



**Entwurf für Fehlerverfahren und Datenbereinigung  
für die Datenerhebung 2006/2007 zur  
Weiterentwicklung des RSA gemäß § 30 RSAV sowie  
der Daten für die Übergangsregelungen zur  
Einführung des Gesundheitsfonds gemäß § 34 RSAV**

29. Oktober 2008

<b>KAPITEL 1: ÜBERSICHT</b>	<b>2</b>
<b>KAPITEL 2: STUFE A - TECHNISCHE PRÜFUNGEN</b>	<b>4</b>
<b>KAPITEL 3: STUFE B - DATENPRÜFUNGEN</b>	<b>7</b>
<b>KAPITEL 4: STUFE C - MEDIZINISCH-INHALTLICHE PRÜFUNGEN</b>	<b>11</b>
<b>KAPITEL 5: STUFE D - AUFFÄLLIGKEITSPRÜFUNGEN</b>	<b>13</b>
<b>ANHANG - BERECHNUNG DES TOLERANZBEREICHES DER AUSSCHÖPFUNGSQUOTEN LAUT GUTACHTEN ZUR STICHPROBENERHEBUNG NACH § 267 ABS. 7 NR. 1 UND 2 SGB V VON PROF. FAULBAUM</b>	<b>14</b>

# Kapitel 1: Übersicht

Das Fehlerverfahren ist ein zentraler Bestandteil im Prozess zur Datenbereinigung der nach § 30 und 34 RSAV von den Spitzenverbänden der Krankenkassen gemeldeten Daten. Im Folgenden wird das Konzept für die Datenbereinigung und das Fehlerverfahren zu Zwecken der Transparenz beschrieben. Der Umgang mit fehlerhaften Daten soll somit für alle Beteiligten nachvollziehbar gehalten werden.

Die Übermittlung der von den Spitzenverbänden zu liefernden Daten erfolgt im Lieferjahr 2008 in Form von USB-Festplatten. Jede Datei muss einen auf die Datensätze der jeweiligen Datei bezogenen entsprechenden Vorlauf- und Nachlaufsatz enthalten. Die Spitzenverbände liefern entsprechend der Vereinbarung nach § 267 Abs. 7 Nr. 1 und 2 SGB V in der Fassung vom 02.06.2008 für die Jahre 2006 und 2007 jeweils die folgenden Satzarten:

- für die Versichertenstammdaten die Satzart 100,
- für die Versichertenstammdaten des monatlichen Abschlagsverfahren die Satzart 111,
- für die Daten der Arzneimittelversorgung die Satzart 400,
- für die Daten der Krankenhausversorgung die Satzart 500,
- für die Daten der ambulanten ärztlichen Versorgung die Satzart 600,
- für die personenbezogenen berücksichtigungsfähigen Leistungsausgaben die Satzart 700, (nur für 2007)
- für die erweiterten Versichertenstammdaten die Satzart 800,
- für die länderbezogenen Einnahmen die Satzart 900.

## Aufbau des Prüfverfahrens:

Das weitere Fehlerverfahren für die oben benannten Daten besteht aus den folgenden, grundsätzlichen Stufen:

1. **Stufe A: Technische Prüfungen:**
  - a. Zusammensetzung und fehlerfreie Übermittlung der Datenlieferung,
  - b. Formatprüfung.
2. **Stufe B: Datenprüfungen,**
  - a. Pseudonymbezogene Prüfungen
  - b. Prüfungen auf Kassenebene.
3. **Stufe C: medizinisch-inhaltliche Prüfungen.**
4. **Stufe D: Auffälligkeitsprüfungen.**

**Fehler im Dateiaufbau führen zu einer grundsätzlichen Abweisung der betroffenen Satzarten**, da diese so gravierend sind, dass die Daten nicht korrekt eingelesen werden können. Es erfolgt eine entsprechende Rückmeldung des BVA an den Spitzenverband mit der Benachrichtigung über die Abweisung der betroffenen Daten. **Der Spitzenverband wiederholt die Datenlieferung für die betroffenen Satzarten.** Prüfungen auf der Stufe B können Kassenausschlüsse nach sich ziehen. Der genaue Umgang mit den spezifischen Fehlern auf den Stufen B-D wird in den jeweiligen Kapiteln näher beschrieben. Übersichten über die auf den einzelnen Stufen möglichen Fehler finden sich in den jeweiligen Kapiteln.

Die Regeldatenlieferung besteht aus einer vollständigen Lieferung der acht Satzarten für einen Spitzenverband. **Im Falle einer Korrekturlieferung übermittelt der Spitzenverband für die**

betroffene(n) Krankenkasse(n) die betroffene Satzart vollständig neu. Grundsätzlich werden alle Datensätze in der neu gelieferten Satzart ersetzt.

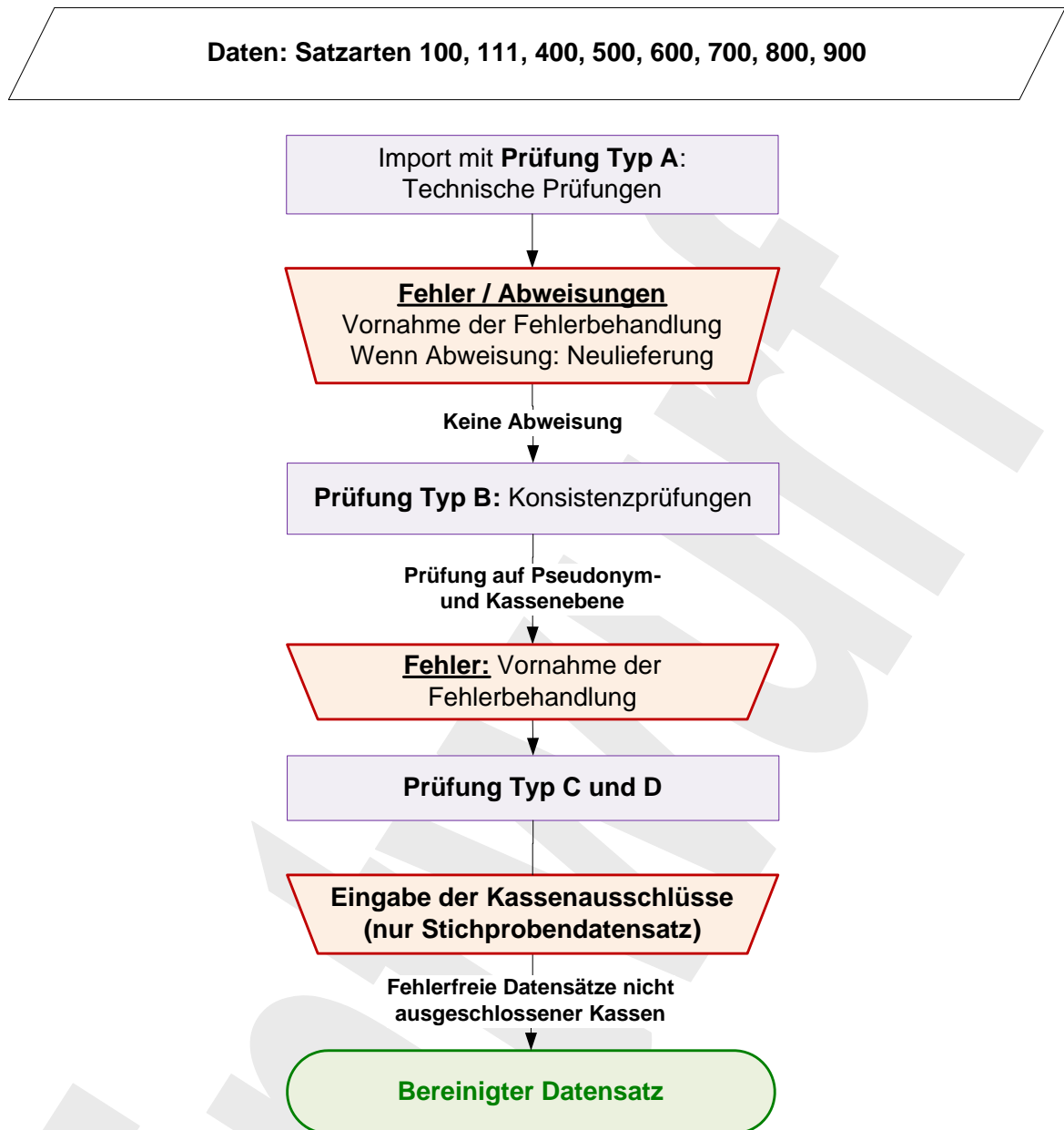


Abbildung 1: Ablauf des Fehlerverfahrens

**Zuständig für Rückfragen:**

Prüfung Typ A  
Prüfung Typ B-D

Ursula Roth  
Dorothee Schmidt

**Kontakt:**

[Ursula.Roth@bva.de](mailto:Ursula.Roth@bva.de)  
[Dorothee.Schmidt@bva.de](mailto:Dorothee.Schmidt@bva.de)

## Kapitel 2: Stufe A - Technische Prüfungen

Der eigentlichen Datenprüfung ist eine technische Prüfung vorgeschaltet, mit der sichergestellt wird, dass die Datenlieferung für die Aufnahme in die Datenbank geeignet ist und die Datenfelder richtig befüllt wurden. **Fehler auf dieser Stufe führen zu einer Abweisung der betroffenen Satzart und erfordern eine vollständige Neulieferung.** In diesem Falle nimmt das BVA Kontakt mit dem Spitzenverband auf, um eine Neulieferung der Daten zu veranlassen. Die technische Prüfung beinhaltet eine Vorprüfung und eine Formatprüfung.

Die Vorprüfung besteht aus dem Abgleich der Vor- und Nachlaufsätze mit den gelieferten Dateien. Mit Hilfe der Vorlauf- und Nachlaufsätze wird sichergestellt, dass die gelieferten Daten für das Einlesen in die Datenbank geeignet sind. Der Vorlaufsatz enthält Angaben zur Dateibezeichnung und zur enthaltenen Satzart. Der Nachlaufsatz enthält Daten, mit denen die Vollständigkeit und die Zusammensetzung der übermittelten Datensätze der gelieferten Satzarten überprüft wird.

Der Formatprüfung liegt die Vereinbarung der Spitzenverbände zu § 267 Abs. 7 Nr. 1 und 2 SGB V in der Fassung vom 2. Juni 2008 zugrunde. An dieser Stelle wird geprüft, ob die in der Anlage 3.5 und Anlage 4.1 beschriebenen Formatanforderungen eingehalten wurden.

### Übersicht der Fehler:

<b>Fehler</b>	<b>Beschreibung</b>	<b>Satzart</b>
A001	Die im Vorlaufsatz angegebene Satzart ist nicht bekannt oder weicht von der im Datensatz / Nachlaufsatz angegebenen Satzart ab.	100, 111, 400, 500, 600, 700, 800, 900
A002	Das im Vorlaufsatz angegebene Berichtsjahr weicht vom dem im Datensatz / Nachlaufsatz angegebenen Jahr ab.	100, 111, 400, 500, 600, 700, 800, 900
A003	Das im Vorlaufsatz angegebene Berichtsjahr ist nicht zulässig.	100, 111, 400, 500, 600, 700, 800, 900
A004	Das im Vorlaufsatz angegebene Lieferjahr ist nicht größer oder gleich dem Berichtsjahr.	100, 111, 400, 500, 600, 700, 800, 900
A005	Das im Vorlaufsatz angegebene Erstellungsdatum ist nicht größer als das Berichtsjahr oder liegt nicht im laufenden Jahr.	100, 111, 400, 500, 600, 700, 800, 900
A006	Die Anzahl der Datensätze stimmt nicht mit der Angabe im Nachlaufsatz überein.	100, 111, 400, 500, 600, 700, 800, 900
A007	Ein oder mehrere Datensatzfelder enthalten keine Daten oder die Daten entsprechen nicht dem vorgegebenen Format. <sup>1</sup>	100, 111, 400, 500, 600, 700, 800, 900
A008	Doppelter Datensatz (vollständiger Satz). Löschung der identischen Datensätze, so dass nur noch einer vorhanden ist.	100, 111, 400, 500, 600, 700, 800, 900
A009	Das angegebene Pseudonym ist in der SA100 nicht vorhanden.	400, 500, 600, 700, 800
A010	Die Meldung der Versichertentage zu diesem Pseudonym umfasst mehr als 365 / 366 Tage bzw. 181 / 182 Tage für die SA111-Meldung für den Zeitraum Jan bis Juni. Für die Satzart 800 bezieht sich der Fehler auf die Summe der Versichertentagen in den Versichertengruppen VG 1, 2 und 4.	100, 111, 800

<sup>1</sup> Wenn dieser Fehler durch Löschung von Datensätzen bei den Einleseroutinen entsteht, erfolgt keine Löschung der restlichen Datensätze.

A011	Die angegebene Betriebsnummer ist unbekannt oder für das Berichtsjahr nicht zugelassen.	100, 111, 400, 500, 600, 700, 800, 900
A012	Der angegebene Wert zum Krankenversicherungs-Kennzeichen ist nicht "0" oder "1".	100, 111
A013	Das Geburtsjahr ist größer als das Berichtsjahr oder liegt über 110 Jahre zurück.	100, 111
A014	Der angegebene Wert zum Geschlecht ist nicht "1" oder "2".	100, 111
A015	Die Anzahl der Tage der Erwerbsminderung (bei SA 800: Tage in VG 4) sind größer als die in SA 100 gemeldeten Versichertentage des Pseudonyms.	100, 800
A016	Der angegebene Wert zur Extrakorporalen Blutreinigung ist nicht "0" oder "1".	100
A017	Der angegebene Wert zum Merkmal Verstorben ist nicht "0" oder "1".	100
A018	Die Anzahl der DMP-Tage ist größer als die Anzahl der in SA 100 gemeldeten Versichertentage des Pseudonyms. Für die SA800 bezieht sich der Fehler auf die Summe der Felder 9-14 (DMP 1-6).	100, 800
A019	Das Verordnungsdatum liegt nicht im Berichtsjahr.	400
A020	Der angegebene Wert zum Entlassungstag liegt nicht im Berichtsjahr.	500
A021	Der angegebene Wert zur Lokalisation ist nicht "0" oder "1".	500, 600
A022	Die Diagnose entspricht nicht den formalen Anforderungen (1.Stelle = A-Z; 2.-3. Stelle = numerisch; 4. - 7. Stelle = Zahlen: 0 bis 9 und vorgegebene Sonderzeichen ("!", ".", "*", "+", "-" und "#"); Gesamtlänge mind. 3, max. 9 Stellen).	500, 600
A023	Die Versichertentage des Pseudonym sind gleich Null.	100, 111
A024	Der angegebene Wert zum Quartal ist nicht "1", "2", "3" oder "4".	600
A025	Der angegebene Wert zur Qualifizierung ist nicht "V", "Z", "A", "G" oder "0".	600
A026	Die Prüfsumme stimmt nicht mit der Summe des letzten Datenfeldes über alle Datensätze überein.	100, 111, 400, 500, 600
A027	Der Betrag Ärzte-Su stimmt nicht mit der Summe über alle Datensätze überein.	700
A028	Der Betrag Zahnärzte-Su stimmt nicht mit der Summe über alle Datensätze überein.	700
A029	Der Betrag Apotheken-Su stimmt nicht mit der Summe über alle Datensätze überein.	700
A030	Der Betrag Krankenhäuser-Su stimmt nicht mit der Summe über alle Datensätze überein.	700
A031	Der Betrag Sonstige LA-Su stimmt nicht mit der Summe über alle Datensätze überein.	700
A032	Der Betrag extrak. Blutr.-SU stimmt nicht mit der Summe über alle Datensätze überein.	700
A033	Der Betrag DMP-Mehrleistung-Su stimmt nicht mit der Summe über alle Datensätze überein.	700
A034	EMR-Tage obwohl das Alter größer 65 ist.	100
A035	KV-NR. = 1 aber Pseudonym hat keine 38 Zeichen. KV-Nr. = 0, aber ersten 8 Zeichen entsprechen nicht der Betriebsnummer. <sup>2</sup>	100, 111
A036	Nicht besetzt	-
A037	Packungsanzahl der PZN = „0“. (Wird auf „1“ umgesetzt.)	400
A038	Der Länderschlüssel liegt nicht zwischen 00 und 16.	800, 900
A039	Für einen Datensatz mit Länderschlüssel 00 enthalten die Felder 6-14 Werte (wird auf „Null“ gesetzt).	800
A040	Für ein Pseudonym der Stichprobe in SA 100 sind keine Ausgaben in SA 700 gemeldet. <sup>3</sup>	100, 700

<sup>2</sup> Ab dem Meldejahr 2009 gilt hier: wenn KV-Nr = 0 Pseudonym hat keine 19 Zeichen oder ersten 8 Zeichen entsprechen keiner Betriebsnummer, wenn KV-Nr = 1 Pseudonym hat keine 38 Zeichen.

<sup>3</sup> Diese Prüfung gilt am dem Meldejahr 2009, da das Stichprobenkennzeichen erst mit dem Meldejahr 2009 geliefert wird. Eine entsprechende Prüfung des Stichprobenkennzeichens wird ebenfalls ab dem Meldejahr 2009 durchgeführt werden.

---

A041 Multiples Pseudonym in der SA 100 und/oder 111 mit unterschiedlichem Inhalt in den anderen Datenfeldern; alle betroffenen Datensätze werden gelöscht. 100, 111

---

Die Fehler A037 und A039 führen nicht zu einer Zurückweisung der gemeldeten Daten. Doppelte Datensätze in der SA 500, die aufgrund der fehlenden Angabe des Fallzählers<sup>4</sup> entstehen können, werden gelöscht.

---

<sup>4</sup> Gilt ab dem Meldejahr 2009

## Kapitel 3: Stufe B - Datenprüfungen

Nach erfolgreich durchlaufener technischer Prüfung werden die Daten im Anschluss der eigentlichen Datenprüfung unterzogen. Die Ebene der Datenprüfungen überwacht die logischen Zusammenhänge zwischen den Satzarten und dient dazu sicherzustellen, dass sich die Daten sinnvoll aggregieren lassen. Sie ist in zwei Schritten aufgebaut. Der erste beschäftigt sich pseudonymbezogen mit der Kompatibilität der Satzarten untereinander. Die zweite prüft die Kompatibilität der Satzarten untereinander auf Kassenebene. Abbildung 1 verdeutlicht das weitere Vorgehen.

### Pseudonymbezogener Abgleich der Satzarten

Dieser Abgleich betrifft insbesondere die Satzarten 100 und 800. Ein Fehler auf dieser Ebene kann eine Löschung oder Änderung der betroffenen Datensätze nach sich ziehen. Für den Fall, dass eine Fehlerhäufung bei einer Krankenkasse auffällt, wird das BVA den Spitzenverband kontaktieren, um die Ursache zu klären und ggf. eine Neulieferung zu veranlassen. In diesem Fall wird die **komplette Satzart** neu geliefert und ersetzt die bisher übermittelten Daten. Für den Fall, dass eine Neulieferung nicht möglich ist, wird mit den Datensätzen nach der unten aufgeführten Fehlerübersicht verfahren.

### Übersicht der pseudonymbezogenen Fehler:

<i>Fehler</i>	<i>SA</i>	<i>Variable / Beschreibung</i>	<i>Fehlerdefinition</i>	<i>Konsequenz</i>
B001	100	verstorben 2006=/ lebendig 2007	Pseudonym im Folgejahr in den Kassendaten wieder auf, obwohl es im Vorjahr als verstorben geführt wird. (Gilt auch für Kassenwechsler).	Hinweis, verstorben auf „0“ setzen
B002	100, 111	Versichertentage Pseudonym > 365	Nach Aggregation auf GKV-Ebene wurden für das Pseudonym, das aus einer einheitlichen KV-Nr. erzeugt wurde, mehr als 365 Versichertentage ausgewiesen.	Hinweis, Klärung, ggf. Neulieferung sonst proportionale Kürzung von VT, EMT und DMPT
B003	100, 111	Abweichung in Alter oder Geschlecht	Das Pseudonym weist nach Aggregation auf GKV-Ebene Abweichungen in Alter oder Geschlecht auf.	Löschung Datensatz
B004	800	Versichertentage Pseudonym > 365	Die Summe der Versicherungstage (nach Aggregation auf GKV-Ebene) der Felder 6-8 (VG 1, 2 u. 4), bzw. 9-14 (DMP 1 bis 6) ist größer als 365 (im Schaltjahr 366) Tage.	Löschung Datensatz
B005	100, 800	Abweichung DMP- Tage	Die Summe der DMP-Tage in SA 800 (Felder 9-14) ist größer als die SA 100 ausgewiesenen DMP-Tage.	Löschung Datensatz
B006	100, 800	Abweichung EMR- Tage	Die Anzahl der Tage in Feld 8 ist größer als die in SA 100 ausgewiesenen EMR-Tage.	Löschung Datensatz



## Kassenbezogener Abgleich der Satzarten

In diesem Schritt wird überprüft, ob systematische Fehler im Abgleich der Satzarten 100, 400, 500 und 600 mit der Satzart 700 vorliegen. Für den Fall, dass ein solcher Fehler auftritt, klärt das BVA mit dem zuständigen Spitzenverband, ob eine Unplausibilität vorliegt und wenn ja, ob die Daten neu geliefert werden können. Für den Fall, dass letzteres nicht möglich ist, entscheidet das BVA nach Anhörung des zuständigen Spitzenverbandes darüber, ob die **Fehlerverletzung so gravierend** ist, dass **sämtliche Daten der Kasse aus dem Verfahren für die Erstellung des Stichprobendatensatzes zur Ermittlung der Kostengewichte ausgeschlossen** werden.<sup>5</sup> In diesem Fall werden die betroffenen Datensätze gelöscht und die übrigen Daten wieder dem Verfahren zugeleitet. Für den Fall, dass eine Neulieferung möglich ist, wird auch in diesem Fall die **komplette Satzart** neu geliefert und ersetzt die bisher übermittelten Daten.

### Übersicht der kassenbezogenen Fehler:

<i>Fehler</i>	<i>SA</i>	<i>Variable / Beschreibung</i>	<i>Fehlerdefinition</i> <sup>6</sup>	<i>Konsequenz</i>
B007	100	Ausschöpfungsquote Versicherungstage auffällig	Die Ausschöpfungsquote der Versicherungstage einer Kasse ist unplausibel. <sup>7</sup>	Hinweis, ggf. Ausschluss der Kasse
B008	700	Ausschöpfungsquote der Leistungsaus- gaben auffällig	Die Ausschöpfungsquote der Leistungs- ausgaben einer Kasse ist unplausibel. <sup>2</sup>	Hinweis, ggf. Ausschluss der Kasse
B009	100	Persistenz der Pseudonyme od. Quote Kassenwechsler auffällig	Der Anteil der Pseudonyme des Vorjahres, die auch im Ausgleichsjahr bei derselben Krankenkasse gemeldet werden, oder die Quote der Kassenwechsler weisen auf einen Fehler beim Pseudonymisierungsverfahren hin. Krankenkassen werden dabei einheitlich auf dem Fusionsstand des Ausgleichsjahres betrachtet.	Hinweis, ggf. Ausschluss der Kasse
B010	400, 700	Abgleich Ausgaben unstimmig	Bei mehr als $x^{\text{krit}}$ % der Versicherten einer Kasse wurden Ausgaben im HLB Apotheke der SA 700 vermerkt, obwohl im gleichen Jahr keine Rezepte in SA 400 gemeldet wurden.  Bzw. bei mehr als $x^{\text{krit}}$ % der Versicherten einer Kasse sind keine Ausgaben in SA 700 vermerkt, obwohl Rezepte geltend gemacht wurden.	Hinweis, ggf. Ausschluss der Kasse

<sup>5</sup> Aus dem Verfahren ausgeschlossen bedeutet in diesem Falle, dass die Daten dieser Kassen nicht für die Berechnungen der Morbiditätszuschläge (Regression) verwendet werden. In die Berechnungen der Zuweisungen (Beitragsbedarf der Kasse) fließen die Daten, die auf Ebene des einzelnen Pseudonyms fehlerfrei geblieben sind, dennoch ein.

<sup>6</sup> Der in dieser Spalte kritische Wert  $x^{\text{krit}}$  wird nach Eingang der Daten ermittelt. Er ergibt sich aus dem Durchschnitt der jeweils relevanten Variable über alle Kassen zuzüglich der zweifachen Standardabweichung.

<sup>7</sup> Die entsprechenden Prüfkriterien beziehen sich auf die jeweiligen Variablen, so wie sie im Gutachten zur Stichprobenerhebung nach § 267 Abs. 7 Nr. 1 und 2 SGB V im Auftrag der Spitzenverbände der Krankenkassen der GKV, März 1997, von Prof. Faulbaum definiert wurden. Die hierbei verwendeten Formeln befinden sich im Anhang. Da die dafür benötigten Informationen in der Datenlieferung 2006 noch nicht zur Verfügung standen, wurden die Prüfkriterien für das Fehlerverfahren 2006 für jeden Leistungsbereich einzeln festgelegt. Die Kriterien des Faulbaumgutachtens werden ab dem Meldejahr 2009 angewandt.

B011	500, 700	Abgleich Ausgaben unstimmig	Bei mehr als $x^{\text{krit}}$ % der Versicherten einer Kasse wurden in der SA 500 Diagnosen gestellt, aber in SA 700 des gleichen Jahres für den HLB KH keine Ausgaben geltend gemacht. Bzw: bei mehr als $x^{\text{krit}}$ % der Versicherten einer Kasse wurden in der SA 700 für den Bereich KH Ausgaben geltend gemacht, obwohl keine Diagnosen, keine Behandlung od. stationäre Anschlussreha in SA 500 des gleichen Jahres angezeigt wurden.	Hinweis, ggf. Ausschluss der Kasse
B012	600, 700	Abgleich Ausgaben unstimmig.	Bei mehr als $x^{\text{krit}}$ % der Versicherten einer Kasse wurden in der SA 600 Diagnosen gestellt, aber in SA 700 des gleichen Jahres für den HLB Ärzte keine Ausgaben geltend gemacht. Bei mehr als $x^{\text{krit}}$ % der Versicherten einer Kasse wurden in der SA 700 für den ambulanten Bereich Ausgaben geltend gemacht, obwohl keine Diagnosen in SA 600 des gleichen Jahres angezeigt wurden.	Hinweis, ggf. Ausschluss der Kasse

### Übersicht über Hinweise:

Hinweiskode	Satzart	Variable / Beschreibung	Definition	Konsequenz
B013	100	Verstorben	Im selben Jahr ist das Pseudonym nach Aggregation auf GKV-Ebene als lebend und als verstorben im Datensatz enthalten. (Kassenwechsler ist verstorben).	Hinweis

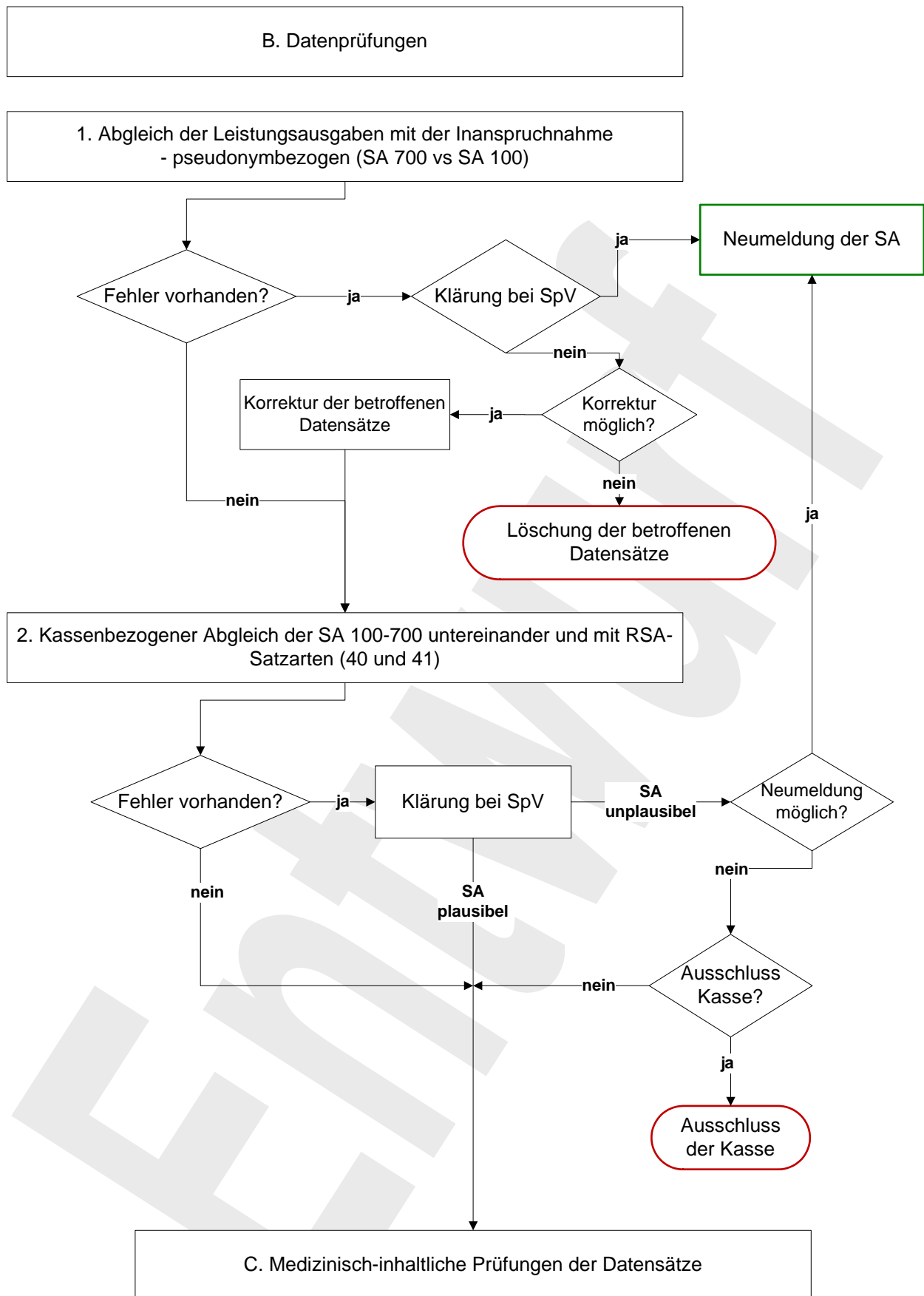


Abbildung 2: Ablauf der Datenprüfung

## Kapitel 4: Stufe C - Medizinisch-inhaltliche Prüfungen

Auf dieser Stufe werden medizinisch-inhaltlichen Prüfungen vorgenommen. Dies betrifft insbesondere die Überprüfung, ob die übermittelten Codes (ICD/PZN) im Berichtsjahr gültig sind, sowie den Abgleich mit den Versichertenstammdaten aus der SA 100. Auch in diesem Fall gilt wieder, dass die Häufung von Fehlern in diesem Bereich bei einer Krankenkasse zu einer Rücksprache des BVA mit dem zuständigen Spitzenverband führt, in der geklärt wird, ob eine Neulieferung fehlerfreier Daten möglich ist. In diesem Fall müssen die betroffenen Satzarten **vollständig neu geliefert** werden. Für den Fall, dass dies nicht möglich ist, werden die entsprechenden Datensätze gelöscht.

Bei Diagnoseangaben ist für Diagnosedaten eines Jahres die ICD-10 GM Version des jeweiligen Jahres, sowie die des Vorjahres und für die ambulanten Diagnosen die des Folgejahres zugelassen. So wären für die ambulanten Diagnosen des Jahr 2007 die Versionen ICD-10-GM Version 2006, 2007 und 2008 akzeptabel, für die stationären hingegen nur die für 2006 und 2007. Grundlage der Prüfungen sind die Listen, wie sie in den Metadateien vom Deutschen Institut für medizinische Dokumentation und Information (DIMDI) im Auftrag des BMG für das jeweilige Jahr herausgegeben wurden.<sup>8</sup> Ambulante und stationäre Diagnosen werden basierend auf den Schlüssel, die vom DIMDI gemäß §§ 295, 301 SGB V veröffentlicht werden, geprüft. Sonderzeichen („+“, „\*“, „!“, „#“) sind analog zur Vereinbarung, die von den Spitzenverbänden der Krankenkassen, der Deutschen Krankenhausgesellschaft, dem Verband der privaten Krankenversicherung und dem Institut für das Entgeltsystem im Krankenhaus (InEK) zu den Kodierrichtlinien zum § 301 SGB V getroffen wurde, weiterzugeben. Der Trennpunkt („.“) kann angegeben werden.

*Beispiele (ICD-10-GM Version 2006):*

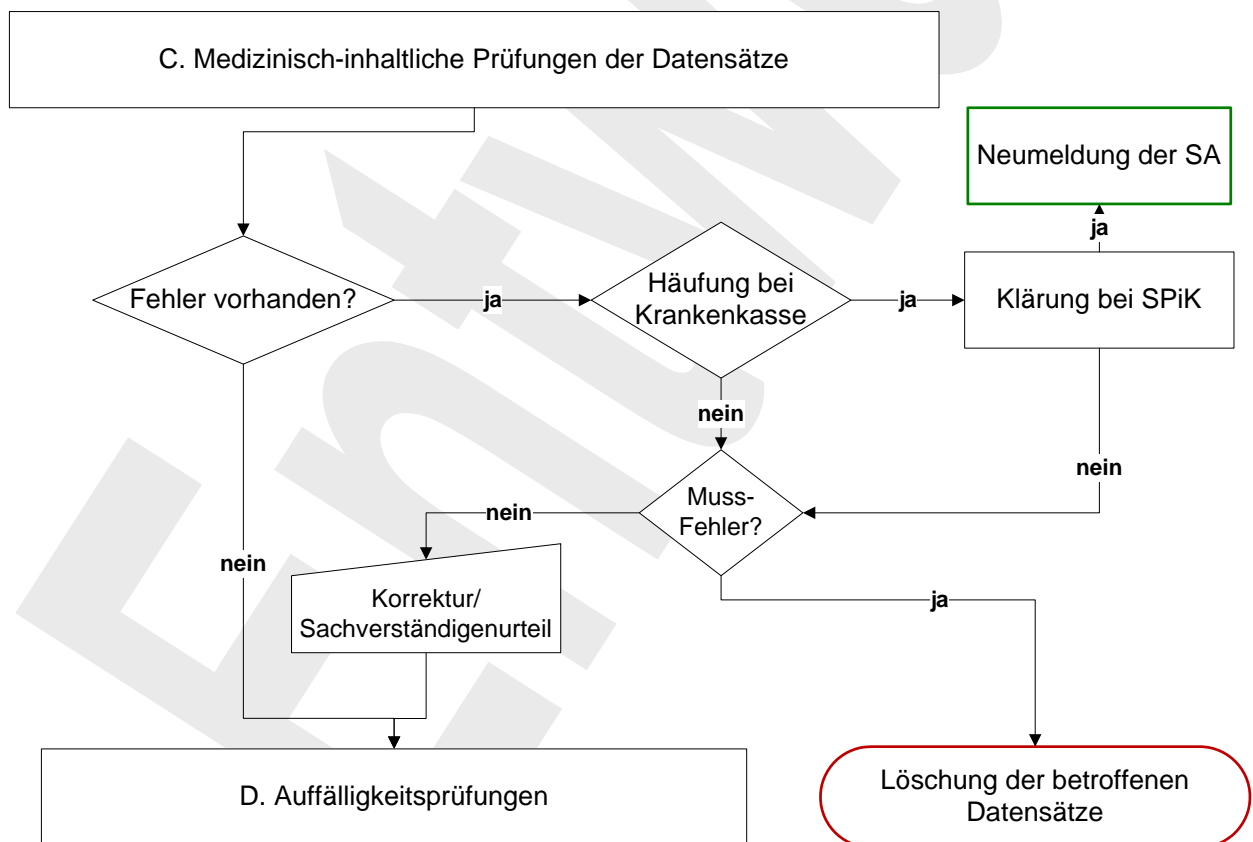
A35.	o.k.
A041	o.k.
A41.51	o.k.
A4151	o.k.
A415	für ambulante Diagnosen: o.k. für stationäre Diagnosen: Fehler - Nicht terminaler Code für Diagnose nach § 301 SGB V.
X77X	Fehler: Diagnose unbekannt.

---

<sup>8</sup> Die entsprechenden Meta-Dateien finden sich im Downloadcenter des DIMDI unter:  
<http://www.dimdi.de/dynamic/de/klasi/downloadcenter/icd-10-gm> .

## Übersicht der Fehler:

Fehler	SA	Beschreibung	Fehlerdefinition	Konsequenz
C001	400	PZN ungültig	PZN entspricht keinem gültigen Kode (entsprechend dem GKV-Arzneimittelindex des WIDO).	Löschung Datensatz
C002	500, 600	ICD ungültig	ICD entspricht keinem für den Bereich gültigen Kode (entsprechend der DIMDI-Liste)	Muss-Fehler: Löschung DS
C003	500, 600	ICD unstimmig	Der Abgleich zwischen ICD-Kode und Versichertenstammdaten ergab eine Unstimmigkeit (Kann-Fehler) für entweder Alter oder Geschlecht. (entsprechend der DIMDI-Liste)	Hinweis
C004	500, 600	ICD unstimmig	Der Abgleich zwischen ICD-Kode und Versichertenstammdaten ergab einen Muss-Fehler für entweder Alter oder Geschlecht. (entsprechend der DIMDI-Liste)	Löschung DS
C005 <sup>9</sup>	400	PZN Zulassungsfehler	Das Marktzugangsdatum der PZN ist unplausibel (entsprechend dem GKV-AI in der Version vom Dezember des auf das Berichtsjahr folgenden Jahres).	Löschung DS



**Abbildung 3: Ablauf der medizinisch-inhaltlichen Prüfungen**

<sup>9</sup> Dieser Fehler wird im Meldejahr 2008 über den Grouper geprüft, ab dem Meldejahr 2009 über die Stufe C der Plausibilisierungsprüfungen.

## Kapitel 5: Stufe D - Auffälligkeitsprüfungen

Auf der letzten Stufe werden die Daten auf sonstige Auffälligkeiten überprüft. Dazu dienen Kennzahlen, welche die Charakteristika der gelieferten Daten zusammenfassen. Dies betrifft zum einen die Prüfungen auf Pseudonymebeine, als auch einen Abgleich der gemeldeten Daten mit den im Jahresausgleich gemeldeten Daten der SA 40 und den aggregierten Meldungen des Vorjahres. Wenn Auffälligkeiten auf Kassenebene bemerkt werden, hält auch hier das BVA mit dem zuständigen Spitzenverband Rücksprache, um die Ursachen zu klären.

### Übersicht der Prüfungen:

Warnung	Beschreibung	SA	Prüfroutine
D001	Auffällige Häufung von PZN	400	Anzahl Packungen je Pseudonym, weicht vom Kassenmittelwert um mehr als drei Standardabweichungen nach oben ab.
D002	Auffällige Häufung von Diagnosen	500, 600	Die Anzahl der Diagnosen pro Pseudonym weicht vom Kassenmittelwert um mehr als drei Standardabweichungen nach oben ab.
D003	Auffällige Häufung von KH-Aufenthalten	500	Die Anzahl der Krankenhausaufenthalte (KH-Aufenthalte mit Diagnoseart=1) weicht um mehr als drei Standardabweichungen von den durchschnittlichen KH-Aufenthalten nach oben ab.
D004	Abgleich SA 100 und 40 unstimmg	100, 40	Der Anteil einer bestimmten Versicherungsgruppe (Alter, Geschlecht, Erwerbsminderungstatus, DMP-Tage) in SA 100 weicht um mehr als 5%-Punkte von dem in SA 40 gemeldeten Anteil ab.
D005	Prävalenz auffällig	600	Die Prävalenz einer Krankheit weicht um mehr als 10% Punkte von der im Vorjahr gemeldeten Krankheitsprävalenz ab.
D006	Abgleich Künstlersozialkasse unstimmg	900, KJ1	Summe der Beiträge über alle Länder der Kasse stimmt nicht mit der in der Schlüssel-Nr. 2400 der KJ1 gemeldeten Beträge überein
D007	Abgleich Beiträge freiwilliger Mitglieder unstimmg	900, KJ1	Summe der Beiträge über alle Länder der Kasse stimmt nicht mit der in der Schlüssel-Nr. 2100 und 2300 der KJ1 gemeldeten Beträge überein
D008	Abgleich Beiträge Bezieher ALG I unstimmg	900, KJ1	Summe der Beiträge über alle Länder der Kasse stimmt nicht mit der in der Schlüssel-Nr. 2011 der KJ1 gemeldeten Beträge überein
D009	Abgleich gemeldete Bezieher ALG II unstimmg	900, KM1	Summe der gemeldeten Bezieher von ALGII über alle Länder der Kasse stimmt nicht mit der in der Schlüssel-Nr. 10031 der KM1 zum 01.07. gemeldeten Bezieher überein
D010	Abgleich Beiträge versicherungspflichtiger Studenten unstimmg	900, KJ1	Summe der Beiträge über alle Länder der Kasse stimmt nicht mit der in der Schlüssel-Nr. 2040 der KJ1 gemeldeten Beträge überein.

# Anhang - Berechnung des Toleranzbereiches der Ausschöpfungsquoten laut Gutachten zur Stichprobenerhebung nach § 267 Abs. 7 Nr. 1 und 2 SGB V von Prof. Faulbaum

Ziel: Ermittlung jener Kassen, deren Ausschöpfungsquoten systematische Verzerrungen nahe legen. Dazu wird ein Toleranzbereich festgelegt. Ausschöpfungsquoten innerhalb des Toleranzbereiches werden als akzeptabel eingestuft.

Grundsätzlich gilt: Ausschöpfungsquoten außerhalb des Toleranzbereiches sollten abgelehnt werden, da solche Ausschöpfungsquoten systematische Verzerrungen nahe legen und somit die Stichprobe nicht mehr repräsentativ wäre.  $S$  kennzeichnet die Größe der Stichprobe, in diesem Fall also 25 Tage.

## 1. Ausschöpfungsquote Versicherungstage:

Variablen:

- $AQT$  - Kurzschreibweise für die Ausschöpfungsquote der Versichertentage einer nach Alter und Geschlecht differenzierten Versicherungsgruppe  $VG$  und eine Einzelkasse  $EK$  mit Stichprobengröße  $S$ :  $AQ^{VG}_{S,EK}$ .
- $\hat{V}$  - die geschätzte Varianz der Ausschöpfungsquote.
- $VZ^{VG}_{g,EK}$  - Summe der Versicherungszeiten für eine Kasse und einen bestimmten Geburtstag  $g$  der Stichprobe.
- $VZ^{VG}_{S,EK}$  - Summe der Versicherungszeiten über alle Geburtstage der Stichprobe  $S$  für eine Versichertengruppe einer Kasse.
- $VZ^{VG}_{EK}$  - Summe der Versicherungszeiten einer Kasse für eine Versichertengruppe.

Toleranzbereich:

$$AQT \in \left[ \frac{S}{365} - t_{\frac{\alpha}{2}} \sqrt{\hat{V}}, \frac{S}{365} + t_{\frac{\alpha}{2}} \sqrt{\hat{V}} \right]$$

mit

$$\hat{V} = \frac{S(1 - \frac{S}{365})}{(VZ^{VG}_{EK})^2} \cdot \frac{1}{S-1} \sum_{g \in S} (VZ^{VG}_{g,EK} - \frac{VZ^{VG}_{S,EK}}{S})^2$$

Für die Stichprobengröße ( $S=25$ ) und für  $\alpha=5\%$  ist  $t_{\alpha/2} = 2,064$ .

D.h. der Toleranzbereich entspricht dem 95%igen Konfidenzintervall der geschätzten Ausschöpfungsquote.

## 2. Ausschöpfungsquote Leistungsausgaben:

Variablen:

- $AQL$  - Kurzschreibweise für die Ausschöpfungsquote der Leistungsausgaben,
- $L^{VG}_{S,EK}$  - Summe der Leistungsausgaben der Versichertengruppe in der Stichprobe.
- $R^{VG}_{EK}$  - Schätzer für den Populations-Pro-Tag-Wert:  $R^{VG}_{EK} = \frac{L^{VG}_{S,EK}}{VZ^{VG}_{S,EK}}$

- $L_{EK}^{VG}$  - geschätzte Summe der Leistungen einer Versichertengruppe in der Population:  
 $L_{EK}^{VG} = R_{EK}^{VG} \cdot VZ_{EK}^{VG}$ .

*Toleranzbereich:*

Da die tatsächliche Summe der Versicherungsleistungen für die Population nicht bekannt ist, kann diese über  $L_{EK}^{VG}$  nur geschätzt werden. Das sich dabei ergebende Problem für die Ermittlung des Toleranzbereiches ist, dass die Varianz des Schätzers nur approximiert werden kann, da  $R_{EK}^{VG}$  ein Verhältnisschätzer ist. Im Wesentlichen entspricht der Toleranzbereich aber auch hier wieder dem Konfidenzintervall mit der gewählten Fehlerwahrscheinlichkeit  $\alpha$ :

$$L_{EK}^{VG} \in [L_{EK}^{VG} - \Delta, L_{EK}^{VG} + \Delta],$$

mit

$$\Delta = \left( t_{\frac{\alpha}{2}} + \frac{365}{S} \sqrt{\hat{V}} \right) \sqrt{AV(\hat{R}_{EK}^{VG})} \cdot VZ_{EK}^{VG} \text{ wobei die geschätzte Varianz für den Populations-Pro-}$$

Tag-Wert-Schätzer  $R_{EK}^{VG}$  wie folgt aussieht: <sup>10</sup>

$$\begin{aligned} AV(\hat{R}_{EK}^{VG}) = & \left( \frac{365}{VZ_{EK}^{VG}} \right)^2 \cdot \frac{1 - \frac{S}{365}}{S} \left( \left( \frac{1}{S-1} \sum_{g \in S} (L_{g,EK}^{VG} - \frac{L_{S,EK}^{VG}}{S})^2 \right) \right. \\ & + \left( R_{EK}^{VG} \right)^2 \cdot \left( \frac{1}{S-1} \sum_{g \in S} (VZ_{g,EK}^{VG} - \frac{VZ_{S,EK}^{VG}}{S})^2 \right) \\ & \left. - 2R_{EK}^{VG} \left( \frac{1}{S-1} \sum_{g \in S} (L_{g,EK}^{VG} - \frac{L_{S,EK}^{VG}}{S}) (VZ_{g,EK}^{VG} - \frac{VZ_{S,EK}^{VG}}{S}) \right) \right). \end{aligned}$$

<sup>10</sup> Der letzte Term des folgenden Ausdrucks entspricht der Kovarianz zwischen Leistungsausgaben und Versicherungstage.