,3	0,01	0,14	891 €	15 €
5	41.293	20,1%	31,2	18,2	0,01	0,12	952 €	21 €
6	38.150	78,3%	33,0	14,8	0,02	0,18	1.030 €	40 €
7	34.935	68,6%	41,1	10,3	0,03	0,21	1.125 €	23 €
8	32.427	96,9%	37,6	10,1	0,04	0,21	1.212 €	27 €
9	30.009	61,7%	41,3	10,5	0,09	0,32	1.310 €	28 €
10	26.187	60,0%	49,2	14,5	0,41	0,56	1.501 €	97 €
11	22.074	45,5%	39,4	20,3	0,76	0,52	1.781 €	80 €
12	19.382	55,6%	41,0	20,6	0,93	0,47	2.028 €	57 €
13	17.474	59,2%	45,0	18,1	1,05	0,57	2.249 €	91 €
14	14.846	54,5%	46,7	19,9	1,38	0,70	2.648 €	125 €
15	12.459	49,8%	49,8	21,4	1,69	0,87	3.155 €	174 €
16	10.268	49,8%	49,2	23,6	1,99	1,09	3.828 €	213 €
17	8.244	43,0%	55,0	21,7	2,71	1,21	4.768 €	348 €
18	6.155	45,0%	58,0	18,8	3,52	1,60	6.386 €	618 €
19	4.247	43,6%	55,3	18,8	3,90	2,32	9.256 €	1.184 €
20	2.009	40,7%	54,8	19,0	4,92	3,08	19.569 €	11.419 €
Kasse C	552.350	48,0%	33,5	18,2	0,38	0,92	1.423 €	1.775 €

Tabelle 8-4 Beispielskasse D nach ausgewählten Kennziffern

Schicht	Pseudonyme		Alter		Zahl der HMGs		Zuweisung	
	Anzahl	Anteil	Mittel	StdAbw.	Mittel	StdAbw.	Mittel	StdAbw.
1	9.483	29,8%	23,6	13,3	0,06	0,48	640 €	175 €
2	6.746	32,1%	26,4	14,0	0,01	0,22	900 €	31 €
3	5.814	57,8%	33,8	18,2	0,03	0,24	1.045 €	53 €
4	4.902	70,9%	43,1	12,3	0,10	0,43	1.239 €	71 €
5	3.953	46,7%	55,8	13,7	0,31	0,59	1.537 €	108 €
6	3.279	47,3%	50,6	21,4	0,72	0,65	1.852 €	98 €
7	2.874	44,8%	52,3	20,4	0,92	0,66	2.113 €	76 €
8	2.474	46,5%	60,2	18,4	1,32	0,71	2.455 €	110 €
9	2.129	49,2%	61,7	16,4	1,64	0,71	2.852 €	120 €
10	1.853	47,4%	63,5	18,2	1,95	0,97	3.278 €	136 €
11	1.620	47,0%	65,1	17,3	2,20	0,95	3.749 €	127 €
12	1.427	46,5%	66,3	17,5	2,70	1,01	4.255 €	164 €
13	1.257	46,8%	68,6	15,4	3,02	1,23	4.834 €	181 €
14	1.103	46,5%	68,9	14,8	3,50	1,22	5.505 €	215 €
15	962	48,1%	70,2	14,1	4,12	1,29	6.313 €	268 €
16	824	45,9%	69,7	13,9	4,68	1,60	7.374 €	334 €
17	690	44,6%	67,4	17,5	5,03	2,10	8.803 €	512 €
18	560	45,4%	68,4	15,9	5,83	2,49	10.836 €	703 €
19	414	39,6%	68,0	14,6	6,14	3,14	14.650 €	1.805 €
20	182	41,8%	64,8	16,6	6,47	3,15	33.461 €	13.426 €
Kasse D	52.546	45,2%	44,9	23,1	0,97	1,65	2.312 €	2.990 €

Tabelle 8-5 Beispielskasse E nach ausgewählten Kennziffern

Schicht	Pseudonyme Frauen		Alter		Zahl der HMGs		Zuweisung	
	Anzahl	Anteil	Mittel	StdAbw.	Mittel	StdAbw.	Mittel	StdAbw.
1	895	21,2%	26,3	8,4	0,01	0,14	470 €	210 €
2	588	36,2%	23,4	11,8	0,01	0,13	716 €	18 €
3	519	45,7%	21,7	16,0	0,01	0,15	811 €	33 €
4	481	20,0%	17,7	11,3	0,00	0,06	877 €	13 €
5	449	35,6%	34,4	13,5	0,00	0,05	937 €	21 €
6	419	44,4%	20,2	18,9	0,02	0,15	1.003 €	33 €
7	379	79,9%	40,7	8,1	0,02	0,17	1.111 €	22 €
8	360	98,6%	28,6	6,1	0,02	0,13	1.171 €	14 €
9	344	99,1%	37,3	10,2	0,01	0,11	1.222 €	20 €
10	327	92,0%	36,4	7,9	0,02	0,16	1.287 €	20 €
11	312	81,7%	39,2	12,3	0,12	0,34	1.349 €	38 €
12	261	64,4%	48,5	18,3	0,45	0,51	1.608 €	80 €
13	219	62,6%	40,8	20,0	0,90	0,43	1.921 €	89 €
14	199	70,9%	36,5	15,5	0,89	0,39	2.124 €	41 €
15	179	68,7%	41,4	18,9	1,01	0,58	2.345 €	110 €
16	152	73,7%	43,0	21,0	1,28	0,78	2.763 €	129 €
17	126	46,0%	50,9	24,4	1,53	1,09	3.354 €	238 €
18	99	50,5%	43,1	27,7	1,85	1,26	4.251 €	294 €
19	67	55,2%	53,2	19,0	2,81	1,46	6.269 €	984 €
20	32	56,3%	50,6	17,8	2,56	2,12	13.204 €	4.973 €
Kasse E	6.407	54,3%	31,6	16,9	0,25	0,66	1.314 €	1.274 €

8.2 Beispielsklassen für die Stichprobenplanung nach § 42 Abs. 1 S. 1 Nr. 2 RSAV

Tabelle 8-6 Beispielskasse F nach ausgewählten Kennziffern

Schicht	Pseudonyme Frauen		Alter		Zahl der HMGs		Zuweisung	
	Anzahl	Anteil	Mittel	StdAbw.	Mittel	StdAbw.	Mittel	StdAbw.
1	373.932	58,09%	43,1	18,9	1,18	0,89	1.501 €	388 €
2	280.518	70,24%	49,2	20,8	1,05	0,37	2.001 €	59 €
3	254.578	73,25%	49,5	18,8	1,12	0,51	2.204 €	72 €
4	227.718	61,12%	62,5	17,0	1,28	0,61	2.465 €	74 €
5	205.590	64,50%	59,3	16,9	1,52	0,70	2.730 €	85 €
6	186.217	64,34%	63,7	16,3	1,69	0,79	3.014 €	78 €
7	169.140	63,36%	65,6	15,1	1,95	0,88	3.318 €	98 €
8	153.705	61,02%	65,1	16,7	2,20	0,90	3.651 €	101 €
9	140.286	62,52%	65,9	15,8	2,39	0,94	4.001 €	108 €
10	127.556	61,58%	67,7	14,9	2,78	0,99	4.400 €	122 €
11	115.717	60,86%	68,4	15,3	3,05	1,06	4.850 €	138 €
12	104.427	59,75%	68,9	15,3	3,45	1,14	5.374 €	164 €
13	94.038	61,04%	67,1	16,8	3,62	1,43	5.968 €	180 €
14	83.952	59,90%	69,2	15,3	4,19	1,46	6.685 €	237 €
15	73.808	59,37%	67,5	16,3	4,39	1,87	7.604 €	291 €
16	64.412	59,33%	65,4	18,1	4,60	2,24	8.713 €	367 €
17	54.923	59,07%	65,6	17,8	5,09	2,48	10.218 €	539 €
18	44.717	55,54%	64,7	17,5	5,35	2,90	12.550 €	845 €
19	33.206	56,36%	64,4	16,9	5,64	3,23	16.901 €	1.788 €
20	15.764	50,06%	61,9	18,0	6,44	3,36	35.602 €	13.844 €
Kasse A	2.804.204	62,98%	59,1	19,6	2,19	1,71	4.003 €	3.730 €

Tabelle 8-7 Beispielskasse G nach ausgewählten Kennziffern

Schicht	Pseudonyme		Alter		Zahl der HMGs		Zuweisung	
	Anzahl	Anteil	Mittel	StdAbw.	Mittel	StdAbw.	Mittel	StdAbw.
1	116.981	60,61%	48,0	21,3	1,36	1,22	1.582 €	449 €
2	82.616	61,03%	59,9	20,4	1,22	0,69	2.240 €	126 €
3	70.378	56,47%	64,4	16,7	1,53	0,84	2.630 €	109 €
4	61.836	57,07%	69,0	15,3	1,80	0,87	2.993 €	90 €
5	55.035	60,61%	69,1	14,7	2,21	0,93	3.363 €	117 €
6	49.528	55,60%	69,0	16,0	2,33	1,04	3.736 €	105 €
7	44.899	58,14%	69,4	14,8	2,64	1,10	4.122 €	117 €
8	40.905	57,87%	70,9	14,4	2,99	1,07	4.524 €	118 €
9	37.356	57,27%	71,4	14,1	3,31	1,10	4.954 €	130 €
10	34.151	55,91%	71,2	14,5	3,64	1,17	5.419 €	140 €
11	31.225	57,30%	70,5	15,3	3,89	1,33	5.927 €	154 €
12	28.453	56,92%	71,8	13,9	4,35	1,33	6.504 €	180 €
13	25.793	56,05%	71,7	13,9	4,69	1,53	7.175 €	209 €
14	23.259	55,59%	71,0	14,3	4,96	1,81	7.956 €	242 €
15	20.812	54,56%	69,3	16,5	5,29	2,08	8.892 €	297 €
16	18.340	54,22%	69,8	15,7	5,83	2,22	10.091 €	413 €
17	15.693	52,62%	68,8	16,1	6,24	2,64	11.792 €	588 €
18	12.859	49,76%	68,6	15,7	6,71	2,99	14.392 €	986 €
19	9.396	48,48%	67,7	15,4	7,13	3,45	19.697 €	2.646 €
20	4.078	45,68%	67,2	15,5	7,08	3,37	45.381 €	9.407 €
Kasse B	783.593	57,60%	65,1	18,6	2,79	2,06	4.723 €	4.388 €

Tabelle 8-8 Beispielskasse H nach ausgewählten Kennziffern

Schicht	Pseudonyme		Alter		Zahl der HMGs		Zuweisung	
	Anzahl	Anteil	Mittel	StdAbw.	Mittel	StdAbw.	Mittel	StdAbw.
1	17.619	41,29%	37,1	15,1	1,06	0,59	1.108 €	419 €
2	11.583	53,86%	37,0	16,2	1,02	0,32	1.686 €	65 €
3	10.461	53,37%	33,6	18,0	1,02	0,36	1.866 €	47 €
4	9.743	67,59%	42,3	18,0	1,06	0,36	2.004 €	38 €
5	9.265	58,09%	35,5	16,5	1,08	0,31	2.107 €	28 €
6	8.710	66,96%	41,6	16,4	1,14	0,47	2.242 €	49 €
7	7.968	49,76%	50,7	18,0	1,32	0,58	2.450 €	68 €
8	7.302	55,33%	48,1	17,1	1,45	0,66	2.674 €	63 €
9	6.672	55,20%	51,3	17,9	1,63	0,69	2.926 €	79 €
10	6.057	47,04%	53,5	17,6	1,87	0,78	3.223 €	99 €
11	5.445	49,29%	53,4	18,2	2,10	0,83	3.586 €	107 €
12	4.903	50,32%	55,2	18,2	2,30	0,95	3.982 €	125 €
13	4.355	46,98%	57,6	17,9	2,70	0,97	4.483 €	165 €
14	3.818	44,05%	58,7	18,2	3,05	1,13	5.114 €	210 €
15	3.297	46,74%	56,8	19,2	3,27	1,49	5.921 €	249 €
16	2.786	43,36%	59,3	18,2	3,86	1,66	7.009 €	370 €
17	2.321	45,11%	54,6	18,6	3,62	2,17	8.409 €	466 €
18	1.863	41,65%	55,6	19,1	4,24	2,46	10.482 €	799 €
19	1.348	42,21%	55,7	18,1	4,77	2,92	14.473 €	1.931 €
20	594	39,06%	51,6	20,7	5,53	3,43	32.894 €	14.418 €
Kasse C	126.110	52,09%	45,1	19,3	1,65	1,27	3.096 €	3.138 €

Tabelle 8-9 Beispielskasse I nach ausgewählten Kennziffern

Schicht	Pseudonyme		Alter		Zahl der HMGs		Zuweisung	
	Anzahl	Anteil	Mittel	StdAbw.	Mittel	StdAbw.	Mittel	StdAbw.
1	3.062	46,93%	42,4	18,5	1,18	0,90	1.460 €	382 €
2	2.251	56,95%	46,4	20,9	1,07	0,40	1.987 €	69 €
3	1.995	50,08%	53,6	19,6	1,21	0,61	2.242 €	93 €
4	1.764	48,30%	60,1	17,1	1,42	0,70	2.536 €	84 €
5	1.575	50,41%	61,5	16,5	1,60	0,70	2.840 €	94 €
6	1.422	47,68%	64,5	15,4	1,88	0,81	3.145 €	96 €
7	1.276	46,39%	65,7	15,5	2,21	0,90	3.503 €	105 €
8	1.163	48,58%	66,4	15,8	2,25	0,94	3.846 €	96 €
9	1.054	45,83%	68,3	13,7	2,76	0,92	4.244 €	123 €
10	958	46,76%	68,8	14,1	2,92	1,19	4.668 €	122 €
11	872	46,22%	70,1	13,3	3,33	1,18	5.130 €	137 €
12	791	48,67%	68,3	15,0	3,62	1,18	5.653 €	163 €
13	716	47,49%	69,9	14,3	4,09	1,28	6.247 €	191 €
14	639	46,17%	70,9	13,4	4,56	1,41	7.006 €	243 €
15	565	47,61%	69,1	14,4	4,80	1,85	7.917 €	280 €
16	492	45,53%	66,3	18,8	5,09	2,14	9.087 €	393 €
17	424	43,87%	69,4	15,0	5,92	2,37	10.552 €	492 €
18	351	38,18%	67,2	15,4	5,98	2,94	12.710 €	829 €
19	259	40,54%	67,0	16,0	6,46	3,34	17.302 €	2.188 €
20	113	44,25%	65,4	16,3	6,59	2,99	39.615 €	11.882 €
Kasse D	21.742	48,39%	59,5	19,6	2,35	1,84	4.114 €	3.889 €

Tabelle 8-10 Beispielskasse J nach ausgewählten Kennziffern

Schicht	Pseudonyme		Alter		Zahl der HMGs		Zuweisung	
	Anzahl	Anteil	Mittel	StdAbw.	Mittel	StdAbw.	Mittel	StdAbw.
1	141	48,23%	35,4	15,3	1,08	0,71	1.217 €	378 €
2	103	60,19%	37,3	16,0	0,96	0,19	1.695 €	65 €
3	91	62,64%	36,7	19,6	1,04	0,25	1.884 €	41 €
4	87	78,16%	41,6	15,1	1,03	0,24	2.009 €	36 €
5	82	67,07%	34,2	13,9	0,98	0,27	2.094 €	21 €
6	81	88,89%	35,6	10,1	0,95	0,35	2.154 €	21 €
7	77	87,01%	37,5	12,4	1,03	0,28	2.230 €	29 €
8	73	46,58%	45,9	18,3	1,18	0,67	2.392 €	49 €
9	67	62,69%	47,0	18,1	1,16	0,67	2.574 €	53 €
10	63	69,84%	46,4	17,1	1,48	0,62	2.755 €	55 €
11	58	72,41%	53,9	15,2	1,41	0,94	2.971 €	76 €
12	53	41,51%	56,0	18,8	1,70	0,70	3.241 €	103 €
13	48	43,75%	53,6	17,0	2,10	0,90	3.614 €	105 €
14	43	46,51%	52,6	16,0	2,02	0,91	4.054 €	130 €
15	37	64,86%	57,9	17,4	2,70	0,91	4.647 €	235 €
16	30	50,00%	55,7	22,5	2,73	1,48	5.698 €	291 €
17	24	62,50%	51,3	16,4	2,88	1,42	7.170 €	685 €
18	18	83,33%	48,6	17,2	2,17	1,15	9.680 €	562 €
19	13	53,85%	51,8	15,4	2,54	2,47	13.215 €	1.720 €
20	8	50,00%	52,1	20,7	3,75	2,60	22.453 €	5.181 €
Kasse E	1.197	62,99%	43,2	17,9	1,37	0,92	2.894 €	2.468 €

8.3 Prüfungsaffine Kennziffern für die Beispielskassen und alle im Ausgleichsjahr 2009 am Morbi-RSA beteiligten Krankenkassen („GKV-Kassen“)

8.3.1 Grundgesamtheiten für die Nr. 1-Stichproben

In Tabelle 8-11 werden folgende Kennziffern ausgewiesen:

- K₁ Mittelwert der Versichertentage im Ausgleichsjahr
- K₂ Anteil der Versicherten mit Krankengeldanspruch im Ausgleichsjahr
- K₃ Mittelwert der Versichertentage mit Krankengeldanspruch im Ausgleichsjahr
- K₄ Anteil der Versicherten mit DMP-Einschreibung im Ausgleichsjahr
- K₅ Mittelwert der Versichertentage mit DMP-Einschreibung im Ausgleichsjahr
- K₆ Anteil der Versicherten mit Erwerbsminderung im Ausgleichsjahr
- K₇ Mittelwert der Versichertentage mit Erwerbsminderung im Ausgleichsjahr

Tabelle 8-11 Beispielskassen und GKV-Kassen nach prüfungsaffinen Kennziffern, Nr. 1-Stichproben

Kasse	Kennziffer						
	K ₁	K ₂	K ₃	K ₄	K ₅	K ₆	K ₇
A	352,6	42,80%	142,0	7,8%	25,6	2,5%	8,3
B	348,2	23,92%	71,9	16,8%	54,5	5,4%	18,2
C	339,0	56,95%	185,7	3,2%	9,5	1,1%	3,6
D	354,0	40,79%	137,9	12,6%	40,9	1,7%	5,7
E	343,0	63,35%	212,5	0,0%	0,0	0,3%	1,0
GKV	349,1	42,22%	138,5	8,1%	25,2	2,5%	8,3

8.3.2 Grundgesamtheiten für die Nr. 2-Stichproben

Für die Grundgesamtheiten, die zu den Nr. 2-Stichproben gehören, werden folgende Kennziffern zur Darstellung gebracht:

K ₁	Anteil der Versicherten mit Auslandsaufenthalt im Vorjahr
K ₂	Mittelwert der Versichertentage mit Auslandsaufenthalt im Vorjahr
K ₃	Anteil der Versicherten mit Erwerbsminderung im Vorjahr
K ₄	Mittelwert der Versichertentage mit Erwerbsminderung im Vorjahr
K ₅	Anteil der Versicherten mit Kostenerstattung im Vorjahr
K ₆	Mittelwert der Versichertentage mit Kostenerstattung im Vorjahr
K ₇	Mittelwert der Versichertentage im Ausgleichsjahr
K ₈	Mittlere Zahl der belegten HMGs
K ₉	Mittlere Zahl der unterschiedlichen relevanten stationären Hautdiagnosen
K ₁₀	Mittlere Zahl der unterschiedlichen relevanten stationären Nebendiagnosen
K ₁₁	Mittlere Zahl der unterschiedlichen gesicherten ambulanten Diagnosen
K ₁₂	Mittlere Zahl der Arzneimittelverordnungen mit unterschiedlichen zuordenbarem ATC-Code

Tabelle 8-12 Beispielskassen und GKV-Kassen nach prüfungsaffinen Kennziffern, Nr. 2-Stichproben

Kasse	Kennziffer											
	K ₁	K ₂	K ₃	K ₄	K ₅	K ₆	K ₇	K ₈	K ₉	K ₁₀	K ₁₁	K ₁₂
F	0,43%	1,5	6,08%	20,8	0,77%	20,8	357,1	2,2	0,19	0,78	4,3	4,8
G	0,11%	0,3	10,18%	35,2	0,01%	35,2	353,5	2,8	0,24	0,98	5,4	5,5
H	2,11%	7,4	4,30%	14,3	0,13%	14,3	348,2	1,7	0,17	0,53	3,4	4,0
I	1,07%	3,8	4,01%	13,5	0,02%	13,5	356,3	2,4	0,19	0,75	4,7	5,1
J	1,00%	3,7	1,50%	4,8	5,51%	4,8	355,9	1,4	0,14	0,49	3,0	3,0
GKV	0,94%	3,3	6,43%	21,9	0,31%	21,9	354,7	2,2	0,20	0,73	4,4	4,9

8.4 Stichprobenumfangsplanung in Abhängigkeit vom Variationskoeffizienten

Tabelle 8-13 Erforderliche Stichprobenumfänge in den fünf Beispielskassen für die Nr. 1-Stichproben in Abhängigkeit vom Variationskoeffizient (V) des Korrekturbeitrags und der gewünschten (relativen) Genauigkeit ε

Kasse A	N =	7.058.189				
V \ ε	5%	10%	15%	20%	25%	30%
1	1.536	384	171	96	61	43
3	13.802	3.456	1.536	864	553	384
6	54.887	13.802	6.141	3.456	2.212	1.536
Kasse B	N =	1.495.360				
V \ ε	5%	10%	15%	20%	25%	30%
1	1.535	384	171	96	61	43
3	13.703	3.449	1.535	864	553	384
6	53.344	13.703	6.121	3.449	2.209	1.535
Kasse C	N =	552.350				
V \ ε	5%	10%	15%	20%	25%	30%
1	1.532	384	171	96	61	43
3	13.491	3.436	1.532	863	553	384
6	50.281	13.491	6.079	3.436	2.204	1.532
Kasse D	N =	52.546				
V \ ε	5%	10%	15%	20%	25%	30%
1	1.493	381	170	96	61	43
3	10.948	3.244	1.493	850	547	381
6	26.948	10.948	5.503	3.244	2.123	1.493
Kasse E	N =	6.407				
V \ ε	5%	10%	15%	20%	25%	30%
1	1.239	362	166	95	61	42
3	4.378	2.246	1.239	762	509	362
6	5.742	4.378	3.137	2.246	1.645	1.239

Tabelle 8-14 Erforderliche Stichprobenumfänge in den fünf Beispielskassen für die Nr. 2-Stichproben in Abhängigkeit vom Variationskoeffizient (V) des Korrekturbeitrags und der gewünschten Genauigkeit ε

Kasse F	N =	2.804.204				
V \ ε	5%	10%	15%	20%	25%	30%
1	1.536	384	171	96	61	43
3	13.761	3.453	1.536	864	553	384
5	37.895	9.571	4.262	2.399	1.536	1.067
Kasse G	N =	783.593				
V \ ε	5%	10%	15%	20%	25%	30%
1	1.534	384	171	96	61	43
3	13.589	3.442	1.534	863	553	384
5	36.619	9.487	4.245	2.394	1.534	1.066
Kasse H	N =	126.110				
V \ ε	5%	10%	15%	20%	25%	30%
1	1.518	383	171	96	61	43
3	12.463	3.365	1.518	858	551	383
5	29.445	8.924	4.129	2.356	1.518	1.058
Kasse I	N =	21.742				
V \ ε	5%	10%	15%	20%	25%	30%
1	1.435	377	169	96	61	43
3	8.453	2.983	1.435	831	539	377
5	13.884	6.661	3.568	2.162	1.435	1.017
Kasse J	N =	1.197				
V \ ε	5%	10%	15%	20%	25%	30%
1	673	291	149	89	58	41
3	1.102	889	673	502	378	291
5	1.161	1.064	935	799	673	564

9 Literatur/Quellen

- Anscombe FJ (1949) Large sample theory of sequential estimation. *Biometrika* 36: 455-458
- Anscombe FJ (1952) Large sample theory of sequential estimation. *Proc. Camb. Phil. Soc.* 48: 600- 607
- Anscombe FJ (1953) Sequential estimation. *J. Roy. Statist. Soc. B* 15: 1-29
- Bauer P, Scheiber V, Wohlzogen FX (1986) *Sequentielle statistische Verfahren*. Gustav Fischer Verlag, Stuttgart
- BVA (2013) Gutachten nach § 42 RSAV: Inhaltliche Anforderungen (Neufassung). Bonn
- BVA (2013a) Korrekturbeträge nach § 15a RSAV. Bonn
- Chatterjee SK (1991) Two-Stage and Multistage Procedures. In: Ghosh BK, Sen P (1991) *Handbook of Sequential analysis*. Marcel Dekker, Inc., New York
- Cochran WG (1972) *Stichprobenverfahren*. Berlin New York
- Cox DR (1952) Estimation by double sampling. *Biometrika* 39: 217-227.
- Dalenius T, Gurney M (1951) The problem of optimal stratification II. *Scandinavisk Aktuariatidskrift* 34: 198-203
- Degen H, Lorscheid P (2002) *Statistik-Lehrbuch mit Wirtschafts- und Bevölkerungsstatistik*, 2. Auflage, Oldenbourg Verlag, München
- Drexl A (1982) *Geschichtete Stichprobenverfahren*. Verlagsgruppe Athenäum / Hain / Scriptor / Hanstein
- Ekman G (1959) An approximation useful in univariate stratification. *Ann. Math. Statist.* 30: 219-229
- Fischer L (2013) Berechnung des Stichprobenumfangs für die Prüfungen nach § 15a RSAV. Persönliche Mitteilung
- Ghosh BK, Sen P (1991) *Handbook of Sequential analysis*. Marcel Dekker, Inc., New York
- Ghosh N, Mukhopadhyay N, Sen PK (1997) *Sequential estimation*. John Wiley & Sons, New York
- Ghosh BK (1991) A brief history of sequential analysis. In: Ghosh BK, Sen P (1991) *Handbook of Sequential analysis*. Marcel Dekker, Inc., New York
- Kirsch M (2013) Potentielle Änderungen der Fehlerquote in den Prüfungen nach § 42 Abs. 1 S.1 Nr. 1 RSAV gegenüber denjenigen nach § 15a RSAV. Persönliche Mitteilung
- Mukhopadhyay N (1982) Stein's two-stage procedure and exact consistency. *Skandinavisk Aktuariatidskrift*: 110-122

- Polaszek U (2013) Personalressourcen Prüfdienste. Persönliche Mitteilung
- Schäfer T (2004) Stichprobenverfahren. In: Voß W, et al. (Hrsg.) Taschenbuch der Statistik, 2. verbesserte Auflage. Fachbuchverlag Leipzig (im Carl Hanser Verlag), Leipzig
- Sinha BK (1991) Sequential Methods for Finite Populations. In: Ghosh BK, Sen P (1991) Handbook of Sequential analysis. Marcel Dekker, Inc., New York
- Stein C (1945) A two-sample test for a linear hypothesis whose power is independent of the variance. Ann. Math. Statist. 16: 243-258
- Sukhatme PV, Sukhatme BV (1970) Sampling Theory of Surveys with Applications. Ames, Iowa, U.S.A