

Berechnung von Durchschnitten und Standardfehlern unter Berücksichtigung gewichteter Daten am Beispiel der Finanzen im Zi-Praxis-Panel

Andreas Nastansky



www.zi.de

**Zentralinstitut für die
kassenärztliche Versorgung
in der Bundesrepublik Deutschland**

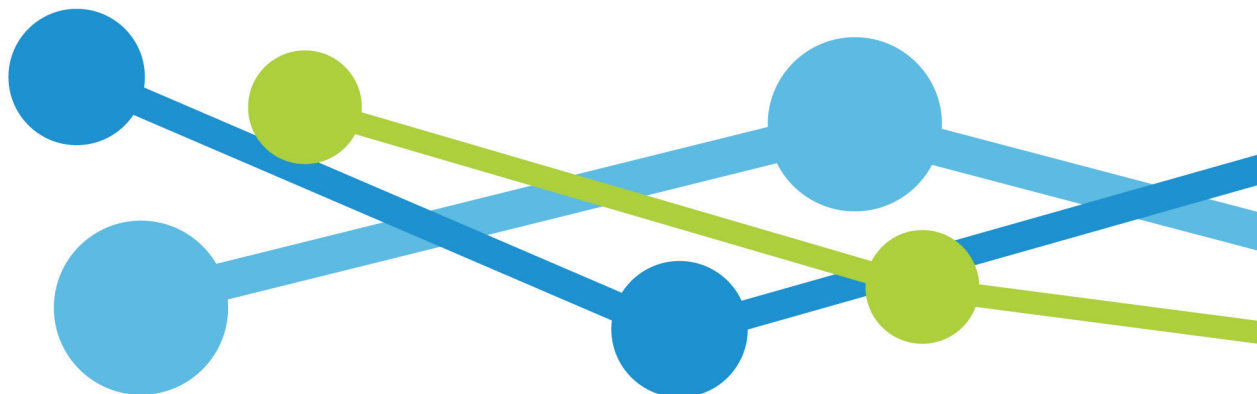
Herbert-Lewin-Platz 3
10623 Berlin

E-Mail: anastansky@zi.de

Tel. +49 30 4005 2450

Fax +49 30 4005 2490

ISSN 2199-1480 (online)



Berechnung von Durchschnitten und Standardfehlern unter Berücksichtigung gewichteter Daten am Beispiel der Finanzen im Zi-Praxis-Panel

Andreas Nastansky

Inhaltsverzeichnis

Kurzfassung.....	5
1 Einleitung.....	6
2 Standardfehler	7
3 Durchschnitt und Standardfehler bei Verhältniszahlen	8
4 Durchschnitt und Standardfehler bei Verhältniszahlen und gewichteten Daten	11
5 Durchschnitt und Standardfehler bei Nicht-Verhältniszahlen und gewichteten Daten	14
6 Finanzkennzahlen im Zi-Praxis-Panel	16
7 Fazit.....	22
Literatur	23

Kurzfassung

Verhältniszahlen wie der Jahresüberschuss je Inhaber repräsentieren im Zi-Praxis-Panel (ZiPP) zentrale Indikatoren der wirtschaftlichen Situation der Praxen in der vertragsärztlichen Versorgung. Der relative Standardfehler bildet dabei die Grundlage zur Einschätzung der Güte der berichteten Mittelwerte. Bei der Anwendung von Statistikprogrammen zur Ermittlung der Stichprobenfunktionen und deren Gütemaße muss beachtet werden, ob Verhältniszahlen und gewichtete Daten vorliegen. Beides trifft auf die Finanzkennzahlen des ZiPP zu. Aus der Nichtberücksichtigung dieser Faktoren können verzerrte Varianzschätzer und Standardfehler resultieren. Als Ausweg wird eine Normierung der Gewichte vorgestellt – mit dem Ziel, unverzerrte Stichprobenfunktionen zu erhalten. Die empirische Analyse für die Erhebung 2014 weist geringe relative Standardfehler der zentralen Finanzkennzahlen aus, die auf eine hohe Belastbarkeit der ZiPP-Daten hindeuten.

1 Einleitung

Mittelwerte wie der Durchschnitt sind statistische Maßzahlen zur Beschreibung des mittleren Niveaus bzw. der mittleren Lage der empirischen Häufigkeitsverteilung eines Merkmals.¹ Im Fall einer Stichprobenerhebung wie das Zi-Praxis-Panel (ZiPP) wird der Stichprobendurchschnitt als Schätzer für den Erwartungswert in der Grundgesamtheit herangezogen. Der Standardfehler repräsentiert dabei ein Maß für die Güte des Schätzers. In empirischen Untersuchungen wird daher in vielen Fällen der Mittelwertschätzer zusammen mit seinem Standardfehler angegeben. Der Fehler des Schätzers kommt dadurch zustande, dass bei der Berechnung des Stichprobendurchschnitts eine Zufallsstichprobe vorliegt, wobei das mögliche oder wahrscheinliche Auftreten eines Schätzfehlers das Wesensmerkmal einer Schätzung darstellt.² In diesem Zusammenhang muss berücksichtigt werden, ob der Durchschnitt über eine Verhältniszahl gebildet wurde.

Am Beispiel der zentralen Finanzkennziffer im ZiPP – dem Jahresüberschuss – wird der Unterschied deutlich. So kann der Jahresüberschuss je Praxis bzw. je Inhaber berechnet werden. Im letzteren Fall liegt eine Verhältniszahl vor. Bei der Anwendung von Statistikprogrammen zur Ermittlung der Stichprobenfunktionen und deren Gütemaße muss diesem Fakt Rechnung getragen werden. Darüber hinaus kann die Nutzung eines Gewichtungsverfahrens die Komplexität der Durchschnittsbildung und Standardfehlerberechnung weiter erhöhen. Als Lösung bietet sich eine Normierung der Gewichte an, um korrekte Stichprobenfunktionen zu erhalten.

Im Berichtswesen des Zi-Praxis-Panel werden zentrale Kennzahlen wie Durchschnitte für die Finanzdaten oder die Arbeitszeiten der Praxen sowie daraus abgeleitete Kennziffern mit Graumarkierung angegeben, wenn der relative Standardfehler der zentralen Kennzahl einen Schwellenwert von 15 Prozent überschreitet. Aufgrund des stark divergierenden Teilnahmeverhaltens auf Fachgebietsebene mit zum Teil nur geringen Fallzahlen kommt dem Standardfehler eine große Bedeutung bei der Einschätzung der Güte der berichteten Mittelwerte bei.

Im Folgenden wird die Berechnung des Standardfehlers des arithmetischen Mittels dargelegt und es werden mögliche Fehlerquellen bei der Durchschnitts- und Standardfehlerbildung bei Verhältniszahlen aufgezeigt. Anschließend werden die Probleme korrekter Standardfehler bei Anwendung eines Gewichtungsverfahrens vorgestellt und ein Lösungsweg präsentiert, der auch programmiertechnisch einfach umzusetzen ist. Dem folgt die Darstellung der Berechnung von Normgewichten auf Subgruppenebene für Verhältnis- und Nicht-Verhältniszahlen. Das Arbeitspapier endet mit einer Vergleichsrechnung am Beispiel der wesentlichen Finanzkennzahlen im ZiPP für das Berichtsjahr 2013.

1 Vgl. Rönz / Strohe (1994), S. 247.

2 Vgl. Schira (2012), S. 427.

2 Standardfehler

Der Standardfehler oder Stichprobenfehler ist ein durch die Stichprobenziehung bedingter Auswahlfehler. Statistisch repräsentiert der Standardfehler $\hat{\sigma}_{\bar{x}}$ ein Streuungsmaß für die durchschnittliche Größe des Stichprobenfehlers der Stichprobenstatistik (z.B. arithmetisches Mittel \bar{x} der Stichprobe). Bei einem erwartungstreuen Schätzer reduziert sich der mittlere quadratische Fehler auf die Varianz der Schätzfunktion.³ Damit liefert er eine Aussage über die Güte des geschätzten Parameters in der Stichprobe.

$$(1) \quad \hat{\sigma}_{\bar{x}} = \frac{\hat{\sigma}}{\sqrt{n}} \quad \text{mit} \quad \bar{x} = \frac{\sum_{i=1}^n x_i}{n} \quad \text{und} \quad \hat{\sigma} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}$$

In (1) ist der Standardfehler des arithmetischen Mittels abgebildet. Dessen Standardfehler ist üblicherweise unbekannt und muss aus der Stichprobe geschätzt werden. Als Schätzer für die unbekannte Standardabweichung der Grundgesamtheit wird die Stichprobenstandardabweichung $\hat{\sigma}$ herangezogen. Der Standardfehler hängt vom Umfang der Stichprobe n und der geschätzten Streuung des zu untersuchenden Merkmals ab. Es gilt allgemein: Je größer der Stichprobenumfang bzw. je kleiner die geschätzte Varianz des Merkmals, desto kleiner ist der Standardfehler. Die Berechnung des Standardfehlers setzt jedoch voraus, dass die Stichprobe das Ergebnis einer Zufallsauswahl ist.

Der relative Standardfehler $\hat{\sigma}_{\bar{x}}^{rel.}$ normiert den Stichprobenfehler auf den geschätzten Parameter (z.B. Stichprobenmittelwert) und erzeugt so ein dimensionsloses und damit auch für Vergleiche geeignetes Gütemaß. Der Ausdruck (2) entspricht dem Variationskoeffizienten des Stichprobendurchschnitts.

$$(2) \quad \hat{\sigma}_{\bar{x}}^{rel.} = \frac{\hat{\sigma}_{\bar{x}}}{\bar{x}} = \frac{\frac{\hat{\sigma}}{\sqrt{n}}}{\bar{x}}$$

Bei der Berechnung des (relativen) Standardfehlers wird bei einer Stichprobe ohne Zurücklegen zusätzlich eine Endlichkeitskorrektur vorgenommen.

$$(3) \quad \hat{\sigma}_{\bar{x}}^{rel.KORR} = \hat{\sigma}_{\bar{x}}^{rel.} \sqrt{\frac{N-n}{N-1}}$$

Für einen hinreichend großen Umfang der Grundgesamtheit N bzw. für einen kleinen Auswahlsatz n/N nimmt der Term der Endlichkeitskorrektur einen Wert nahe Eins an und der relative Standardfehler in (3) entspricht dem in (2). Liegt der Auswahlsatz unter 5%, so ist die Endlichkeitskorrektur in der Regel vernachlässigbar.⁴

³ Vgl. Schlittgen (2011), S. 296.

⁴ Vgl. Hartung (2009), S. 275.

3 Durchschnitt und Standardfehler bei Verhältniszahlen

Im Rahmen des ZiPP werden für unterschiedliche Merkmale Verhältniszahlen berichtet, die Auskunft über die wirtschaftliche Situation oder über Strukturmerkmale der Praxen für ein bestimmtes Berichtsjahr oder über einen Zeitraum geben. Zum Beispiel wird die Verhältniszahl Jahresüberschuss je Inhaber selbst nicht im Fragebogen erhoben, sondern aus den validierten Finanzangaben der Teilnehmer und den Angaben zur Anzahl der Inhaber gebildet. Eine Verhältniszahl ist dabei als Quotient zweier sachlogisch verbundener statistischer Zahlen definiert.⁵ Der Standardfehler dieser abgeleiteten Kennzahl wird über dessen geschätzte Varianz berechnet. Bei der Nutzung von statistischer Software wie SPSS können folgende Probleme bei der Berechnung von Mittelwerten von Verhältniszahlen und deren Standardfehlern auftreten:

- Wird die Mean-Funktion ohne Gewichtung angewendet, d.h. ein einfaches arithmetisches Mittel der Verhältniszahl berechnet, erhält eine Praxis mit einer Million Euro Jahresüberschuss und 10 Inhabern das gleiche Gewicht bei der Mittelwertbildung wie eine Praxis mit einem Inhaber und 100000 Euro Jahresüberschuss. In der Folge reflektiert ein derart gebildeter Mittelwert nicht das mittlere Niveau der empirischen Häufigkeitsverteilung des Merkmals Jahresüberschuss je Inhaber.
- Dieser ungewogene Mittelwert fließt auch in die Berechnung der Stichprobenvarianz bzw. der Stichprobenstandardabweichung und im Weiteren auch in den Standardfehler ein.

Bei der Berechnung des Durchschnitts und des (relativen) Standardfehlers von Verhältniszahlen ($X = A / B$) kann theoretisch wie folgt vorgegangen werden:

Die Ermittlung des Stichprobendurchschnitts kann über zwei Wege erfolgen: zum einen über die Zähler- und Nennersumme der Verhältniszahl X der n Praxen der Stichprobe und zum anderen über ein gewogenes arithmetisches Mittel mit der Nennergröße der i -ten statistischen Einheit als Gewichte.

$$(4) \quad \bar{x} = \frac{\sum_{i=1}^n a_i}{\sum_{i=1}^n b_i} \quad \text{bzw.} \quad \bar{x} = \frac{\sum_{i=1}^n x_i b_i}{\sum_{i=1}^n b_i}$$

Berechnung der Stichprobenstandardabweichung mit dem (gewogenen) Stichprobendurchschnitt.

$$(5) \quad \hat{\sigma} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2 b_i}{\sum_{i=1}^n b_i - 1}}$$

Berechnung des relativen Standardfehlers des Stichprobenmittelwertes der Verhältniszahl unter Berücksichtigung der gültigen Fallzahl n dieser Kennzahl.

$$(6) \quad \hat{\sigma}_{\bar{x}}^{rel.} = \frac{\hat{\sigma}_{\bar{x}}}{\bar{x}} = \frac{\hat{\sigma}}{\bar{x} \sqrt{n}}$$

⁵ Vgl. Rönz / Strohe (1994), S. 394.

Berücksichtigung der Endlichkeitskorrektur.

$$(7) \quad \hat{\sigma}_{\bar{x}}^{rel. KORR} = \hat{\sigma}_{\bar{x}}^{rel.} \sqrt{\frac{N-n}{N-1}}$$

Bei der Anwendung der Funktion gewichteter Mittelwert in vielen statistischen Programmen wird der Mittelwert korrekt berechnet, jedoch nicht die Varianz bzw. Standardabweichung und der Standardfehler. Die Ursache ist darin begründet, dass bei der Standardfehlerberechnung durch einen falschen Stichprobenumfang (Summe der Gewichte b_i) geteilt wird. In unserem Beispiel entspricht dies der Summe der Inhaber. Da in der Stichprobe Gemeinschaftspraxen enthalten sind, übersteigt die Summe der Inhaber die Anzahl der Praxen. Richtigerweise muss bei der Stichprobenvarianz jedoch durch die Anzahl der statistischen Einheiten (hier: Praxen) abzüglich Eins dividiert werden. Bei der Standardfehlerberechnung muss die Wurzel aus dem Stichprobenumfang n und nicht aus der Summe der Gewichte b_i gezogen werden.

Als Lösung bietet sich eine einfache Transformation der Gewichte an. Über eine Normierung (8) wird sichergestellt, dass die Summe der Gewichte (Inhaberzahlen = Nennergröße der Verhältniszahl) mit der Anzahl der statistischen Einheiten (hier: Anzahl der Praxen) übereinstimmt.

$$(8) \quad g = \frac{n}{\sum_{i=1}^n b_i}$$

Durch Multiplikation der praxisindividuellen Nennerwerte der Verhältniszahl mit dem Normfaktor g werden die normierten Gewichte erzeugt.

$$(9) \quad \gamma_i = gb_i$$

Diese gehen anschließend in die Berechnung des Stichprobenmittelwertes (10), der Stichprobenstandardabweichung (11) und der (relativen) Standardfehler (12 bzw. 13) wie folgt ein:

$$(10) \quad \bar{x}^{NORM} = \frac{\sum_{i=1}^n x_i \gamma_i}{\sum_{i=1}^n \gamma_i} = \frac{\sum_{i=1}^n x_i b_i}{\sum_{i=1}^n b_i} \quad \text{d.h.} \quad \bar{x}^{NORM} = \bar{x}$$

$$(11) \quad \hat{\sigma}^{NORM} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2 \gamma_i}{\sum_{i=1}^n \gamma_i - 1}}$$

$$(12) \quad \hat{\sigma}_{\bar{x}}^{rel. NORM} = \frac{\hat{\sigma}_{\bar{x}}^{NORM}}{\bar{x}} = \frac{\hat{\sigma}^{NORM}}{\bar{x} \sqrt{n}}$$

$$(13) \quad \hat{\sigma}_{\bar{x}}^{rel. NORM KORR} = \hat{\sigma}_{\bar{x}}^{rel. NORM} \sqrt{\frac{N-n}{N-1}}$$

An dieser Stelle sei darauf hingewiesen, dass für die Berechnung des arithmetischen Mittels gemäß (10) keine Normierung der Nennerwerte der Verhältniszahl vorgenommen werden muss. Die Anwendung normierter und nicht-normierter Gewichte führen zu identischen Mittelwertschätzungen, da der sowohl im Zähler als auch im Nenner auftretenden Faktor b_i herausgekürzt wird. Im Kontrast dazu resultiert aus der Schätzung der Standardabweichung mit nicht-normierten Gewichten bei Verhältniszahlen eine verzerrte Varianzschätzung des gewogenen Durchschnitts, da die Summe der Gewichte vom Stichprobenumfang abweicht. Aufgrund der Subtraktion von Eins im Nennerterm in (5) ist der Faktor b_i nicht kürzbar. Diese Korrektur dient der Herstellung der Erwartungstreue der Stichprobenvarianz.⁶

In Anlehnung an das Anwendungsbeispiel in Abschnitt 6 bedeutet dies, dass sich bei der Existenz von Gemeinschaftspraxen in der Erhebung die Anzahl der Praxisinhaber (Summe der Gewichte) von der Anzahl der Praxen (Stichprobenumfang) unterscheidet. Bei großen Stichprobenumfängen ist die Verzerrung in (5) gegenüber (11) jedoch vernachlässigbar. Auf Subgruppenebene, d.h. im ZiPP auf Fachgebieten- oder Fachgruppenebene, kann es infolge der geringen Teilnehmerzahlen in einzelnen Subgruppen hingegen zu deutlichen Abweichungen in den geschätzten Varianzen und Standardfehlern kommen.

6 Vgl. Schira (2012), S. 430f.

4 Durchschnitt und Standardfehler bei Verhältniszahlen und gewichteten Daten

Zusätzlich zur Fragestellung der Mittelwertbildung und Standardfehlerberechnung bei Verhältniszahlen tritt im Zi-Praxis-Panel das Problem der Gewichtung der Praxen nach Strukturmerkmalen der Grundgesamtheit auf. Aufgrund der Stichprobenplanung und der über die Fachgebiete differierende Rücklaufquoten kommt es zu Abweichungen in der Verteilung zentraler Strukturmerkmale (Fachgebietszusammensetzung, Praxisgröße, regionale Aspekte) zwischen den Praxen des ZiPP und den Praxen der Grundgesamtheit. Um diese Abweichungen auszugleichen, wird eine Gewichtung (Hochrechnung) der teilnehmenden Praxen vorgenommen.⁷ Damit werden die Angaben von einigen ZiPP-Praxen stärker und von anderen Praxen schwächer gewichtet. In diesem Fall wird die Hochrechnung dazu genutzt, um bestimmte Häufigkeiten als Eckwerte an die Struktur der Grundgesamtheit anzupassen.⁸ Dieses Vorgehen wirkt sich nicht nur auf die Häufigkeitsverteilung, sondern auch auf beschreibende statistische Kennzahlen wie Durchschnitte, Streuungsmaße und Gütekennzahlen der Schätzer aus. Ein Beispiel für eine derartige Kennzahl im ZiPP repräsentiert der (gewichtete) Durchschnitt des Jahresüberschusses je Inhaber.

Bei der Mittelwertbildung von Verhältniszahlen unter Einbeziehung von Gewichten einer Hochrechnung kann in Anlehnung an Abschnitt 3 folgendes Vorgehen gewählt werden:

Berechnung des (gewichteten) Stichprobendurchschnitts als gewogenes arithmetisches Mittel mit den Nennergrößen der Verhältniszahl und den Gewichtungsfaktoren der Strukturgewichtung als Gewichte. Der praxisindividuelle Faktor w_i ergibt sich dabei aus dem Gewichtungsverfahren.

$$(14) \quad \bar{x}^G = \frac{\sum_{i=1}^n x_i b_i w_i}{\sum_{i=1}^n b_i w_i}$$

Berechnung der (gewichteten) Stichprobenstandardabweichung mit dem (gewichteten) Stichprobendurchschnitt.

$$(15) \quad \hat{\sigma}^G = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x}^G)^2 b_i w_i}{\sum_{i=1}^n b_i w_i - 1}}$$

Berechnung des relativen Standardfehlers des (gewichteten) Stichprobenmittelwertes der Verhältniszahl unter Berücksichtigung der gültigen Fallzahl n dieser Kennzahl.

$$(16) \quad \hat{\sigma}_{\bar{x}^G}^{rel.} = \frac{\hat{\sigma}_{\bar{x}^G}}{\bar{x}^G} = \frac{\hat{\sigma}^G}{\bar{x}^G} \frac{1}{\sqrt{n}}$$

Berücksichtigung der Endlichkeitskorrektur.

$$(17) \quad \hat{\sigma}_{\bar{x}^G}^{rel.KORR} = \hat{\sigma}_{\bar{x}^G}^{rel.} \sqrt{\frac{N-n}{N-1}}$$

⁷ Vgl. Zentralinstitut für die kassenärztliche Versorgung in Deutschland (2016), S. 106f.

⁸ Vgl. von der Lippe (2012), S. 20.

Erneut wird zwar das arithmetische Mittel korrekt berechnet, jedoch nicht die Varianz bzw. Standardabweichung und daraus abgeleitet der (relative) Standardfehler. Zum einen wird bei der Bestimmung der Stichprobenstandardabweichung durch ein Summenprodukt $w_i b_i$ geteilt, das nicht dem Stichprobenumfang entspricht und zu einem verzerrten Varianzschätzer in der Stichprobe führt. Zum anderen tritt bei der Anwendung von Funktionen gewichteter Mittelwert in Programmen wie SPSS bei der Standardfehlerberechnung das Problem auf, dass durch einen falschen Stichprobenumfang (Summe der Gewichte $w_i b_i$) dividiert wird. In unserem Beispiel entspricht dies der Summe der gewichteten Inhaberzahlen. Richtigerweise muss jedoch durch die Anzahl der statistischen Einheiten (Praxen) bzw. dessen Wurzel dividiert werden.

Als Lösung bietet sich erneut eine Normierung der Gewichte an. Über (18) wird sichergestellt, dass die Summe der Gewichte (Produkt aus der Nennergröße der Verhältniszahl, b_i , und den Gewichtungsfaktoren aus der Strukturgewichtung, w_i) mit der Anzahl der statistischen Einheiten übereinstimmt.

$$(18) \quad g^* = \frac{n}{\sum_{i=1}^n b_i w_i}$$

Durch Multiplikation der praxisindividuellen Nennerwerte der Verhältniszahl und den Gewichtungsfaktoren w_i mit dem Normfaktor g^* werden die normierten Gewichte erzeugt.

$$(19) \quad \gamma_i^* = g^* b_i w_i$$

Diese gehen anschließend in die Berechnung des Stichprobenmittelwertes (20), der Stichprobenstandardabweichung (21) und der (relativen) Standardfehler (22 bzw. 23) wie folgt ein:

$$(20) \quad \bar{x}^{G \text{ NORM}} = \frac{\sum_{i=1}^n x_i \gamma_i^*}{\sum_{i=1}^n \gamma_i^*} = \frac{\sum_{i=1}^n x_i b_i w_i}{\sum_{i=1}^n b_i w_i} \quad \text{d.h.} \quad \bar{x}^{G \text{ NORM}} = \bar{x}^G$$

$$(21) \quad \hat{\sigma}^{G \text{ NORM}} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x}^G)^2 \gamma_i^*}{\sum_{i=1}^n \gamma_i^* - 1}}$$

$$(22) \quad \hat{\sigma}_{\bar{x}^G}^{G \text{ rel. NORM}} = \frac{\hat{\sigma}_{\bar{x}^G}^{G \text{ NORM}}}{\bar{x}^G} = \frac{\hat{\sigma}^{G \text{ NORM}}}{\bar{x}^G \sqrt{n}}$$

$$(23) \quad \hat{\sigma}_{\bar{x}^G}^{G \text{ rel. NORM KORR}} = \hat{\sigma}_{\bar{x}^G}^{G \text{ rel. NORM}} \sqrt{\frac{N-n}{N-1}}$$

Wenn die Summe der Gewichte aus dem Strukturgewichtungsverfahren dem Stichprobenumfang entspricht, d.h. es gilt $\sum_{i=1}^n w_i = n$, reduziert sich das Problem der korrekten Standardfehlerberechnung auf den Fall im Abschnitt 3. Werden hingegen Mittelwerte und Standardfehler auf Subgruppenebene ausgewiesen, ist das Vorgehen (20) bis (23) zu befolgen.

Auf Fachgebietsebene weicht im Zi-Praxis-Panel die Summe der Gewichte der Praxen eines Fachgebietes vom Stichprobenumfang (Anzahl der Praxen) des Fachgebietes n_k ab, da eine Hoch- bzw. Untergewichtung von Praxen bestimmter Fachgebiete im ZiPP erfolgt. Die Gleichungen (18) bis (23) müssen für eine Berechnung auf Subgruppenebene wie folgt angepasst werden:

$$(24) \quad g_k^* = \frac{n_k}{\sum_{i=1}^n w_i^{(k)} b_i}$$

Durch (24) wird sichergestellt, dass die Summe der Gewichte (Produkt aus Nennergröße der Verhältniszahl und Strukturgewicht) auf Subgruppenebene mit der Anzahl der statistischen Einheiten (hier: Anzahl der Praxen des Fachgebiets) übereinstimmt.

Mittels der Multiplikation der praxisindividuellen Nennerwerte der Verhältniszahl b_i und den Gewichtungsfaktoren w_i mit den k Normfaktoren g_k^* werden die normierten Gewichte auf Praxisebene erzeugt.

$$(25) \quad \gamma_i^* = g_k^* b_i w_i$$

Diese gehen anschließend in die Berechnung des Stichprobenmittelwertes (26), der Stichprobenstandardabweichung (27) und der (relativen) Standardfehler (28 bzw. 29) auf Subgruppenebene wie folgt ein:

$$(26) \quad \bar{x}_k^{G \text{ NORM}} = \frac{\sum_{i=1}^{n_k} x_i \gamma_i^*}{\sum_{i=1}^{n_k} \gamma_i^*} = \frac{\sum_{i=1}^{n_k} x_i g_k^* b_i w_i}{\sum_{i=1}^{n_k} g_k^* b_i w_i} = \frac{\sum_{i=1}^{n_k} x_i b_i w_i}{\sum_{i=1}^{n_k} b_i w_i} \quad \text{d.h.} \quad \bar{x}_k^{G \text{ NORM}} = \bar{x}_k^G$$

$$(27) \quad \hat{\sigma}_k^{G \text{ NORM}} = \sqrt{\frac{\sum_{i=1}^{n_k} (x_i - \bar{x}_k^G)^2 \gamma_i^*}{\sum_{i=1}^{n_k} \gamma_i^* - 1}}$$

$$(28) \quad \hat{\sigma}_{\bar{x}_k^G}^{G \text{ rel. NORM}} = \frac{\hat{\sigma}_k^{G \text{ NORM}}}{\bar{x}_k^G} = \frac{\hat{\sigma}_k^{G \text{ NORM}}}{\sqrt{n_k}}$$

$$(29) \quad \hat{\sigma}_{\bar{x}_k^G}^{G \text{ rel. NORM KORR}} = \hat{\sigma}_{\bar{x}_k^G}^{G \text{ rel. NORM}} \sqrt{\frac{N_k - n_k}{N_k - 1}}$$

Wie (26) darlegt, stimmen die normierten gewichteten Durchschnitte auf Subgruppenebene (hier: Fachgebietsebene) mit den nicht-normierten gewichteten Mittelwerten überein. Demgegenüber würden aus einem Verzicht der Normierung der Gewichte auf Fachgebietsebene bei Verhältniszahlen wie dem Jahresüberschuss je Inhaber verzerrte Varianzschätzer und Standardfehler resultieren. Aufgrund der Normierung entsprechen auf Fachgebietsebene die Summe der normierten Gewichte der Anzahl Praxen des jeweiligen Fachgebiets, d.h. es gilt:

$$\sum_{i=1}^{n_k} \gamma_i^{*(k)} = n_k.$$

5 Durchschnitt und Standardfehler bei Nicht-Verhältniszahlen und gewichteten Daten

Als Folge des Gewichtungsverfahrens im ZiPP tritt die Problematik der fehlerhaften Standardfehlerberechnung auf Subgruppenebene auch bei Nicht-Verhältniszahlen auf, da durch die Anzahl der Summengewichte auf der Subgruppenebene (hier: Fachgebiete) geteilt wird. Diese weichen aber von der Anzahl der statistischen Einheiten der jeweiligen Subgruppe ab. Bei der Ermittlung des Standardfehlers des mittleren Jahresüberschusses je Praxis auf Fachgebietsebene müssen ebenso normierte Gewichte verwendet werden. Im Gegensatz zum Vorgehen bei Verhältniszahlen wird auf die Gewichtung mit der Nennergröße b_i verzichtet, sodass bei der Berechnung der Standardfehler wie folgt vorgegangen werden kann:

Berechnung des (gewichteten) Stichprobendurchschnitts als gewogenes arithmetisches Mittel mit den Gewichtungsfaktoren der Strukturgewichtung als Gewichte. Der praxisindividuelle Faktor w_i ergibt sich dabei aus dem Gewichtungsverfahren.

$$(30) \quad \bar{x}^W = \frac{\sum_{i=1}^n x_i w_i}{\sum_{i=1}^n w_i}$$

Berechnung der (gewichteten) Stichprobenstandardabweichung mit dem (gewichteten) Stichprobendurchschnitt.

$$(31) \quad \hat{\sigma}^W = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x}^W)^2 w_i}{\sum_{i=1}^n w_i - 1}}$$

Berechnung des relativen Standardfehlers des Stichprobenmittelwertes der Nicht-Verhältniszahl unter Berücksichtigung der gültigen Fallzahl n dieser Kennzahl.

$$(32) \quad \hat{\sigma}_{\bar{x}^W}^{rel.} = \frac{\hat{\sigma}_{\bar{x}^W}}{\bar{x}^W} = \frac{\hat{\sigma}^W}{\bar{x}^W \sqrt{n}}$$

Berücksichtigung der Endlichkeitskorrektur.

$$(33) \quad \hat{\sigma}_{\bar{x}^W}^{rel. KORR} = \hat{\sigma}_{\bar{x}^W}^{rel.} \sqrt{\frac{N-n}{N-1}}$$

Das Vorgehen (30) bis (33) führt zu unverzerrten Stichprobenvarianzen und Standardfehlern, sofern die Summe der Gewichte w_i aus dem Strukturgewichtungsverfahren sich zum Stichprobenumfang n ergibt. Soll zum Beispiel der Jahresüberschuss je Praxis über alle Praxen im ZiPP berechnet werden, können die nicht-normierten Gewichte des Gewichtungsverfahrens genutzt werden. Im Gegensatz dazu bedarf es einer Normierung der Gewichte, wenn Mittelwerte und deren Gütemaße auf Subgruppenebene (z.B. Fachgebietsebene) bestimmt werden. Durch (34) wird sichergestellt, dass die Summe der Gewichte auf Subgruppenebene mit der Anzahl der statistischen Einheiten (hier: Anzahl der Praxen des Fachgebiets) übereinstimmt.

$$(34) \quad g'_k = \frac{n_k}{\sum_{i=1}^n w_i^{(k)}}$$

Dabei gilt auf Subgruppenebene üblicherweise: $\sum_{i=1}^{n_k} w_i^{(k)} \neq n_k$.

Durch Multiplikation der praxisindividuellen Gewichte mit den k Normfaktoren g_k werden die normierten Gewichte erzeugt.

$$(35) \quad \gamma'_i = g'_k w_i$$

Diese gehen anschließend in die Berechnung des Stichprobenmittelwertes (36), der Stichprobenstandardabweichung (37) und der (relativen) Standardfehler (38 bzw. 39) auf Subgruppenebene wie folgt ein:

$$(36) \quad \bar{x}_k^{W \text{ NORM}} = \frac{\sum_{i=1}^{n_k} x_i \gamma'_i}{\sum_{i=1}^{n_k} \gamma'_i} = \frac{\sum_{i=1}^{n_k} x_i g'_k w_i}{\sum_{i=1}^{n_k} g'_k w_i} = \frac{\sum_{i=1}^{n_k} x_i w_i}{\sum_{i=1}^{n_k} w_i} \quad \text{d.h.} \quad \bar{x}_k^{W \text{ NORM}} = \bar{x}_k^W$$

$$(37) \quad \hat{\sigma}_k^{W \text{ NORM}} = \sqrt{\frac{\sum_{i=1}^{n_k} (x_i - \bar{x}_k^W)^2 \gamma'_i}{\sum_{i=1}^{n_k} \gamma'_i - 1}}$$

$$(38) \quad \hat{\sigma}_{\bar{x}_k^W}^{W \text{ rel. NORM}} = \frac{\hat{\sigma}_{\bar{x}_k^W}^{W \text{ NORM}}}{\bar{x}_k^W} = \frac{\hat{\sigma}_k^{W \text{ NORM}}}{\bar{x}_k^W \sqrt{n_k}}$$

$$(39) \quad \hat{\sigma}_{\bar{x}_k^W}^{W \text{ rel. NORM KORR}} = \hat{\sigma}_{\bar{x}_k^W}^{W \text{ rel. NORM}} \sqrt{\frac{N_k - n_k}{N_k - 1}}$$

Wie (36) zeigt, stimmen die normierten gewichteten Durchschnitte auf Subgruppenebene mit den nicht-normierten gewichteten Mittelwerten überein. Im Gegensatz dazu würden aus einem Verzicht der Normierung der Gewichte auf Fachgebietsebene bei Nicht-Verhältniszahlen wie dem Jahresüberschuss je Praxis verzerrte Varianzschätzer und Standardfehler resultieren.

6 Finanzkennzahlen im Zi-Praxis-Panel

Anhand der zentralen Finanzkennzahlen im Zi-Praxis-Panel – Einnahmen, Aufwendungen und Jahresüberschuss – wird die Problematik der Standardfehlerberechnung beispielhaft demonstriert. Wie in den vorangegangenen Abschnitten dargelegt wurde, resultiert die Problemstellung aus den folgenden zwei Faktoren: Berechnung von Durchschnitten von Verhältniszahlen und Anwendung eines Strukturgewichtungsverfahrens.

Die heterogenen Strukturen in der vertragsärztlichen Versorgung mit dem Vorhandensein von Gemeinschaftspraxen, d.h. mehr als ein Inhaber ist pro Praxis tätig, trägt dazu bei, dass die Finanzkennzahlen je Inhaber (Verhältniszahl) vom den entsprechenden Werten auf Praxisebene abweichen. So lag der Anteil der ZiPP-Praxen im Längsschnitt in der Erhebung 2014, die als Gemeinschaftspraxis organisiert waren, bei rund 20%.⁹ Im Falle von ausschließlich Einzelpraxen würden identische Finanzkennzahlen auf Praxis- und Inhaberebene ausgewiesen werden.

Aufgrund der Zielsetzung, verlässliche Aussagen über die Entwicklung der Wirtschaftslage in möglichst allen Fachgebieten mit vertretbarem Aufwand tätigen zu können, wird eine geschichtete Stichprobe gezogen.¹⁰ Kleine Fachgebiete sind deshalb über einen höheren Auswahlatz bei der Stichprobe in der Regel überproportional im ZiPP vertreten, große Fachgebiete hingegen unterproportional. Um auf Basis dieser Daten Aussagen über die Entwicklung der vertragsärztlichen Versorgung in Deutschland insgesamt treffen zu können, werden die Angaben aus der Erhebung mithilfe bundesweiter Abrechnungsdaten nach Fachgebietszugehörigkeit, KV-Umsatzklassen und Regionstypen gewichtet.¹¹

Die Hochrechnung im Rahmen des Strukturgewichtungsverfahrens führt im Zi-Praxis-Panel zu einer „Höhergewichtung“ der Praxen – speziell der Fachgebiete Allgemeinmedizin und Innere Medizin (hausärztlich) und der Psychotherapie. Dies ergibt sich aus dem Umstand, dass für diese Fachgebiete Stichproben mit einem geringen Auswahlatz gezogen werden. Als Folge ist die Anzahl der an der Stichprobe teilnehmenden Praxen sehr viel kleiner als die Anzahl der gewogenen Praxen nach der Hochrechnung.¹² Entsprechend verringert sich das Gewicht von Praxen in Fachgebieten, die im ZiPP „überrepräsentiert“ sind. Zu diesen zählen u.a. die Chirurgie und die Urologie.

In Tabelle 1 werden für die zentralen Finanzkennzahlen mit Inhaberbezug des Berichtsjahres 2013 der Längsschnittstichprobe die normierten (22) den nicht-normierten (16) relativen Standardfehlern gegenübergestellt. Mit einem Niveau von jeweils unter zwei Prozent bzw. für den Jahresüberschuss von unter einem Prozent zeichnen sich die Mittelwertschätzer der ZiPP-Finanzdaten auf Gesamtebene durch eine hohe statistische Güte aus. Über alle Praxen liegen die normierten Standardfehler in einer Spanne von 0,1 bis 0,16 Prozentpunkten geringfügig über den nicht-normierten Werten. Dementsprechend nimmt die statistische Güte der geschätzten Mittelwerte geringfügig ab. Dies ist auf u.a. darauf zurückzuführen, dass die Summe der Nennergewichte (hier: Inhaberzahlen) größer ist als die Anzahl der Praxen. Im Mittel sind 1,25 Inhaber in den 4701 Praxen der Längsschnittauswahl tätig. Über alle Praxen entspricht die Summe der nicht-normierten Gewichte der Strukturgewichtung mit der Anzahl der Praxen (Stichprobenumfang) überein. Auf Subgruppenebene (Fachgebietsebene) gilt dies nicht.

Aus Tabelle 1 ist deutlich zu entnehmen, dass nur in einem Fachgebiet (Nuklearmedizin) die relativen Standardfehler der Durchschnitte der Finanzkennzahlen je Inhaber einen Wert von 15% überschreiten. Hier wirkt sich primär der geringe Stichprobenumfang ($n=8$) aus. Demgegenüber zeichnen sich die Stichprobenmittelwerte stark besetzter Fachgebiete wie der Allgemeinmedizin und der Psychotherapie durch niedrigere Standardfehler aus.

9 Vgl. Zentralinstitut für die kassenärztliche Versorgung in Deutschland (2016), S. 131.

10 Vgl. Zentralinstitut für die kassenärztliche Versorgung in Deutschland (2016), S. 101f.

11 Vgl. Zentralinstitut für die kassenärztliche Versorgung in Deutschland (2016), S. 106f.

12 Vgl. von der Lippe (2012), S. 21.

Tabelle 1: Vergleich der normierten und nicht-normierten relativen Standardfehler der Finanzen je Inhaber für das Berichtsjahr 2013

Fachgebiet	Relativer Standardfehler				Differenz zwischen normierter und nicht-normierter Berechnung				
	Einnahmen je Inhaber		Aufwendungen je Inhaber		Jahresüberschuss je Inhaber		absolute Differenz in %-Punkten		Relative Differenz in %
	normiert	nicht normiert	normiert	nicht normiert	normiert	nicht normiert	Einnahmen	Aufwendungen	
Gesamt	1,08%	0,96%	1,48%	1,32%	0,95%	0,85%	0,12	0,16	12,00%
Allgemeinmedizin und Innere Medizin (hausärztlich)	1,62%	0,90%	2,00%	1,12%	1,85%	1,03%	0,72	0,88	79,14%
Anästhesiologie	5,29%	4,86%	7,92%	7,29%	4,67%	4,30%	0,43	0,64	8,75%
Augenheilkunde	4,88%	5,51%	6,06%	6,84%	4,89%	5,52%	-0,63	-0,78	11,36%
Chirurgie	3,35%	4,41%	3,81%	5,01%	3,63%	4,78%	-1,06	-1,20	23,94%
Dermatologie	3,60%	4,56%	3,90%	4,94%	4,08%	5,16%	-0,96	-1,04	21,02%
Gynäkologie	2,04%	2,20%	2,49%	2,69%	2,37%	2,56%	-0,16	-0,19	7,23%
Hals-Nasen-Ohren-Heilkunde	2,49%	3,16%	2,57%	3,26%	3,34%	4,24%	-0,67	-0,69	21,15%
Innere Medizin - Gastroenterologie	6,39%	6,25%	6,01%	5,89%	9,66%	9,46%	0,13	0,12	2,11%
Innere Medizin - Kardiologie	6,59%	5,28%	8,06%	6,46%	7,38%	5,91%	1,31	1,60	24,79%
Innere Medizin - Pneumologie	6,11%	7,04%	7,39%	8,52%	7,83%	9,03%	-0,93	-1,13	13,26%
Innere Medizin - ohne bzw. mit mehreren Schwerpunkten	6,09%	6,63%	7,82%	8,51%	6,11%	6,65%	-0,54	-0,69	8,14%
Innere Medizin - sonstige Fachgebiete	8,49%	5,97%	11,42%	8,03%	5,54%	3,90%	2,52	3,39	42,22%
Kinder- und Jugendmedizin	1,76%	2,15%	2,22%	2,73%	2,06%	2,53%	-0,40	-0,50	18,47%
Kinder- und Jugendpsychiatrie und psychotherapie	6,51%	11,69%	7,93%	14,24%	5,97%	10,73%	-5,18	-6,32	44,34%
Nervenheilkunde	3,60%	5,04%	4,18%	5,85%	3,95%	5,54%	-1,44	-1,67	28,56%
Neurologie	5,95%	7,76%	6,57%	8,56%	7,82%	10,20%	-1,81	-1,99	23,29%
Neurologie, Nervenheilkunde, Psychiatrie	4,83%	4,33%	8,17%	7,33%	4,43%	3,97%	0,50	0,85	11,54%
Nuklearmedizin	18,20%	14,80%	20,20%	16,43%	21,84%	17,77%	3,39	3,77	22,91%
Orthopädie	2,92%	3,49%	3,04%	3,64%	3,83%	4,58%	-0,57	-0,60	16,42%
Physikalische und rehabilitative Medizin	7,93%	12,74%	8,79%	14,12%	9,24%	14,85%	-4,81	-5,33	37,77%
Psychiatrie	4,28%	5,65%	5,57%	7,36%	4,25%	5,61%	-1,37	-1,79	24,30%
Psychosomatische Medizin und Psychotherapie	1,96%	3,54%	3,11%	5,61%	2,16%	3,89%	-1,58	-2,50	44,61%
Psychotherapie	2,48%	1,47%	3,25%	1,93%	2,76%	1,64%	1,01	1,32	68,58%
Radiologie	6,86%	4,26%	8,07%	5,02%	9,21%	5,73%	2,59	3,05	60,82%
Urologie	3,34%	3,69%	3,42%	3,78%	4,36%	4,82%	-0,35	-0,36	9,47%
Übergreifend tätige Praxen	8,75%	5,14%	9,74%	5,73%	9,29%	5,46%	3,60	4,01	70,00%

Quelle: Zi-Praxis-Panel 2014.

Auf Fachgebietsebene beeinflusst das Strukturgewichtungsverfahren das Vorzeichen der Differenz. Praxen von Fachgebieten, die stark höhergewichtet werden wie die Allgemeinmedizin oder die Übergreifend tätigen Praxen, weisen mit 0,72 bzw. 3,6%-Punkten eine positive Vorzeichendifferenz auf, wobei bereinigt um das Niveau des relativen Standardfehlers sich die prozentuale Abweichung in einer ähnlichen Größenordnung bewegt (rd. 79% zu 70%). In beiden Fällen nimmt die gewichtete Summe der Praxen größere Werte an als die ungewichtete Summe (vgl. Tabelle 2), was nach (1) zu den niedrigeren nicht-normierten Standardfehlern führt. Im Kontrast dazu nehmen Praxen des Fachgebietes Kinder- und Jugendpsychiatrie und -psychotherapie überproportional häufig am ZiPP teil, sodass im Rahmen der Hochrechnung eine Untergewichtung vorgenommen wird. Als Ergebnis erhalten Praxen dieser Subgruppe durchschnittlich ein Gewicht von 0,28 – mit der Folge, dass der normierte Standardfehler der Praxisfinanzkennzahlen merklich unter dem nicht-normierten liegt.

Praxen von Fachgebieten mit einem hohen Anteil an Gemeinschaftspraxen (54%) und höheren mittleren Inhaberzahlen je Praxis (2,4) wie die Radiologie werden ebenso tendenziell größere Werte im normierten Gütemaß der Finanzmittelwerte annehmen, da – unter Nicht-Beachtung der Strukturgewichtung – die gewichtete Summe die ungewichtete Summe (Anzahl

Tabelle 2: Summen der Strukturgewichte und Inhaberzahlen nach Fachgebieten

Fachgebiet	Summen			
	Anzahl Praxen	$\sum w_i \cdot b_i$	$\sum w_i$	$\sum b_i$
Gesamt	4701	5896,82	4701,00	5911
Allgemeinmedizin und Innere Medizin (hausärztlich)	672	2154,26	1688,82	850
Anästhesiologie	105	123,99	75,70	184
Augenheilkunde	259	203,70	161,94	312
Chirurgie	223	129,43	93,75	301
Dermatologie	225	140,73	113,52	278
Gynäkologie	509	438,15	357,23	619
Hals-Nasen-Ohren-Heilkunde	307	191,25	142,83	402
Innere Medizin - Gastroenterologie	35	36,45	28,25	43
Innere Medizin - Kardiologie	42	64,84	52,22	50
Innere Medizin - Pneumologie	49	37,11	29,34	61
Innere Medizin - ohne bzw. mit mehreren Schwerpunkten	47	39,82	30,50	59
Innere Medizin - sonstige Fachgebiete	42	83,93	47,16	58
Kinder- und Jugendmedizin	384	255,57	193,60	508
Kinder- und Jugendpsychiatrie und -psychotherapie	112	35,39	31,29	130
Nervenheilkunde	151	77,55	68,28	168
Neurologie	40	23,95	21,45	47
Neurologie, Nervenheilkunde, Psychiatrie	68	84,35	47,88	112
Nuklearmedizin	8	11,58	7,90	12
Orthopädie	296	207,06	149,68	396
Physikalische und rehabilitative Medizin	35	14,17	13,50	37
Psychiatrie	110	63,47	57,66	123
Psychosomatische Medizin und Psychotherapie	387	119,44	114,54	398
Psychotherapie	350	992,78	977,42	358
Radiologie	39	99,28	35,84	92
Urologie	157	128,86	91,09	222
Übergreifend tätige Praxen	49	139,72	69,63	91

Quelle: Zi-Praxis-Panel 2014.

Tabelle 3: Vergleich der normierten relativen Standardfehler ohne bzw. mit Endlichkeitskorrektur der Finanzen je Inhaber für das Berichtsjahr 2013

Fachgebiet	Relativer Standardfehler				Differenz zwischen normiert und normiert mit Endlichkeitskorrektur					
	Einnahmen je Inhaber		Aufwendungen je Inhaber		Jahresüberschuss je Inhaber		absolute Differenz in %-Punkten		Relative Differenz in %	
	normiert	normiert mit Korrektur	normiert	normiert mit Korrektur	normiert	normiert mit Korrektur	Einnahmen	Aufwendungen		Jahresüberschuss
Gesamt	1,08%	1,05%	1,48%	1,45%	0,95%	0,93%	0,03	0,04	0,02	2,41%
Allgemeinmedizin und Innere Medizin hausärztlich)	1,62%	1,60%	2,00%	1,98%	1,85%	1,83%	0,02	0,02	0,02	0,95%
Anästhesiologie	5,29%	5,10%	7,92%	7,64%	4,67%	4,51%	0,19	0,28	0,17	3,54%
Augenheilkunde	4,88%	4,69%	6,06%	5,82%	4,89%	4,70%	0,19	0,24	0,19	3,98%
Chirurgie	3,35%	3,16%	3,81%	3,59%	3,63%	3,42%	0,20	0,22	0,21	5,84%
Dermatologie	3,60%	3,43%	3,90%	3,72%	4,08%	3,88%	0,17	0,19	0,20	4,81%
Gynäkologie	2,04%	1,97%	2,49%	2,41%	2,37%	2,29%	0,07	0,09	0,08	3,49%
Hals-Nasen-Ohren-Heilkunde	2,49%	2,36%	2,57%	2,43%	3,34%	3,16%	0,14	0,14	0,18	5,51%
Innere Medizin - Gastroenterologie	6,39%	6,17%	6,01%	5,81%	9,66%	9,33%	0,22	0,20	0,33	3,40%
Innere Medizin - Kardiologie	6,59%	6,44%	8,06%	7,87%	7,38%	7,21%	0,15	0,18	0,17	2,28%
Innere Medizin - Pneumologie	6,11%	5,85%	7,39%	7,09%	7,83%	7,51%	0,25	0,30	0,32	4,10%
Innere Medizin - ohne bzw. mit mehreren Schwerpunkten	6,09%	5,91%	7,82%	7,59%	6,11%	5,93%	0,18	0,23	0,18	2,94%
Innere Medizin - sonstige Fachgebiete	8,49%	8,31%	11,42%	11,17%	5,54%	5,42%	0,19	0,25	0,12	2,20%
Kinder- und Jugendmedizin	1,76%	1,67%	2,22%	2,11%	2,06%	1,96%	0,09	0,11	0,10	5,00%
Kinder- und Jugendpsychiatrie und psychotherapie	6,51%	5,91%	7,93%	7,21%	5,97%	5,43%	0,59	0,72	0,54	9,10%
Nervenheilkunde	3,60%	3,40%	4,18%	3,95%	3,95%	3,74%	0,20	0,23	0,22	5,53%
Neurologie	5,95%	5,72%	6,57%	6,31%	7,82%	7,52%	0,23	0,25	0,30	3,88%
Neurologie, Nervenheilkunde, Psychiatrie	4,83%	0,00%	8,17%	0,00%	4,43%	0,00%				
Nuklearmedizin	18,20%	17,80%	20,20%	19,76%	21,84%	21,37%	0,39	0,44	0,47	2,16%
Orthopädie	2,92%	2,77%	3,04%	2,89%	3,83%	3,64%	0,14	0,15	0,19	4,88%
Physikalische und rehabilitative Medizin	7,93%	7,44%	8,79%	8,24%	9,24%	8,66%	0,50	0,55	0,58	6,24%
Psychiatrie	4,28%	4,07%	5,57%	5,30%	4,25%	4,04%	0,21	0,28	0,21	4,99%
Psychosomatische Medizin und Psychotherapie	1,96%	1,78%	3,11%	2,83%	2,16%	1,96%	0,18	0,28	0,20	9,07%
Psychotherapie	2,48%	2,46%	3,25%	3,22%	2,76%	2,73%	0,02	0,03	0,02	0,86%
Radiologie	6,86%	6,57%	8,07%	7,74%	9,21%	8,83%	0,28	0,33	0,38	4,11%
Urologie	3,34%	3,20%	3,42%	3,28%	4,36%	4,17%	0,14	0,15	0,19	4,29%
Übergreifend tätige Praxen	8,75%	8,69%	9,74%	9,68%	9,29%	9,22%	0,06	0,07	0,06	0,69%

Quelle: Zi-Praxis-Panel 2014.

der Praxen) übersteigt. Da in einer Vertragsarztpraxis im ZiPP mindestens ein Inhaber pro Praxis tätig sein muss, verfügen alle Fachgebiete über eine durchschnittliche Inhaberzahl von mehr als Eins.

Ferner wird sichtbar, dass das Fachgebiet Innere Medizin - Gastroenterologie nur geringe Unterschiede in den relativen Standardfehlern ausweist. In diesem Fall treten zwei sich einander aufhebende Faktoren gegenüber: Einerseits sind Vertragsarztpraxen der Gastroenterologie im ZiPP leicht „überrepräsentiert“, was im Fachgebietsdurchschnitt zu Gewichten w_i kleiner Eins (0,8) führt, sodass die gewichtete Summe der Praxen kleiner ist als die ungewichtete Summe. Daraus resultieren zu niedrigere nicht-normierte Standardfehler. Andererseits sind in den Praxen dieses Fachgebietes im Mittel 1,23 Inhaber tätig. In der Folge läge die gewichtete Summe der Praxen über der ungewichteten Summe, daraus folgt ein höherer relativer Standardfehler. Im Zusammenspiel heben sich beide Effekte nahezu gegenseitig auf.

Tabelle 4: Übersicht über Endlichkeitskorrekturfaktoren nach Fachgebieten

Fachgebiet	Endlichkeitskorrektur		
	Praxen ZiPP	Praxen Grundgesamtheit	Faktor
Gesamt	4701	98788	0,98
Allgemeinmedizin und Innere Medizin (hausärztlich)	672	35530	0,99
Anästhesiologie	105	1497	0,96
Augenheilkunde	259	3306	0,96
Chirurgie	223	1958	0,94
Dermatologie	225	2385	0,95
Gynäkologie	509	7410	0,97
Hals-Nasen-Ohren-Heilkunde	307	2854	0,94
Innere Medizin - Gastroenterologie	35	509	0,97
Innere Medizin - Kardiologie	42	911	0,98
Innere Medizin - Pneumologie	49	598	0,96
Innere Medizin - ohne bzw. mit mehreren Schwerpunkten	47	795	0,97
Innere Medizin - sonstige Fachgebiete	42	942	0,98
Kinder- und Jugendmedizin	384	3932	0,95
Kinder- und Jugendpsychiatrie und -psychotherapie	112	640	0,91
Nervenheilkunde	151	1396	0,94
Neurologie	40	514	0,96
Neurologie, Nervenheilkunde, Psychiatrie	68		
Nuklearmedizin	8	165	0,98
Orthopädie	296	3101	0,95
Physikalische und rehabilitative Medizin	35	282	0,94
Psychiatrie	110	1121	0,95
Psychosomatische Medizin und Psychotherapie	387	2231	0,91
Psychotherapie	350	20331	0,99
Radiologie	39	473	0,96
Urologie	157	1860	0,96
Übergreifend tätige Praxen	49	3486	0,99

Quelle: Zi-Praxis-Panel 2014.

Tabelle 3 beinhaltet einen Vergleich der normierten relativen Standardfehler (22) vor und nach der Endlichkeitskorrektur entsprechend dem Korrekturfaktor in (23). Entsprechend (29) liegen die relativen Standardfehler nach der Endlichkeitskorrektur unter den Standardfehlern gemäß (2). Über alle Praxen weichen die relativen Standardfehler der drei zentralen Finanzkennzahlen um 0,02 bis 0,04%-Punkte bzw. um 2,4% von den korrespondierenden Werten vor der Endlichkeitskorrektur ab. Aus dem Auswahlsatz von 4,8% über alle Praxen resultiert ein Korrekturfaktor von 0,98, der das Ergebnis auf Gesamtebene unwesentlich verändert (vgl. Tabelle 4).

Lediglich für Fachgebiete mit reger Beteiligung (u.a. Kinder- und Jugendpsychiatrie und -psychotherapie und die Psychosomatische Medizin und Psychotherapie) beeinflusst die Berücksichtigung der Endlichkeitskorrektur das Gütemaß der geschätzten Mittelwerte spürbar (vgl. Tabelle 3). In diesen Fachgebieten haben sich an der Erhebung 2014 17,5% bzw. 17,3% der Praxen beteiligt, woraus Endlichkeitskorrekturfaktoren in Höhe von jeweils 0,91 berechnet werden. Neben dem Auswahlsatz wirkt auch der Umfang der Grundgesamtheit auf die ermittelten Differenzen ein. Für in der Grundgesamtheit stark vertretene Fachgebiete wie die Allgemeinmedizin und die Psychotherapie ergeben sich relative Abweichungen von unter einem Prozent.

Insgesamt führt die Berücksichtigung der Endlichkeitskorrektur bei der Berechnung der relativen Standardfehler für die meisten Fachgebiete zu einer geringen Reduktion der Varianz des Stichprobenmittelwertes und damit nur zu einer leichten Erhöhung der Schätzgenauigkeit.

7 Fazit

Verhältniszahlen wie der Jahresüberschuss je Inhaber repräsentieren im Zi-Praxis-Panel zentrale Indikatoren der wirtschaftlichen Situation der Praxen in der vertragsärztlichen Versorgung. Der relative Standardfehler bildet dabei die Grundlage zur Einschätzung der Güte der berichteten Mittelwerte. Bei der Anwendung von Statistikprogrammen zur Ermittlung der Stichprobenfunktionen und deren Gütemaße muss beachtet werden, ob Verhältniszahlen und gewichtete Daten vorliegen. Beides trifft auf die Finanzkennzahlen des ZiPP zu. Aus der Nichtberücksichtigung dieser Faktoren können verzerrte Varianzschätzern und Standardfehler resultieren. Als Ausweg wurde eine Normierung der Gewichte vorgestellt – mit dem Ziel, unverzerrte Stichprobenfunktionen zu erhalten.

Allgemein gilt der Zusammenhang, dass der Standardfehler umso kleiner wird, je größer der Stichprobenumfang ist. Dies zeigt sich auch in den ZiPP-Daten. Insgesamt werden geringe relative Standardfehler der zentralen Finanzkennzahlen ausgewiesen, die auf eine hohe Belastbarkeit der ZiPP-Daten hindeuten. Auf der Fachgebietsebene werden infolge der geringen Teilnehmerzahlen in einzelnen Subgruppen wie der Nuklearmedizin vergleichsweise hohe Standardfehler aufgezeigt. Aufgrund der Strukturgewichtung und der Teilnahme von Gemeinschaftspraxen wurden auf Fachgebietsebene zum Teil merkliche Abweichungen in den geschätzten Varianzen und damit zwischen den normierten und nicht-normierten Standardfehlern beobachtet. Über alle Praxen waren die Unterschiede hingegen vernachlässigbar.

Die Berücksichtigung der Endlichkeitskorrektur bei der Berechnung der Standardfehler führt für die meisten Fachgebiete nur zu einer geringen Reduktion der Varianz des Stichprobenmittelwertes und damit nur zu einer leichten Erhöhung der Schätzgenauigkeit. Bei großen Grundgesamtheiten und geringem Auswahlsatz wie im Zi-Praxis-Panel ist die Unterscheidung hinsichtlich der Art des Ziehens der Stichprobe, d.h. die Berücksichtigung der Endlichkeitskorrektur praktisch nicht erforderlich. Für die Mehrzahl der Fachgebiete im ZiPP enthält die Stichprobe nicht mehr als zehn Prozent der Praxen der Grundgesamtheit.

Literatur

Hartung, J. (2009): Statistik. 15. Aufl., München.

Rönz, B./ Strohe, H.G. (Hrsg.)(1994): Lexikon Statistik. Wiesbaden.

Schira, J. (2012): Statistische Methoden der VWL und BWL: Theorie und Praxis. 4. Aufl., München.

Schlittgen, R. (2012): Einführung in die Statistik: Analyse und Modellierung von Daten. 12. Aufl., München.

von der Lippe, P. (2012): Zi-Praxis-Panel und amtliche Kostenstrukturerhebung: ein Methodenvergleich. In: Zi-Paper, Nr. 1.

Zentralinstitut für die kassenärztliche Versorgung in Deutschland (2016): Jahresbericht 2014 - Wirtschaftliche Situation und Rahmenbedingungen in der vertragsärztlichen Versorgung der Jahre 2010 bis 2013. In: Zi-Praxis-Panel Jahresbericht, 5. Jahrgang.