

Evaluation der Datengrundlagen und der Datenaufbereitung für das Kostenmodell KOREG



Schlussbericht

zuhanden der ats-tms AG

Olten, 29. September 2023

Impressum

Bibliographische Angaben

Titel: Evaluation der Datengrundlagen und der Datenaufbereitung für das Kostenmodell KOREG

Autorinnen und Autoren: Schoch, T., Hulliger, B., Spasova, T. und Thees, O.

Auftraggeber: ats-tms AG

Ort: Olten

Datum: 29. September 2023

Projektteam

Tobias Schoch

Beat Hulliger

Tsvetana Spasova

Oscar Thees

Kontakt

Prof. Dr. Tobias Schoch

Fachhochschule Nordwestschweiz

Hochschule für Wirtschaft, Institut ICC

Riggenbachstrasse 16

CH-4600 Olten, Schweiz

E-Mail: tobias.schoch@fhnw.ch

Tel. (direkt): +41 (0)62 957 21 02

Anmerkung

Der Bericht gibt die Auffassung des Projektteams wieder, die nicht notwendigerweise mit derjenigen des Auftraggebers bzw. der Auftraggeberin oder der Begleitorgane übereinstimmen muss. Für den Inhalt ist allein der Auftragnehmer / die Auftragnehmerin verantwortlich.

Bildrechte

Titelbild: Freepik.com (this cover has been designed using assets from Freepik.com)

Inhaltsverzeichnis

Kurzfassung	v
Abkürzungsverzeichnis	viii
1 Einleitung	1
1.1 Ausgangslage und Rahmenbedingung.....	1
1.2 Zielsetzung und Gegenstand der Evaluation.....	2
1.3 Grundzüge von TARMED und TARDOC.....	3
1.4 Terminologie.....	4
1.5 Repräsentativität und Total Survey Quality.....	7
1.6 Weitere Anmerkungen.....	13
1.7 Aufbau des Berichts.....	13
2 Offenlegung der Rohdaten zu den Datensätzen RoKo und Spartenbetriebsdauern	15
3 Rollende Kostenstudie	16
3.1 Steckbrief zur RoKo.....	17
3.2 Aspekte der Repräsentation.....	18
3.3 Aspekte der Messung.....	30
3.4 Zwischenfazit.....	32
4 Spartenbetriebsdauern – Datenbank der nationalen Konsolidierungsstelle NAKO	34
4.1 Steckbrief.....	34
4.2 Aspekte der Repräsentation.....	35
4.3 Aspekte der Messung.....	38
4.4 Zwischenfazit.....	42
5 Alignment zwischen RoKo und NAKO	43
6 Schlussfolgerungen und Empfehlungen	45
Literaturverzeichnis	49
Anhang A: TARDOC	51
Anhang B: Rollende Kostenstudie RoKo	52
B.1 Datenmodell und Übersicht.....	52
B.2 Datenaufbereitung und Datenbereinigung.....	56
B.3 Korrektur/ Herausrechnen von A215.....	57
B.4 Analyse der fehlenden Werte.....	60
B.5 Ausreisserentdeckung und Ausschluss.....	63

B.6 Gewichtung – Expansion/ Vervielfachung.....	68
B.7 Kanton VS: Kalibrierung.....	68
Anhang C: Spartenbetriebsdauern – NAKO.....	70

Kurzfassung

Der Prüfbericht des Bundesamts für Gesundheit (BAG) zur Tarifstruktur TARDOC 1.0 und 1.1 hat kritische Punkte und Empfehlungen betreffend das Kostenmodell KOREG und die Erhebungsinstrumente Rollende Kostenstudie (RoKo, Ärztekasse) und Spartenbetriebsdauern (Daten der Nationalen Konsolidierungsstelle, NAKO, NewIndex AG) ausgesprochen. Der vorliegende Bericht greift die kritischen Punkte und Empfehlungen auf und diskutiert diese aus einer methodisch/statistischen Perspektive.

Die Kritikpunkte und Anmerkungen aus dem Prüfbericht lassen sich in zwei Gruppen aufteilen: (1) Forderung nach Herstellung vollständiger Transparenz; (2) Untersuchungen zu Repräsentativität (inkl. möglicher Verzerrungen infolge präferenzierter Nonresponse und selbst deklarerter Angaben). Der Forderung nach Herstellung von Transparenz kommt der vorliegende Bericht nach, indem er die Datenbearbeitung zur RoKo und NAKO reproduziert, untersucht und bewertet. Die Untersuchung zur «Repräsentativität» orientiert sich an einem Quality Framework zum Total Survey Error. Dieses Quality Framework ist in der Forschung zur Stichprobentheorie weit verbreitet und Teil des European Statistical Systems, dem auch die amtliche Statistik in der Schweiz untersteht. Das Framework stellt die Fitness-for-Use in den Vordergrund, um zu beurteilen, ob ein Schätzer oder eine Statistik sich in einem bestimmten Kontext für statistische Inferenz eignet. Die Beurteilung der Eignung ist kriterienbasiert. Die Orientierung an einem Quality Framework drängt sich auf, weil der Begriff der «Repräsentativität» – entgegen einer weitverbreiteten Meinung – kein Begriff aus der Stichprobentheorie darstellt und dessen Verwendung, legt man wissenschaftliche Kriterien zugrunde, ungenau und unnötig ist.

RoKo und NAKO dienen – im Kontext der Einführung des TARDOC – als Erhebungsinstrumente, um Kenngrößen zu schätzen, die anschliessend für die Parametrisierung des Kostenmodells KOREG verwendet werden. Das Konzept der Fitness-for-Use wird im vorliegenden Kontext angewendet, um zu beurteilen, ob die geschätzten Mittelwerte zu den relevanten Variablen einen möglichst kleinen Mean Square Error (bezgl. des Total Survey Errors) besitzen. Das heisst, ob die Schätzer möglichst unverzerrt und effizient sind.

Rollende Kostenstudie

Die Grundgesamtheit der RoKo umfasst die berufstätigen, selbständig praktizierenden Ärztinnen und Ärzte im ambulanten Bereich (ohne Spitalambulatorien) und ist für jeden Kanton separat definiert (gekoppelt an die Erteilung der Berufsausübungsbewilligungen durch die Kantone). Die Zugehörigkeit zu einer kantonalen Grundgesamtheit wird an zwei konstitutiven Merkmalen des TARMED festgemacht: A1) Die Ärztin/ der Arzt besitzt eine Berufsausübungsbewilligung zur Erbringung von OKP-Leistungen. A2) Sie/ er ist Mitglied der kantonalen Ärztesgesellschaft (inkl. Beitritt zum nationalen Rahmenvertrag TARMED und dem kantonalen Tarifvertrag zwischen der Ärztesgesellschaft und den Versicherern) oder sie/ er ist Nicht-Mitglied, hat sich jedoch dem Rahmen- und Anschlussvertrag des TARMED angeschlossen. Die Definition der (Ziel-) Grundgesamtheit wird an die Mitgliedschaft/ Partizipa-

tion in der kantonalen Ärztegesellschaft geknüpft. Nur Mitglieder der kantonalen Ärztegesellschaften werden der Zielgrundgesamtheit zugerechnet. Nichtmitglieder der kantonalen Ärztegesellschaft (und damit auch nicht der FMH angeschlossen) partizipieren nicht an der Aushandlung der kantonalen Tarifverträge. Sie «schweigen» und schliessen sich dem verhandelten Vertrag an.

Die RoKo (2017–2019) deckt die Grundgesamtheit für 21 Kantone ab. Die vier Zentralschweizer Kantone Luzern, Nidwalden, Obwalden, Uri und der Kanton Jura sind nicht beteiligt. Mit Hilfe der FMH-Ärztestatistik kann die Abdeckung der RoKo grob abgeschätzt werden. Sie beträgt 94.9 % aller Ärztinnen und Ärzten der Grundgesamtheit.

In Kantonen mit freiwilliger Teilnahme fiel die Teilnahmebereitschaft an der RoKo in der Tendenz geringer aus als in Kantonen mit einer Teilnahmepflicht. Das kann zu Verzerrungen führen – muss es aber nicht. Am Beispiel des Kantons Wallis (VS) haben wir detailliert aufgezeigt, dass es mitunter grosse Diskrepanzen bei den Anteilswerten zwischen den RoKo-Erhebungsdaten und dem kantonalen Mitgliederverzeichnis (nach Alter, Spezialisierung, Geschlecht und Urbanität) gibt. Um diese Diskrepanzen genauer abzuschätzen und deren Effekt auf die Schätzung der relevanten Variablen zu beurteilen, wurde eine Kalibrierung für die RoKo im Kanton VS durchgeführt. Die Kalibrierung ist ein Verfahren, um mögliche Verzerrungen infolge von Coverage und Nonresponse Error zu verringern (allenfalls zu beheben). Die geschätzten Mittelwerte der relevanten Variablen (= Parameter für KOREG) wurden ungewichtet (Originaldaten) und mit den kalibrierten Gewichten berechnet. Die mittlere absolute Abweichung zwischen den ungewichteten und gewichteten Mittelwerten beträgt nur 2.4 Prozentpunkte. Das Ausmass der Diskrepanzen kann auch mithilfe des Standardfehlers des Mittelwertschätzers (zu den Originaldaten) beurteilt werden. Diese Metrik erlaubt es, die den Schätzern inhärente Variabilität in der Beurteilung der Diskrepanzen zu berücksichtigen. Im Mittel (über alle relevanten Variablen) haben die Diskrepanzen ein Ausmass von 0.27 Standardfehlern. Die Diskrepanzen sind also gering (im Gegensatz zu den im Prüfbericht des BAG geäusserten Vermutungen). Für die RoKo-Daten des Kantons VS erscheint uns – angesichts der geringen Diskrepanzen – eine Kalibrierung im vorliegenden Kontext nicht zwingend.

Die RoKo wird im Kontext von KOREG als Erhebungsinstrument zur Schweiz verwendet. Es stellt sich also die Frage, ob die kantonale Zusammensetzung in der RoKo getreu abgebildet ist. Die Analyse zeigt, dass Kantone, deren Ärztegesellschaft ein Teilnahmeobligatorium ausgesprochen hat, überproportional in der RoKo vertreten sind. Um das Ausmass der Überrepräsentation abzuschätzen, wurden approximative Hochrechnungsgewichte berechnet, um die kantonalen Unterschiede bei der Teilnahmebereitschaft und damit der Abdeckung auszugleichen. Die Abweichungen zwischen gewichteter und ungewichteter Schätzung sind jedoch gering und liegen in der Grössenordnung von 0.9 %–3.1 % beim Modell X bzw. 5.4 %–7.0 % beim Modell Y. Weil die Gewichtung die geschätzten mittleren Aufwände und Erträge in vergleichbarem Ausmass reduziert, fällt der Effekt auf das Praxisergebnis gering aus. Diese Erkenntnisse sprechen dafür, dass die kantonale Zusammensetzung der RoKo die Schweiz (in Bezug auf die geschätzten Charakteristika) hinreichend gut abbildet.

Die RoKo dient – im Kontext der Einführung des TARDOC – als Erhebungsinstrument, um die Mittelwerte zu den Kosten und Erträgen (differenziert nach weiteren Positionen) der Praxen zu schätzen, die anschliessend für die Parametrisierung des Kostenmodells KOREG verwendet werden. Für diesen Zweck erachten wir die RoKo – auch angesichts fehlender alternativer Datenquellen – für geeignet (im Sinne des Konzepts der Fitness-for-Use bzw. Fitness-for-Purpose).

Spartenbetriebsdauern NAKO/ NewIndex

Die Betriebsdauerdaten der NAKO werden – im Kontext der Einführung des TARDOC – verwendet, um die Betriebsdauern zu schätzen, die anschliessend für die Parametrisierung des Kostenmodells KOREG verwendet werden. Die NAKO ist ein Data Stream of Invoices. Deshalb gibt es keine potenziellen Verzerrungen infolge von Unit- oder Item-Nonresponse, wie dies bei Befragungen der Fall sein kann. Weil die NAKO überdies keine Stichprobenerhebung ist, entfallen Untersuchungen zum Stichprobenfehler gänzlich.

Alignment von RoKo und NAKO

Die Erhebungsinstrumente RoKo und NAKO sind nicht direkt miteinander gekoppelt. Es handelt sich um separate Erhebungsinstrumente. Beide Erhebungsinstrumente erfassen nur anonymisierte Daten, sodass es nicht möglich ist, mithilfe eines Identifikators einzelne Subjekte zwischen den Erhebungen zu identifizieren. Die Unterschiedlichkeit der Erhebungsinstrumente beeinträchtigt deren Vergleichbarkeit nicht per se. Entscheidend ist, ob die Erhebungsinstrumente geeignet sind, (zu den relevanten Variablen) möglichst unverzerrte und effiziente Schätzungen von Charakteristika/ Parametern zur Grundgesamtheit zu liefern. Mit anderen Worten, es ist ausreichend zu prüfen, ob jedes Erhebungsinstrument diesen Anforderungen genügt. Diese Untersuchungen wurden für RoKo und NAKO separat durchgeführt und kommen zum Schluss, dass die beiden Erhebungsinstrumente (im Kontext des KOREG) ein gutes Alignment aufweisen.

Fazit

Die Erhebungsinstrumente RoKo und NAKO sind, bezogen auf das Quality Framework (Total Survey Error), jeweils für sich genommen von ausreichender Qualität (für den angestrebten Zweck). Die Unterschiede zwischen den Instrumenten sind angesichts der Komplexität der zu erfassenden Phänomene gering. Beide Instrumente zielen im Wesentlichen auf die gleiche Grundgesamtheit ab und führen zu relativ geringen Abweichungen bei den Schlüsselindikatoren. Damit ist auch die Kombination der beiden Instrumente ein hinreichend valides Instrument. Verbesserungsmöglichkeiten bestehen immer. Die Antwortrate der Kantone ohne Teilnahmepflicht könnte durch eine Pflicht verbessert werden. Durch die Beteiligung der Zentralschweizer Kantone und des Kantons JU, die nicht an der RoKo partizipieren, könnte die schweizweite Abdeckung verbessert werden.

Abkürzungsverzeichnis

AL	Ärztliche Leistungen
BAG	Bundesamt für Gesundheit
BFS	Bundesamt für Statistik
BStatG	Bundesstatistikgesetz
ESQRS	ESS Standard for Quality Reports Structure
ESS	European Statistical System
ES COP	European Statistics Code of Practice
FMH	Verbindung der Schweizer Ärztinnen und Ärzte
IPL	Infrastruktur- und Personalleistung
KOREG	Kostenmodell KOREG
KVG	Krankenversicherungsgesetz (Bundesgesetz über die Krankenversicherung)
MAS	Strukturdaten der Arztpraxen und ambulanten Zentren (Erhebung)
MSE	Mean Square Error (mittlerer quadratischer Fehler)
MTK	Medizinertarif-Kommission UVG
NAKO	Nationale Konsolidierungsstelle (NewIndex)
OECD GoST	OECD Glossary of Statistical Terms
OKP	Obligatorische Krankenpflegeversicherung
QM	Qualitätsmanagement
RHG	Registerharmonisierungsgesetz
RoKo	Rollende Kostenstudie (Ärztekasse)
SIMS	Single Integrated Metadata Structure (Teil des ESS)
TPW	Taxpunktwert
TSE	Total Survey Error
UVG	Bundesgesetz über die Unfallversicherung
UBR	Untersuchungs- und Behandlungsraum

1 Einleitung

In der obligatorischen Krankenpflegeversicherung (OKP) werden ambulante ärztliche Leistungen in Arztpraxen und Spitälern seit 2004 über die Tarifstruktur TARMED abgerechnet. Diese Tarifstruktur wurde seither mehrmals angepasst, jedoch noch nie ganzheitlich revidiert. Alle Bestrebungen einer Gesamt- oder Teilrevision des TARMED sind in den letzten Jahren gescheitert, sodass der Bundesrat mittels Verordnung per 1. Januar 2018 eine gesamtschweizerisch einheitliche Tarifstruktur für die OKP festgelegt hat.¹

Am 12. Juli 2019 legten curafutura (Verband der Krankenversicherer CSS, Helsana, KPT und Sanitas) und die Verbindung der Schweizer Ärztinnen und Ärzte (FMH) dem Bundesrat einen Antrag zur Genehmigung der Tarifstruktur TARDOC (Vers. 1.0) zur Vergütung ambulanter ärztlicher Leistungen vor, die den TARMED ersetzen soll. Das Bundesamt für Gesundheit (BAG) befand den TARDOC Vers. 1.0 im Dezember 2019 für nicht genehmigungsfähig. Im Juni 2020 haben die Tarifpartner dem Bundesrat die Version 1.1 der Tarifstruktur TARDOC vorgelegt.

Im Bericht «Tarifstruktur TARDOC 1.0 und 1.1: Ergebnisse der formellen und materiellen Prüfung» vom 19. November 2020 hielt das BAG die Ergebnisse der formellen und materiellen Prüfung zu den Version 1.0 und 1.1 des TARDOC fest. Dieser Bericht wird im Folgenden als «Prüfbericht» des BAG bezeichnet; siehe BAG (2020).

1.1 Ausgangslage und Rahmenbedingung

Die ats-tms AG hat diesen Bericht bei der Fachhochschule Nordwestschweiz FHNW – Hochschule für Wirtschaft (nachstehend FHNW genannt) in Auftrag gegeben. Die Trägerschaft der ats-tms AG (Stand² 2022) bestellt sich aus

- FMH,
- curafutura und
- Medizinaltarif-Kommission UVG (MTK).

Die Trägerschaft von ats-tms AG wird in Folgenden auch als die Tarifpartner des TARDOC bezeichnet.

Datensätze

Die Parametrisierung des TARDOC folgt bei den Sparten, bei denen die «Tarifführerschaft» bei den Arztpraxen liegt, dem Kostenmodell KOREG, das seinerseits auf Berechnungen zu den Datensätzen

¹ Siehe Verordnung über die Festlegung und die Anpassung von Tarifstrukturen in der Krankenversicherung (SR 832.102.5).

² Der Verband «Die Spitäler der Schweiz / H+» ist nach der Verabschiedung der finalen Leistungsstruktur im September 2018 aus der Gesellschaft ats-tms AG ausgetreten.

- Rollende Kostenstudie (RoKo) der Ärztekasse und
- Spartenbetriebsdauern (Daten der Nationalen Konsolidierungsstelle, NAKO, NewIndex AG)

abstellt. Für die RoKo und die Spartenbetriebsdauern wird im Folgenden die Bezeichnung Datensätze verabredet. Die Datensätze werden im weiteren Verlauf des Berichts ausführlich charakterisiert und diskutiert.

Datennutzung und Datenschutz

Die FHNW hat am 19. Januar 2022 mit der

- ats-tms AG,
- Ärztekasse,
- NewIndex und
- FMH

eine Vereinbarung betreffend die Datenlieferung und -nutzung zu RoKo und den Spartenbetriebsdauern abgeschlossen. Die Vereinbarung sichert der FHNW den Zugang zu Einzel- / Individualdaten zu. Der ats-tms AG (und namentlich der Tarifpartnerin curafutura) wird der Zugang zu den Einzel-/ oder Individualdaten ausdrücklich untersagt.

1.2 Zielsetzung und Gegenstand der Evaluation

Der Prüfbericht des BAG benennt kritische Punkte zum Kostenmodell KOREG bzw. den Datensätzen (RoKo und Spartenbetriebsdauern) und spricht diesbezüglich Empfehlungen zuhanden der Tarifpartner aus. Der vorliegende Bericht greift die kritischen Punkte und Empfehlungen aus dem Prüfbericht auf und diskutiert diese aus einer methodisch/statistischen Perspektive.

Der Prüfbericht des BAG benennt vier kritische Punkte. Diese lassen sich (in paraphrasierter Form) wie folgt umschreiben.

K1 Berechnungen basieren auf unterschiedlichen Datenquellen

Die Kalkulation der Kosten und Betriebsdauern (je Sparte) beruht auf unterschiedlichen Datenquellen. Die Repräsentativität der Daten für die Grundgesamtheit der ambulanten Praxen wird infrage gestellt. Es sei darum nicht klar, ob «Gleiches mit Gleichem» verglichen werde (Prüfbericht, S. 55–56).

K2 Freiwillige Teilnahme

Die Teilnahme an der Roko-Erhebung ist (in vielen Kantonen) freiwillig. Die freiwillige Teilnahme könne infolge von Selektionseffekten (z. B. selektive Teilnahme) zu verzerrten Schätzungen führen. (Prüfbericht, S. 56)

K3 Selbst deklarierte Angaben

Bei den Angaben in der RoKo handelt es sich um selbst deklarierte Angaben von Ärzten und Ärztinnen. Selbstdeklaration können zu Verzerrungen führen. (Prüfbericht, S. 56)

K4 Spartenbetriebsdauer nicht empirisch erhoben

Die Spartenbetriebsdauern werden anhand von abgerechneten TARMED-Positionen bestimmt. Dies entspräche nicht einer empirischen Erhebung der tatsächlichen Betriebsdauern. Die Folgen davon könnten mögliche Verzerrungen sein. (Prüfbericht, S. 56)

Der Prüfbericht des BAG formuliert darüber hinaus die folgenden Anpassungsempfehlungen:

E1 Herstellung vollständiger Transparenz

Der Prüfbericht fordert die Herstellung von vollständiger Transparenz zum Kostenmodell gegenüber allen Tarifpartnern von TARDOC und der zuständigen Prüfbehörde (BAG) ein. Die Rohdaten aus der RoKo seien gegenüber dem BAG und den Tarifpartnern offenzulegen. (Prüfbericht, S. 57)

E2 Untersuchung zu Repräsentativität und Wirtschaftlichkeit

Es sollten Angaben zur Repräsentativität und der Wirtschaftlichkeit der erhobenen Daten gemacht werden. Die Daten sollten gegebenenfalls bereinigt werden. (Prüfbericht, S. 57)

E3 Transparente Herleitung sämtlicher Parameter

Die Herleitung sämtlicher Parameter müsse transparent ausgewiesen werden. (Prüfbericht, S. 57)

Zusätzlich zu den Empfehlungen E1–E3, benennt der Prüfbericht des BAG weitere Anpassungsempfehlungen, die keinen Bezug zu methodisch-statistischen Aspekten aufweisen (z. B. zur Bildung der Spartencluster). Diese Empfehlungen werden im vorliegenden Bericht nicht thematisiert.

1.3 Grundzüge von TARMED und TARDOC

Die Vergütung medizinischer Leistungen im ambulanten Bereich nach TARMED erfolgt auf Grundlage eines Zeittarifs. Jede Tarifposition besitzt einen Kostensatz für die ärztliche Leistung (AL) und die Infrastruktur- und Personalleistung (IPL). Der TARDOC Einzelleistungstarif beruht für die Berechnung der IPL auf zwei Kostenmodellen:

- Kostenmodell KOREG zur Berechnung der Kostensätze für die IPL der Sparten bzw. Leistungen, die vorwiegend im praxisambulanten Bereich Anwendung finden;
- Kostenmodell INFRA zur Berechnung der Kostensätze der IPL der Sparten bzw. Leistungen, die vorwiegend im spitalambulanten Bereich Anwendung finden.

Das Kostenmodell KOREG wurde zur Tarifierung der Leistungen (genauer der Spartenkostensätze) entwickelt, die mehrheitlich durch selbständig praktizierende Ärztinnen und Ärzte abgerechnet werden. Die Parametrisierung des Modells basiert auf Schätzungen für die RoKo.

Das KOREG-Kostenmodell unterscheidet die in Tabelle 1 aufgeführten Kostenarten (diese werden direkt aus der RoKo übernommen).

Tabelle 1: Kostenarten in KOREG

Kostenart	RoKo	Sparte/ Bemessungsbasis
Raumkosten	A3T Raumaufwand	Quadratmeter
Zinskosten	A4T Zinsen	Investitionen
Abschreibungen	A5T Abschreibungen	Abschreibungssätze
Material/ Medikamente	A1T Material/Medi	Apotheke, Labor und Allg. sparte
Unterhaltskosten	A73 Unterhalt und Rep	Unterhaltssätze
Übriges (fix)	A61 Praxisversicherungen; A72 Fortbildung	Umlage auf Allg. sparte
Übriges (variabel)	A71 Verwaltungsaufwand; A74 Fahrzeugaufwand	Umlage auf Allg. sparte
Eigenkapitalzins	B1 und Eigenkapitalzins; LB1z Eigenkapitalzins	Investitionen
Löhne	A2T Personalkosten	Anteilswerte

Quelle: RoKo-Datenbank, eigene Darstellung.

Der TARDOC Einzelleistungstarif formuliert für KOREG zwei Sub-Modelle:

- Modell X: Psychiatrie
- Modell Y: übrige Spezialitäten

Die beiden Sub-Modelle umfassen eine unterschiedliche Anzahl an Sparten, wobei die Sparten Funktionseinheiten einer Arztpraxis (z. B. Untersuchungsräume) sind. Die aus der RoKo erhobenen Kennzahlen werden auf die Sparten verteilt, wobei tariffremde Kostenelemente (z. B. Medikamente) sogenannten Eliminationssparten zugewiesen werden. Für jede Sparte wird anschliessend das Verhältnis der Gesamtkosten zu den Betriebsdauern (aus NAKO / NewIndex) berechnet. Wir verzichten darauf, die Sparten der Modelle aufzuführen; siehe Anhang A.

1.4 Terminologie

In diesem Kapitel werden zentrale Fachbegriffe der Stichprobentheorie eingeführt. Wir orientieren uns an den in der amtlichen Statistik gebräuchlichen Termini (nach BFS, Eurostat,

OECD). Die Verweise zu den einschlägigen Dokumenten sind als Kurzbelege aufgeführt. Letztere sind (um die Lesbarkeit zu erhöhen) nachfolgend aufgeführt.

Kurzbelege

BStatG	Bundesstatistikgesetz (SR 431.01)
ESS	European Statistical System (Eurostat, 2021, Handbook of Quality and Metadata Reports)
ES COP	European Statistics Code of Practice (Eurostat)
OECD GoST	OECD Glossary of Statistical Terms
RHG	Registerharmonisierungsgesetz (Bundesgesetz über die Harmonisierung der Einwohnerregister und anderer amtlicher Personenregister, SR 431.02)

Datenbeschaffung

Nach der Art der Datenbeschaffung werden folgende Daten unterschieden (vgl. bspw. Art. 4 BStatG): 1) Verwaltungsdaten fallen im Kontext des Gesetzesvollzugs an und 2) Erhebungsdaten werden für einen statistischen Zweck erhoben.

Verwaltungsdaten

Verwaltungsdaten (administrative Daten) sind Daten, die bei einer dem Gesetz unterstellten Organisation durch den Vollzug von Bundesrecht anfallen (Art. 4 BStatG). Es werden in der Regel zwei Arten unterschieden (OECD GoST, ESS):

- Register (administrative register with unique identifiers);
- Daten zu Verwaltungsvorgängen (administrative transactions).

Grundgesamtheit

Die (Ziel-) Grundgesamtheit bezeichnet die Gesamtheit aller Einheiten, zu welchen Merkmale erhoben werden sollen.

Merkmal

Ein Merkmal ist eine Eigenschaft einer statistischen Einheit (z. B. Person oder Sache), die objektiv erfasst und beschrieben werden kann (z. B. Zivilstand). Eine Merkmalsausprägung ist ein konkreter Wert, den ein Merkmal annehmen kann (z. B. verheiratet für das Merkmal Zivilstand).

Identifikator

Der Identifikator ist eine nicht sprechende und unveränderliche Nummer, die als funktionales Element in einem Datenbestand die eindeutige Identifikation einer Person oder Sache erlaubt (vgl. Art. 3 RHG).

(Statistische) Einheit

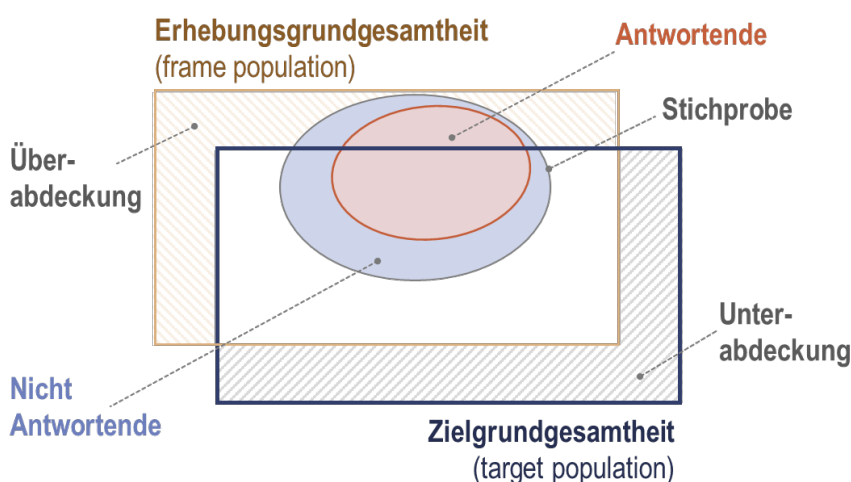
Statistische Einheiten sind die Einheiten (z. B. Personen), zu welchen Informationen gesucht werden (z. B. Merkmale erhoben) und letztlich Statistiken erstellt werden (OECD GoST, ESS).

Diese Einheiten können ihrerseits in Beobachtungseinheiten und Analyseeinheiten unterteilt werden.

Ziel- und Erhebungsgrundgesamtheit

Die Erhebungsgrundgesamtheit (engl. Survey Population) oder auch Rahmengrundgesamtheit genannt weicht in der Regel von der (Ziel-) Grundgesamtheit ab. Die Erhebungsgrundgesamtheit umfasst alle Einheiten, die empirisch tatsächlich erhoben werden können (z. B. zu Befragende besitzen eine gültige Postanschrift). Die Unterschiede zwischen der Zielgrundgesamtheit (engl. Target Population) und Erhebungsgrundgesamtheit werden als Imperfektionen oder Abdeckungsfehler (engl. Coverage Errors) bezeichnet (OCED GoST, ESS). Es wird zwischen Über- und Unterabdeckung unterschieden; siehe Abbildung 1.

Abbildung 1: Schemenhafte Illustration



Quelle: Särndal und Lundström (2005, Kapitel 2.2).

Stichprobenerhebung/ Stichprobe

Als Stichprobe bezeichnet man eine Teilmenge einer Grundgesamtheit, die unter bestimmten Gesichtspunkten ausgewählt wurde (z. B. Zufallsauswahl oder Zufallsstichprobe).

Stichprobenrahmen

Der Stichprobenrahmen (engl. Sampling Frame; siehe Abbildung 1) beinhaltet die Menge aller Einheiten der Erhebungsgrundgesamtheit zusammen mit allen Informationen, die für die Stichprobenziehung und für die Kontaktaufnahme erforderlich sind. Die Modalitäten der Stichprobenziehung sind Gegenstand des Stichprobenplans (ESS).

Zensuserhebung (Vollerhebung)

Die Zensuserhebung ist ein Sonderfall einer Erhebung, weil Daten zu allen Erhebungseinheiten (keine Zufallsauswahl) einer Grundgesamtheit erfasst werden (OECD GoST). Die Zensuserhebung verfolgt einen statistischen Zweck und ist darum kein Register.³

Alle für Stichprobenerhebungen definierten Qualitätsindikatoren finden auch bei der Zensuserhebung Anwendung, mit der Ausnahme, dass der Stichprobenfehler definitionsgemäss null ist (OECD GoST, ESS). Die Abdeckung der Zensuserhebung kann aufgrund von Imperfektionen bei der Erhebungsgrundgesamtheit (over- und undercoverage), Ausfallprozessen (z. B. Teilnahmeausfälle oder -verweigerungen) oder Umsetzungsfehlern, unvollständig oder fehlerhaft sein (ESS).

Nonresponse

Unit-Nonresponse steht für die Teilnahmeverweigerung (gesamthafter Antwortausfall). Unter Item-Nonresponse wird die Antwortverweigerung zu einzelnen Items oder Fragen verstanden; siehe Abbildung 1.

1.5 Repräsentativität und Total Survey Quality

Die empirische Forschung orientiert sich an Gütekriterien. Diese bilden die Grundlage für eine angemessene Bewertung der Forschungsergebnisse. Gütekriterien werden für einzelne Messinstrumente (z. B. standardisierte Fragebogen) und auch für das Forschungsdesign in seiner Gesamtheit formuliert. Für Messinstrumente werden die Gütekriterien Objektivität, Reliabilität und Validität eingefordert (Diekmann, 2009, Kapitel VI.3; Schnell et al. 2018, Kapitel 4.3.2).⁴ Über die zentrale Bedeutung dieser Kriterien, insbesondere deren Anspruch auf Allgemeingültigkeit, besteht in der empirischen Forschung ein breiter Konsens. In ihrer abstrakten, allgemeingültigen Formulierung sind die genannten Gütekriterien für die Forschungspraxis allerdings nicht oder nur bedingt nutzbar. Um den Grad oder das Ausmass an Objektivität, Reliabilität und Validität eines Messinstruments zu bestimmen, müssen die Kriterien einer operationalen Definition zugeführt werden.

³ Nach der Lesart von Eurostat sind die Begriffe Zensus (Volkszählung) und Zensuserhebung keine Synonyme. Der Zensus moderner Prägung kann vollständig oder auch teilweise auf Registerdaten basieren, ohne neue Daten zu erheben. In diesem Fall spricht man nicht von Zensuserhebung.

⁴ Der Grad der Objektivität eines Messinstruments informiert darüber, in welchem Umfang die Ergebnisse unabhängig sind von der Person, die das Instrument anwendet. Die Reliabilität bezieht sich auf die Replizierbarkeit von Messungen und beschreibt, wie präzise oder zuverlässig die Ergebnisse bei Messwiederholungen repliziert werden können. Die Validität gibt den Grad der Genauigkeit an, mit welchem ein Messinstrument das Phänomen, das es zu messen vorgibt, auch tatsächlich misst (Diekmann, 2009, Kapitel VI.3).

1.5.1 Zum Begriff der «Repräsentativität»

Für Stichprobenerhebungen wird häufig – zusätzlich zu den oben genannten Gütekriterien – eingefordert, dass die Stichprobe «repräsentativ» zu sein habe (z. B. Benesch, 2013, S. 12).⁵ Dies sei, so Benesch (2013, *ibid.*), sicher dann der Fall, wenn die «Stichprobe ein verkleinertes, aber sonst wirklichkeitsgetreues Abbild der Grundgesamtheit» (engl. «miniature of the population») darstelle. Diese Sichtweise ist derart weitverbreitet (auch in den Medien), dass eine kritische Besprechung Not tut.

Entgegen einer weit verbreiteten Meinung ist «Repräsentativität» kein Begriff aus der Stichprobentheorie. Der Wortgebrauch in den Medien⁶ und auch in Teilen der Wissenschaften zeichnet sich durch ein hohes Mass an Vagheit und Ambiguität aus. Diese Feststellung trifft heute genauso zu, wie vor über 40 Jahren, als Kruskal und Mosteller (1979a, 1979b, 1979c, 1980) in einer Serie von Artikeln die Verwendung der Begriffe «representative sample» und «representative sampling» untersucht haben. Aus diesen Arbeiten lassen sich zwei Begriffstypen zu «Repräsentativität» synthetisieren:

- Repräsentativität als Eigenschaftsbegriff zur Charakterisierung der Stichprobe (sample), z. B. «miniature of the population»;
- Repräsentativität als Erklärungs begriff zur Ziehung der Stichprobe (sampling); d. h. eine Erklärung, wie die Stichprobe generiert wurde, z. B. «representative sampling as permitting good estimation».

Der Eigenschaftsbegriff setzt an zu hoher Stelle an, weil er die realisierte Stichprobe an sich zu charakterisieren sucht, anstatt auf einzelne Messinstrumente und Teilaspekte zu fokussieren. Eine Stichprobe kann sehr wohl bezüglich eines Teilaspekts nicht «repräsentativ» sein, aber die Population in allen anderen Aspekten «korrekt repräsentieren». Wenn man diesen Gedankengang weiterdenkt, so wird schnell klar, dass der Eigenschaftsbegriff kein (sinnvoll) quantifizierbares Gütekriterium einer Stichprobe darstellt; siehe auch Schnell et al. (2018, Kapitel 6.6).

Repräsentativität als Erklärungs begriff aufzufassen, der den Ziehungsprozess einer Zufallsstichprobe erklärt, ist nicht abwegig. Allerdings ist diese Begriffsbildung für sich genommen nicht erhellend, weil sie stets einer konkreten operationalen Definition bedarf (sonst kann keine Stichprobenziehung erfolgen). Darauf wies bereits W. E. Deming mit der Aussage hin: «[w]hat we select are not representative samples, but probability samples» (zit. nach Kruskal und Mosteller, 1979c, S. 256). Es stellt sich nun zurecht die Frage, was der Begriff des «representative samplings» an begrifflichem Präzisionsgewinn leistet, das nicht bereits durch die Definition der Wahrscheinlichkeitsauswahl gewährleistet wird. Die Antwort ist:

⁵ Benesch (2013) steht in seinem Verständnis stellvertretend für weitere Autorinnen und Autoren. Für weitere Belege siehe Kruskal und Mosteller (1979a und 1979b).

⁶ In der veröffentlichten Meinung wird der Begriff der «repräsentativen Erhebung» häufig als «Gütesiegel» kolportiert, das nicht weiter präzisiert wird, im Sinne von «[t]ake my word, without evidence, that my sample will not lead you astray» (Kruskall und Mosteller, 1979c, S. 246). Dieser Wortgebrauch besitzt «pseudoscientific glamour» (Kruskal und Mosteller, 1979a, S. 13) und ist nur «schmückendes und vorwiegend inhaltsleeres Attribut» (Schnell et al., 2018, S. 296).

Wenig, oder in den Worten von Schnell et al. (2018, S. 298): «[d]ie Verwendung des Begriffs ‚Repräsentativität‘ ist, legt man wissenschaftliche Kriterien zugrunde, ungenau und unnötig. [...] Keine einzelne Masszahl kann eine theoretische Analyse der Ausfallprozesse ersetzen». Wir werden darum im nächsten Kapitel das Konzept der Total Survey Quality einführen, das es erlaubt, die Qualität von Erhebungen umfassend zu untersuchen und zu bewerten.

1.5.2 Total Survey Error und Total Survey Quality

In der Erhebungspraxis kann eine Vielzahl von Fehlern auftreten. Analytisch können zwei Fehlerarten unterschieden werden: zufällige und systematisch auftretende Fehler. Die zufällig auftretenden Fehler resultieren (im Denkmodell der frequentistischen Statistik) aus variierenden Angaben bei wiederholten Stichprobeziehungen. Systematische Fehler haben vielfältige Ursachen (z. B. Nonresponse Error oder Frame Error) und führen zu Verzerrungen (Bias). Von einer Verzerrung wird dann gesprochen, wenn der Schätzer eines Parameters oder einer Charakteristik (z. B. Mittelwert) von dessen wahren Wert abweicht (hier vom Mittelwert in der Population). Der wahre Wert ist (ausser in trivialen Fällen) unbekannt. Unter Umständen ist es trotz Unkenntnis des wahren Werts möglich, das Ausmass der Verzerrung mit indirekten Methoden abzuschätzen.

Total Survey Error

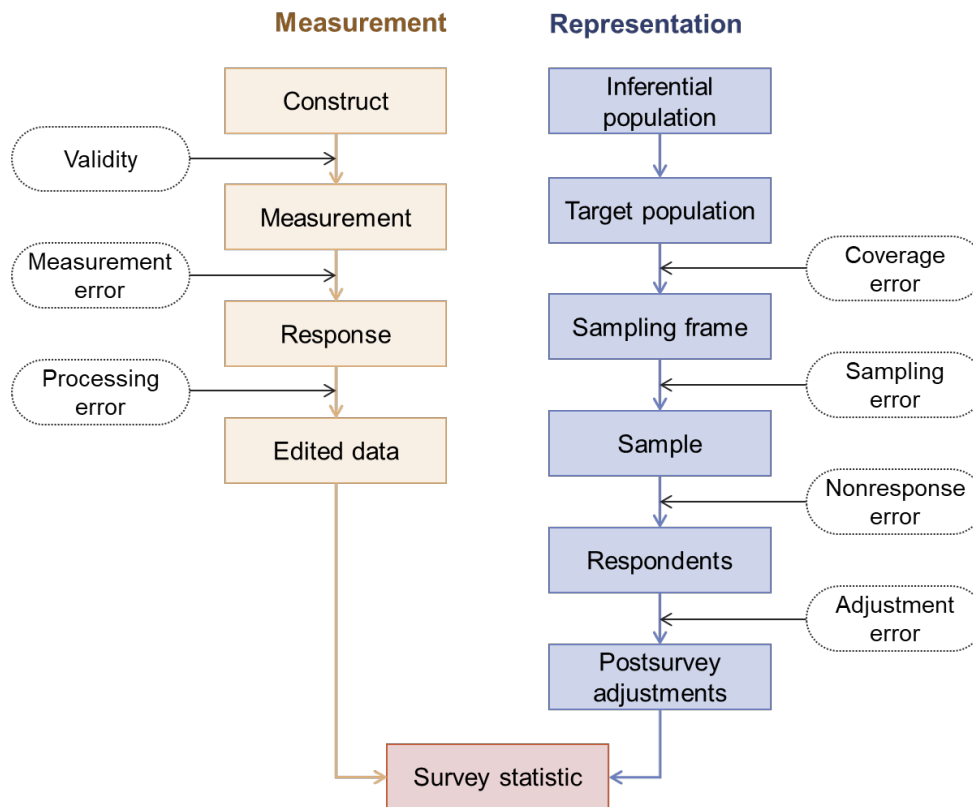
Die Summe aller fehlerhaften Abweichungen zwischen dem wahren Wert und der Schätzung wird z. B. nach Biemer und Lyberg (2003, S. 36) als Total Survey Error (TSE) bezeichnet. Der TSE ist eine theoretische Grösse, die in der Praxis nicht direkt berechnet, allenfalls aber abgeschätzt werden kann. Der TSE kann analytisch in folgende Bestandteile zerlegt werden

$$\text{TSE} = \text{Sampling Error} + \text{Nonsampling Error}$$

Für die Zensuserhebung (Vollerhebung) gilt $\text{Sampling Error} = 0$. Der Nonsampling Error setzt sich aus weiteren Fehlerquellen zusammen, wie sie als Blasen in Abbildung 2 aufgeführt sind. Die grafische Darstellung visualisiert den Survey Life Cycle nach Groves et al. (2009) als zwei Stränge, die beide in die Schätzung münden (Knotenpunkt: Survey Statistic). Der linke Strang illustriert den Messprozess, der vom zu messenden Konstrukt ausgeht und bei der Schätzung endet. Mögliche Fehlerquellen, welche die Messung beeinträchtigen können, sind Measurement und Processing Error (und mangelhafte Validität des Konstrukts). Der rechte Strang illustriert den Repräsentationsaspekt. Ausgehend von der Grundgesamtheit führt der Strang über die Stichprobe und die Teilmenge der antwortenden Subjekte zur Schätzung. Die Etappen entlang dieses Stranges können durch Coverage Error, Sampling Error⁷, Nonresponse Error und Adjustment Errors beeinträchtigt werden.

⁷ Biemer und Lyberg (2003, S. 36) machen eine Unterscheidung zwischen Sampling und Nonsampling Error. Groves et al. (2009, S. 48) hingegen verzichten auf diese Dichotomie und weisen den Sampling Error als einen der Fehler des Gesamtkonzepts aus.

Abbildung 2: Elemente des Total Survey Errors verknüpft mit den Etappen des Measurements and Representational Inference Process



Quelle: Eigene Darstellung nach Groves et al. (2009, Abb. 2.4) und Groves und Lyberg (2010, Abb. 3).

Der TSE wird für jeden zu schätzenden Parameter bzw. jede Charakteristik eigenständig definiert. Sei θ eine (nicht weiter definierte) Charakteristik oder ein Parameter der Grundgesamtheit (z. B. Total der Medikamentenkosten). Der Schätzer wird mit $\hat{\theta}$ bezeichnet. Der mittlere quadratische Fehler (engl. Mean Square Error) des Schätzers ist gegeben durch

$$MSE(\hat{\theta}) = \text{Var}(\hat{\theta}) + \text{Bias}(\hat{\theta})^2$$

wobei Bias für die Verzerrung des Schätzers steht und Var die Varianz bezeichnet, die sich aus variierenden Angaben bei wiederholten Erhebungen ergibt. Der MSE informiert über Präzision eines Schätzers und berücksichtigt alle Fehlerquellen, die in den TSE einfließen; siehe Exkurs. Alle Bemühungen zur Steigerung der Qualität einer Schätzung sind auf die Reduktion des MSE gerichtet.

Der MSE ist eine theoretische Grösse, die vom wahren Wert θ abhängt. Nichtsdestotrotz kann der MSE durch indirekte Analysen und theoretische Überlegungen (insbesondere in Bezug auf einzelne Fehlerquellen, z. B. Nonresponse Error) abgeschätzt werden. Im Gegensatz zum vagen Begriff der Repräsentativität, erlauben es die Konzepte MSE, TSE und die klar definierten Fehlerarten (siehe Abbildung 2) mögliche Verzerrungen zu identifizieren und

entsprechende Gegenmassnahmen einzuleiten oder Korrekturen (z. B. Kalibrierung, siehe nächster Abschnitt) anzubringen.

Exkurs: Schemenhafte Illustration zur Verzerrung (Bias) und Varianz eines Schätzers

Die fünf Pfeile auf den Dartscheiben repräsentieren die Schätzwerte zu fünf (hypothetischen) Replikationen eines Schätzers (diese Interpretation folgt dem Paradigma der frequentistischen Inferenzstatistik). Es sind drei verschiedene Schätzer dargestellt. Die erste Dartscheibe zeigt einen unverzerrten Schätzer (keine Verzerrung, Bias = 0) mit kleiner Varianz. Die mittlere Dartscheibe illustriert einen unverzerrten Schätzer (Bias = 0) mit markanter Streuung. Der dritte Schätzer ist verzerrt («trifft nicht ins Ziel»; Bias > 0), hat aber nur eine geringe Streuung.



Quelle: Biemer und Lyberg (2003, Kapitel 2.4).

Einsatz der Kalibrierung zur Reduktion des TSE

Die Kalibrierung ist ein Verfahren, um mögliche Verzerrungen infolge von Coverage und Nonresponse Error zu verringern (allenfalls zu beheben); siehe bspw. Särndal und Lundström (2005, Kap. 6). Die Kalibrierung wurde ursprünglich für den Einsatz bei Zufallsstichproben entwickelt. Sie wird jedoch auch für Nonprobability Samples eingesetzt; siehe Elliott und Valliant (2017), Mattei (2018) und Valliant (2020). Zur Illustration: Das Bundesamt für Statistik hat die Kalibrierung bei der Zensuserhebung «Strukturdaten der Arztpraxen und ambulanten Zentren» (MAS) verwendet, um Abweichungen infolge von Coverage und Nonresponse Errors zu reduzieren oder zu korrigieren; siehe BFS (2018, 2022). Im folgenden Exkurs ist die Kalibrierung anhand eines Beispiels illustriert.

Exkurs. Beispielhafte Illustration zur Kalibrierung

Wir wollen den Einsatz der Kalibrierung beispielhaft erläutern und nehmen an, dass eine Stichprobe (möglicherweise) durch präferenzielles Antwort-/ Teilnahmeverhalten beeinflusst sei (Problematik: Nonresponse Error/ Bias).

Wir nehmen an, dass das Antwort-/ Teilnahmeverhalten der befragten Personen von Dritt-

variablen abhängt (v. a. soziodemografische Variablen, wie bspw. Alter, Nationalität, etc., aber auch Variablen zum «Mindset» der Befragten). In der Stichprobenerhebung wird (neben den eigentlich interessierenden Variablen) auch eine Vielzahl von Drittvariablen erhoben, deren Populationscharakteristika (z. B. Mittelwert) aus anderen Quellen (z. B. Registerdaten) bereits bekannt sind. Für solche Drittvariablen können allfällige Diskrepanzen zwischen den Charakteristika der Erhebung und denjenigen der Population berechnet werden (z. B. mittleres Alter in der Stichprobe vs. mittleres Alter in der Population).

Durch eine geeignete (Um-) Gewichtung der erhobenen Beobachtungen in der Stichprobe werden die Gewichte derart angepasst (oder erzeugt), dass keine Diskrepanzen mehr zwischen den gewichteten Charakteristika der Erhebung und denjenigen der Population bestehen (d. h. kalibriert). Diese angepassten Gewichte werden nun für alle Berechnungen zur Stichprobe verwendet.

Können damit alle potenziellen Verzerrungen beseitigt werden? Geht man davon aus, dass das Antwort-/ Teilnahmeverhalten der befragten Personen perfekt durch die Drittvariablen erklärt werden kann, dann würde dieses Verfahren alle Verzerrungen beseitigen. In der empirischen Praxis ist dies allerdings sehr unwahrscheinlich. Es ist einerseits nicht a priori bekannt, welche Variablen das Antwort-/ Teilnahmeverhalten tatsächlich determinieren (man hat i. d. R. nur Vermutungen), andererseits werden in den Stichproben nicht alle relevanten Drittvariablen erhoben. Insbesondere verfügt man typischerweise keine Angaben zum «Mindset» der Befragten. Aus diesen Gründen vermag die (Um-) Gewichtung in der Praxis nur einen Teil der Verzerrungen zu beseitigen. Nichtsdestotrotz ist die Kalibrierung eine geeignete Methode, um den Bias wenigstens teilweise zu reduzieren.

1.5.3 Fitness-for-Use/ Fitness-for-Purpose

Das European Statistical System (ESS), dem auch die amtliche Statistik der Schweiz angeschlossen ist, formuliert – als Code of Practice und Quality Assurance Framework – 16 Prinzipien zur Entwicklung, Produktion und Verbreitung von Statistiken; siehe Tabelle 2.

Der Total Survey Error bildet eine wichtige Qualitätsdimensionen innerhalb des Prinzipienkatalogs ab. Diese Dimension findet Eingang in die Prinzipien Sound Methodology, Appropriate Statistical Procedures und Accuracy and Reliability. Daneben sind auch weitere Prinzipien von Bedeutung, wie Relevance, Timeliness, Coherence etc. Einige der Prinzipien sind für die amtliche Statistik zugeschnitten (z. B. Mandate for Data Collection and Access to Data oder Accessibility and Clarity) und können nicht sinnvoll auf nicht staatliche Einheiten, wie die Ärztekasse Genossenschaft als Trägerin der RoKo, übertragen werden. Nichtsdestotrotz sind die 16 Prinzipien hilfreich, um die Fitness-for-Use/ Fitness-for-Purpose einer Statistik oder Schätzung zu bewerten.

Tabelle 2: Die 16 Prinzipien des European Statistical Systems zur Entwicklung, Produktion und Verbreitung von Statistiken

Institutional Environment	Statistical Processes	Statistical Outputs
1 Professional Independence	7 Sound Methodology	11 Relevance
1 ^{bis} Coordination and Cooperation	8 Appropriate Statistical Procedures	12 Accuracy and Reliability
2 Mandate for Data Collection and Access to Data	9 Non-Excessive Burden on Respondents	13 Timeliness and Punctuality
3 Adequacy of Resources	10 Cost Effectiveness	14 Coherence and Comparability
4 Commitment to Quality		15 Accessibility and Clarity
5 Statistical Confidentiality and Data Protection		
6 Impartiality and Objectivity		

Quelle: Eurostat (2021, Kapitel 2.3).

Anmerkung. Die Vielfältigkeit der 16 Prinzipien des ESS zeigt auf, dass Qualitätsbewertungen zu Statistiken durch einen «Prozess des Abwägens» gekennzeichnet sind. Viele der Prinzipien sind komplementär zueinander, sodass eine «Verbesserung» in einer Dimension häufig mit einer «Verschlechterung» in einer anderen einhergeht. So ist es bspw. nicht oder nur schwer möglich, den «Datenhunger» nach mehr, detaillierteren und zeitlich feiner abgelösten Daten zu stillen, ohne dabei Einbussen bei den Aspekten Non-Excessive Burden on Respondents, Cost Effectiveness und Accuracy and Reliability hinnehmen zu müssen. Z. B. ein längerer Fragebogen ist zwar geeignet, um detailliertere Kennzahlen zu erfassen, führt jedoch typischerweise zu einer kleineren Teilnehmerate, woraus Verzerrungen resultieren können, d. h. Einbussen bei Accuracy and Reliability.

1.6 Weitere Anmerkungen

Alle Berechnungen in diesem Bericht wurden mit der R-Statistiksoftware (Language and Environment for Statistical Computing, siehe R Development Core Team, 2023) durchgeführt.

1.7 Aufbau des Berichts

In Kapitel 2 gehen wir auf die Forderung des BAG nach Offenlegung der Rohdaten zu den Datensätzen RoKo und Spartenbetriebsdauern/ NAKO ein. In den Kapiteln 3 und 4 werden die Erhebungsinstrumente RoKo und Spartenbetriebsdauern/ NAKO in Bezug auf Qualität

und Fitness-for-Use diskutiert. In Kapitel 5 wird das Alignment von RoKo und Spartenbetriebsdauern/ NAKO (im Zusammenhang mit KOREG) besprochen. Der Bericht schliesst mit Schlussfolgerungen und Empfehlungen an; siehe Kapitel 6.

Orientierungshilfe

Die Besprechung der im Prüfbericht des BAG aufgeworfenen kritischen Punkte K1–K4 und Empfehlungen E1–E3 erfolgt in unterschiedlichen Kapiteln. Zur besseren Orientierung mag die folgende tabellarische Übersicht eine Hilfestellung bieten.

K1	«Repräsentativität» und Unterschiedlichkeit der Datengrundlagen	Kapitel 3 und Kapitel 5
K2	Freiwillige Teilnahme und mögliche Verzerrungen	Kapitel 3 und Anhang B
K3	Selbst deklarierte Angaben	
K4	Nicht empirisch erhobene Daten	Kapitel 4 und Anhang C
E1	Forderung nach Offenlegung der Rohdaten und Herstellung vollständiger Transparenz	Kapitel 2
E2	Forderung nach Offenlegung der Rohdaten und Herstellung vollständiger Transparenz	Kapitel 3
E3	Transparente Herleitung sämtlicher Parameter	Kapitel 3 und Kapitel 4 und Anhänge B und C

2 Offenlegung der Rohdaten zu den Datensätzen RoKo und Spartenbetriebsdauern

Die Anpassungsempfehlung E1 des BAG zielt darauf ab, «vollständige Transparenz zum Kostenmodell» gegenüber allen Beteiligten und der zuständigen Prüfbehörde zu schaffen. Darüber hinaus empfiehlt der Prüfbericht des BAG die Offenlegung der RoKo-Daten (damit ist der Zugang der Tarifpartner zu den Einzel-/ Mikrodaten gemeint).

Die Forderung nach Offenlegung ist keine notwendige Bedingung, zur Schaffung von Transparenz. Diese Forderung trägt überdies den bestehenden Datennutzungsrechten und dem Datenschutz zu wenig Rechnung. Anstelle der Forderung nach Offenlegung der Einzeldaten, kann Transparenz auch durch den Einbezug eines Intermediärs sichergestellt werden. Im vorliegenden Fall üben die Autoren/innen des Berichts die Rolle des Intermediärs aus. Die Dateneignerin (Ärztekasse) stellt der Datennutzerin (ats-tms AG) nur aggregierte Daten (Mittelwerte/ Modellwerte pro Sparte) zur Verfügung. Der Intermediär ist in die Details der Datenaufbereitung und die Berechnung der Aggregate eingeweiht und verbürgt gegenüber der Datennutzerin, dass diese Arbeiten den vereinbarten Standards genügen.

3 Rollende Kostenstudie

In diesem Kapitel wird das Erhebungsinstrument der Rollenden Kostenstudie (RoKo) charakterisiert und im Hinblick auf die eingangs formulierten Evaluationsfragen untersucht. Die Besprechung orientiert sich an den in Abbildung 2 nach Groves et al. (2009, Abb. 2.4) formulierten Aspekten zur Qualität und zum Total Survey Error:

- Aspekte der Messung (z. B. Processing Errors)
- Aspekte der Repräsentation (z. B. Nonresponse oder Coverage Error)

Nach einem Überblick zu den Metadaten (nach Eurostat) wird die RoKo in einem kurzen Steckbrief beschrieben.

Metadaten zur RoKo

Anmerkung. Die Metadaten zur RoKo sind nach dem ESS Standard for Quality Reports (ESQRS) V2.0 strukturiert; siehe Eurostat (2021, Appendix B)

Contact

Contact organisation Ärztekasse

Statistical presentation

Statistical unit Independently practicing ambulatory medical organisations

Reference area Switzerland

Time coverage 2017–2019

Statistical processing

Source data Full enumeration

Frequency of data collection Annual

Data collection Mode: Online and paper survey (Ärztekasse)

Data compilation Estimation and imputation; see below

Adjustment

Accuracy and reliability

Sampling error Does not apply (full enumeration, not sample)

Non-sampling error See below

Coverage error See below

Measurement error See below

Non response error See below

Processing error See below

Imputation - rate Imputation only for variable A215; see below

Seasonal adjustment Does not apply

Confidentiality

Confidentiality - policy Confidential

3.1 Steckbrief zur RoKo

Die Rollende Kostenstudie ist eine von den kantonalen Ärztegesellschaften und der Ärztekasse durchgeführte Datenerhebung. Die Erhebungssubjekte sind die selbständig praktizierenden Ärztinnen und Ärzte. Der Erhebungsmodus ist eine schriftliche Befragung (online oder papierbasiert).

Die Befragten erfassen jährlich Angaben zu ihrem Material-, Personal-, Raum- und Kapitalaufwand sowie zu den Abschreibungen. Zusätzlich werden die Aufwendungen für Versicherungen und der übrige Aufwand (Verwaltungsaufwand, Weiterbildung, Beiträge an Berufsverbände usw.) erfasst. Neben den Aufwendungen erfasst die RoKo auch die Erträge. In Ergänzung zu den finanziellen Daten werden auch Angaben zur Praxis (Grösse, Auslastung, Öffnungszeiten, Angestellte etc.) erhoben.

Die vorliegende Analyse im Zusammenhang mit der Einführung des TARDOC beruht auf einem Datenauszug aus der RoKo (vom 6.10.2021) für die Referenzjahre 2017–2019. Die RoKo dient – im Kontext der Einführung des TARDOC – als Erhebungsinstrument, um Kenngrößen zu schätzen, die dann für die Parametrisierung des Kostenmodells KOREG verwendet werden. Für die Parametrisierung werden die geschätzten Mittelwerte der in Tabelle 3 aufgeführten Variablen benutzt.

Tabelle 3: Für KOREG relevante Variablen

Variablen		Variablen (Forts.)	
A1T	Total Materialaufwand	A71	Verwaltungsaufwand, Beiträge Berufsverbände, Fortbildung
A2T	Total Personalaufwand	A72	Aufwand aus Praxispartnerschaft
A3T	Total Raumaufwand	A73	Unterhalt und Reparaturen der Praxiseinrichtungen
A4T	Total Kapitalaufwand	A74	Fahrzeugaufwand
A5T	Total Abschreibungen	A75	Praxisversicherungen
B1	Bilanzwerte	E1T	Total Bruttoertrag der Praxistätigkeit

Quelle: RoKo-Datenbank; Überleitungsbogen.

Fitness-for-Use

Das Konzept der Fitness-for-Use wird auf die RoKo angewendet, um zu beurteilen, ob die geschätzten Mittelwerte zu den relevanten Variablen einen möglichst kleinen Mean Square Error (bezgl. des Total Survey Errors) besitzen. Das heisst, ob die Schätzer möglichst unverzerrt und effizient sind.

Umfang der RoKo

Für die Analyse werden die Daten der Referenzjahre 2017–2019 gemeinsam betrachtet. Zur Glättung der geschätzten Kennzahlen hat die Ärztekasse das folgende Vervielfachungs- oder Gewichtungsschema verwendet:

$$3 \times \text{Daten 2019} + 2 \times \text{Daten 2018} + 1 \times \text{Daten 2017}$$

Der resultierende Datensatz umfasst also 3 Kopien der 2019er-Daten und 2 Kopien der 2018er-Daten. Die Daten für das Jahr 2017 werden nur einmal hinzugefügt. Mit dem 3-2-1-Gewichtungsschema will die Ärztekasse den aktuelleren Daten mehr Gewicht geben. Als Folge der Vervielfachung (Expansion) resultieren für die Modelle X und Y eine (Netto-) Anzahl⁸ von 7'501 bzw. 29'263 Einträge; siehe Tabelle 4. Ohne die Vervielfachung resultieren für das Modell X bzw. Y im Mittel eine Anzahl von 1'250 bzw. 4'877 Fälle pro Jahr. Die Fallzahlen der einzelnen Jahre können Tabelle 4 entnommen werden.

Tabelle 4: Fallzahlen (nach Jahr, expandiert und im Mittel)

Modell	Fallzahlen			Expandiert	im Mittel
	2017	2018	2019		
Modell X	1'271	1'267	1'232	7'501	1'250
Modell Y	4'812	4'919	4'871	29'263	4'877

Quelle: RoKo-Datensatz; eigene Berechnungen.

3.2 Aspekte der Repräsentation

3.2.1 Grundgesamtheit

In der Diskussion zum Konzept der Grundgesamtheit wird zwischen der (Ziel-) Grundgesamtheit (d. h. dem Ideal) und der Erhebungsgrundgesamtheit (auch Rahmengrundgesamtheit genannt) unterschieden, wobei letztere der Menge aller Erhebungssubjekte entspricht, die empirisch tatsächlich erhoben werden können. Diskrepanzen zwischen den beiden Ausprägungen der Grundgesamtheit werden als Coverage Errors oder Imperfektionen bezeichnet. Ob eine Erhebungsgrundgesamtheit geeignet ist, um Charakteristika zur Ziel-Grundgesamtheit zu schätzen, bemisst sich am Total Survey Error, d. h. ob die Schätzungen möglichst unverzerrt und effizient sind.⁹

Die Bestimmung der Erhebungsgrundgesamtheit (Rahmengrundgesamtheit) der RoKo ist durch die Trägerschaft der Erhebung motiviert. Dazu muss man sich den Entstehungszusammenhang der RoKo als Erhebungsinstrument der kantonalen Ärztesellschaften und der Ärztekasse (d. h. ärzteigene Datenerhebung) in Erinnerung rufen. Die Erhebungsgrundgesamtheit entspricht den Mitgliederverzeichnissen der kantonalen Ärzteschaften

⁸ Die Nettoanzahl (Nettostichprobengrösse) meint die für die Schätzung verwendete Anzahl an Beobachtung (nach allfälligen Korrekturen, Ausschlüssen u. dgl. m.)

⁹ Gelegentlich wird in der Literatur auch von der Verallgemeinerbarkeit der Resultate (externer Validität) gesprochen. Es ist eine Frage der Perspektive, ob man die Verallgemeinerbarkeit an sich oder die Abwesenheit von Fehlern zur Grundlage der Argumentation macht.

ten. Auf dieser Grundlage wurden die Ärztinnen und Ärzte in den Kantonen zur Teilnahme an der Erhebung eingeladen.

Eine Definition für die (Ziel-) Grundgesamtheit anzugeben, ist anspruchsvoll. Dies liegt vor allem daran, dass sie nicht in Form eines amtlichen Registers o. Ä. erfasst ist (siehe Exkurs). Aus diesem Grund ist eine Konstruktion der Grundgesamtheit erforderlich. Dies soll schrittweise erfolgen.

Im ersten Schritt halten wir fest, dass die berufstätigen, selbständig praktizierenden Ärztinnen und Ärzte im ambulanten Bereich (ohne Spitalambulatorien) die Einheiten der (Ziel-) Grundgesamtheit bilden. In den nächsten Schritten soll die Definition präzisiert werden.

Exkurs: Medizinalberuferegister und MAS-Erhebung

Im Medizinalberuferegister (MedReg) des Bundesamts für Gesundheit sind alle in der Schweiz tätigen Medizinalpersonen verzeichnet. Im Kontext dieser Studie liegt der Fokus auf den Ärztinnen und Ärzten. Das MedReg führt die von den kantonalen Behörden erteilten kantonalen Berufsausübungsbewilligungen (sowie allfällige Auflagen und Einschränkungen der Berufsausübungsbewilligung) und weitere Individualmerkmale (Diplome, Weiterbildungstitel, Berechtigungen im Umgang mit Betäubungsmitteln, Selbstdispensation etc.). Die Daten zu angestellten Ärztinnen und Ärzten müssen im MedReg auch erfasst werden.

Eine Ärztin oder ein Arzt kann Berufsausübungsbewilligungen in mehreren Kantonen besitzen und innerhalb eines Kantons parallele Bewilligungen für mehrere Praxisniederlassungen/ Standorte besitzen (z. B. Mitarbeitende in Praxen, die mehrere Niederlassungen haben).¹⁾ Per 31.12.2022 weist das MedReg über 7'000 Ärzt/innen aus, die in mehr als einem Kanton eine Bewilligung haben.²⁾ Das MedReg nimmt keine Unterscheidung zwischen Haupt- oder Nebenniederlassungen vor. Schwerer wiegt – im Kontext der Konstruktion der Grundgesamtheit –, dass das MedReg keine tragfähigen Angaben zum Erwerbsstatus der Ärzt/innen und den Berufsausübungsmodalitäten enthält. Es ist deshalb nicht ohne weiteres möglich, die freiberuflich tätige Ärzteschaft von den angestellten Ärztinnen und Ärzten bzw. von der in Spitälern tätigen Ärzteschaft zu unterscheiden. Der Versuch, die Grundgesamtheit in Abwesenheit von tragfähigen Angaben zum Erwerbsstatus und zu den Berufsausübungsmodalitäten zu formulieren, dürfte in einem ausnehmend grossen Ausmass durch Coverage Errors (fälschliche Inklusion/ Exklusion in die bzw. aus der Grundgesamtheit) verzerrt sein, dass diese den Total Survey Error dominieren. Weil die genannten Merkmale unbekannt sind, können sie nicht in den (Stichproben-) Rahmen integriert und für das Post-Survey Adjustment (z. B. Kalibrierung oder Post-Stratification) genutzt werden. Das MedReg (Stand 2023) ist darum nicht geeignet, um daraus die Grundgesamtheit zu konstruieren, an welcher die Abdeckung der RoKo beurteilt werden kann.

Die Strukturdaten der Arztpraxen und ambulanten Zentren (MAS-Erhebung) des Bundesamts für Statistik ist eine Unternehmensdatenerhebung. Als solche fokussiert sie auf das Unternehmen (d. h. juristische Einheiten), die eine Rechtspersönlichkeit haben, im Gegen-

satz zur RoKo, welche die Ärztin und den Arzt zum Gegenstand der Erhebung machen. Die MAS-Erhebung beruht auf dem Betriebs- und Unternehmensregister des BFS (Basis: Unternehmensidentifikationsnummer, UID) und verwendet Daten der Zentralen Ausgleichsstelle zur AHV/IV (Basis: AHV-Nummer) und weiteren Quellen (z. B. Mehrwertsteuerregister).³⁾ Die eindeutigen Identifikatoren UID und AHV-Nummer sind aus Gründen des Datenschutzes nicht in der RoKo enthalten, was eine Verknüpfung verunmöglicht. Des Weiteren deckt die MAS sowohl Arztpraxen und Ambulatorien als auch weitere Institutionen des ambulanten Sektors ab, die nicht zulasten des KVG abrechnen⁴⁾. Die selbständig praktizierende Ärzteschaft lässt sich auf Basis der MAS nicht oder nur indirekt und damit nicht genügend präzise identifizieren bzw. von den anderen Leistungserbringern im ambulanten Sektor abgrenzen. Es ist deshalb nicht sinnvoll, die Grundgesamtheit der MAS auf die RoKo anzuwenden.

¹ Verordnung über das Register der universitären Medizinalberufe, SR 811.117.3

² BAG (2021). Ärztinnen und Ärzte. Bundesamt für Gesundheit (Direktionsbereich für Gesundheitsberufe und Verbraucherschutz). Bern.

³ BFS (2018). Erste Erhebung «Strukturdaten der Arztpraxen und ambulanten Zentren» (MAS 2015): Analyse von Teilnahme und Grundgesamtheit, Bundesamt für Statistik, Neuchâtel (BFS-Nummer be-d-14.04.05-01).

⁴ Ibid.

Anmerkung. Zur Vereinfachung der nachfolgenden Diskussion werden die oben genannten Individualmerkmale (aktive Berufstätigkeit als selbständig praktizierende/r Ärztin/Arzt im ambulanten Bereich) stillschweigend vorausgesetzt und nicht ständig wiederholt. Wir sprechen verkürzt von Leistungserbringern und meinen damit die selbständig praktizierenden Ärztinnen und Ärzte im ambulanten Bereich.

Die Erbringung von medizinischen Leistungen zulasten der OKP ist für Ärztinnen und Ärzte an die Erteilung einer kantonalen Berufsausübungsbewilligung gekoppelt und darum nach Kantonen differenziert zu betrachten. Jede kantonale Ärzteschaft konstituiert eine eigenständig kantonale Grundgesamtheit. Es resultieren somit 26 Grundgesamtheiten.

Die Zugehörigkeit zu einer kantonalen Grundgesamtheit (für ein bestimmtes Referenzjahr) wird an zwei konstitutiven Merkmalen des TARMED festgemacht. Der Bezug zum TARMED ist darin begründet, dass die RoKo eine Erhebung zu den nach dem TARMED verrechneten Leistungen ist. Die Merkmale sind wie folgt definiert.

A1 Die Ärztin/ der Arzt besitzt eine – durch die kantonalen Gesundheitsbehörden erteilte – aktuelle (für das Referenzjahr gültige) Berufsausübungsbewilligung zur Erbringung von OKP-Leistungen.

A2 Die Ärztin/ der Arzt ist

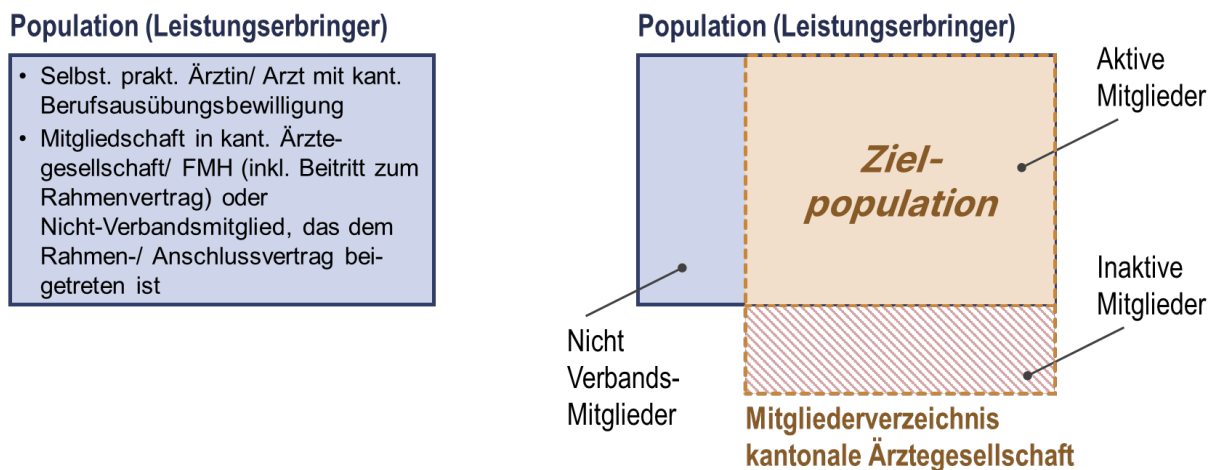
- reguläres Mitglied der kantonalen Ärztesgesellschaft (inkl. Beitritt zum nationalen Rahmenvertrag TARMED und dem kantonalen Tarifvertrag zwischen der Ärztesgesellschaft und den Versicherern) oder
- nicht Mitglied der kantonalen Ärztesgesellschaft, aber dem Rahmen- und Anschlussvertrag des TARMED beigetreten.

Die Bedingungen A1 und A2 bilden gemeinsam eine notwendige und hinreichende Bedingung für die Zugehörigkeit zur (kantonalen) Grundgesamtheit. Nur Ärztinnen und Ärzte, die eines der beiden Teilkriterien unter Lemma A2 erfüllen, können medizinische Leistungen zulasten der OKP nach dem TARMED abrechnen. Darum ist dieses Merkmal für die Zugehörigkeit zur (Ziel-) Grundgesamtheit konstitutiv.

Weitere Anmerkungen.

- Die kantonalen Ärztesgesellschaften bilden die Basisorganisationen der FMH. Einzelmitglieder (Individuen) einer kantonalen Ärztesgesellschaft sind auch Mitglied der FMH. Der Beitritt zur FMH erfolgt über die Basisorganisationen.
- Abbildung 3 zeigt eine schematische Repräsentation der Grundgesamtheit der selbstständig praktizierenden Ärzteschaft im ambulanten Bereich (ohne Spitalambulatorien). Zur Vereinfachung der Darstellung haben wir darauf verzichtet, die Population für jeden Kanton einzeln anzugeben.
- Der TARDOC lehnt sich betreffend die Definition der Leistungserbringer an den TARMED an.

Abbildung 3: Schematische Illustration der Ziel-Population (selbstständig praktizierende Ärzteschaft im ambulanten Bereich ohne Spitalambulatorien)



Quelle: Eigene Darstellung.

Anmerkungen: Zur Vereinfachung der Darstellung haben wir darauf verzichtet, die Population für jeden Kanton einzeln anzugeben.

Der Taxpunktwert (TPW) nach TARMED wird zwischen den kantonalen Ärztesgesellschaften und den Versicherern verhandelt (Privatpraxen und ambulante Spitalpraxen eines Kantons haben i. d. R. unterschiedliche TPW; siehe bspw. Weidmann, 2004). Nichtmitglieder der kantonalen Ärztesgesellschaft partizipieren nicht an der Aushandlung der kantonalen Tarifverträge. Sie «schweigen» und schliessen sich dem verhandelten Vertrag an. Vieles spricht dafür, dass es sich um ein qualifiziertes Schweigen oder zustimmendes Schweigen handelt.¹⁰

Festlegung (Partizipationskriterium). Die Definition der (Ziel-) Grundgesamtheit wird an die Mitgliedschaft/Partizipation in der kantonalen Ärztesgesellschaft geknüpft. Nur Mitglieder der kantonalen Ärztesgesellschaften werden der Zielgrundgesamtheit zugerechnet.

Diskussion

Die Mitglieder der kantonalen Ärztesgesellschaften bilden nur einen Teil der berufstätigen, selbständig praktizierenden Ärztinnen und Ärzte im ambulanten Bereich (ohne Spitalambulatorien) ab. Wie bereits oben ausgeführt, sind die Mitglieder der kantonalen Ärztesgesellschaften auch FMH-Mitglieder. Die FMH besitzt zwar keine 100% Abdeckung aller in der Schweiz approbierten Ärztinnen und Ärzte, allerdings ist die Abdeckung hoch.¹¹

Die Einforderung des Mitgliedschafts-/ Partizipationskriteriums ist auch pragmatischen Gründen geschuldet. Nur Mitglieder der kantonalen Ärztesgesellschaft legen ihre Daten zuhänden der RoKo offen. Für Nichtmitglieder werden keine Daten zur Wirtschaftstätigkeit erhoben.¹²

3.2.2 Abdeckung: Eine erste Betrachtung

Für die RoKo-Erhebung zeichnen die kantonalen Ärztesgesellschaften verantwortlich. In 13 Kantonen haben die Ärztesgesellschaften für die RoKo eine Datenlieferungspflicht ausgesprochen (siehe Tabelle 5). Dabei ist zu beachten, dass die Pflicht keine gesetzliche Grundlage hat, sondern eine vertragliche Verpflichtung ist, welche mit der Mitgliedschaft eingegangen wird. In acht Kantonen ist die Teilnahme fakultativ; sie wird jedoch durch die kantonalen Ärztesgesellschaft mitunter incentiviert (z. B. durch einen Teilerlass der Mitgliederbei-

¹⁰ Nichtmitgliedern steht es frei, sich durch den Beitritt zur kantonalen Ärztesgesellschaft Gehör zu verschaffen, Daten zu ihrer Tätigkeit an die RoKo beizusteuern und ihre Interessen in die Verhandlungen einzubringen.

¹¹ Die Berechnung der Abdeckung gestaltet sich aufgrund unterschiedlicher Definitionen schwierig. Wir wollen darum nur Zahlen vorbringen, die einen groben Eindruck vermitteln: Per 31.12.2020 zählt die FMH 43'061 Mitglieder, inkl. Pensionierte (FMH, Mitgliederstatistik nach Basisorganisationen); das MedReg weist für den gleichen Zeitpunkt 37'761 Ärzt/innen mit erteilter Berufsausübungsbewilligung und fachlicher Verantwortung (inkl. angestellte Ärzt/innen) aus (BAG, 2021).

¹² Um das Partizipationskriterium ernsthaft anzuzweifeln, müssten empirischen Fakten vorgebracht werden, z. B. ein Nachweis, dass sich die Nichtmitglieder bezüglich wirtschaftlicher Daten signifikant von den Mitgliedern unterscheiden. Mit den (gegenwärtig) verfügbaren Daten kann ein solcher Nachweis nicht geleistet werden.

träge). Vier Zentralschweizer Kantone (UR, NW, OW und LU) und der Kanton JU beteiligen sich nicht an der RoKo.

Die Abdeckung durch die RoKo kann mithilfe der FMH-Ärztestatistik 2020 grob abgeschätzt werden. Die Abschätzung stützt sich auf die berufstätige Ärzteschaft mit Hauptberufstätigkeit im Praxissektor¹³. Sie stellt eine Annäherung an die relevante Grundgesamtheit dar. Auf dieser Basis wird der Anteil der berufstätigen Ärzteschaft mit Hauptberufstätigkeit im Praxissektor in den Kantonen UR, NW, OW, LU und JU, die nicht an der RoKo teilnehmen, in der Höhe von 5.1 % abgeschätzt. Die RoKo-Kantone haben somit einen Anteil von 94.9 %.¹⁴ Es ist plausibel aus dieser Abschätzung zu folgern, dass die kantonale Abdeckung der RoKo wohl etwa in der gleichen Grössenordnung liegt.

Wir kommen in Kapitel 3.2.5 noch einmal auf die Besprechung der Abdeckung zurück. Dann diskutieren wir, ob einzelne Kantone disproportional repräsentiert sind.

Tabelle 5: Datenlieferungspflicht in der RoKo (nach Kanton)

Kanton	Bemerkung	Kanton (Forts.)	Bemerkung (Forts.)
AG	Datenlieferungspflicht	NW	Keine Beteiligung an der RoKo
AI	Datenlieferungspflicht	OW	Keine Beteiligung an der RoKo
AR	Datenlieferungspflicht	SG	Datenlieferungspflicht
BE	Datenlieferungspflicht	SH	Datenlieferungspflicht
BL	Keine Pflicht	SO	Datenlieferungspflicht
BS	Datenlieferungspflicht	SZ	Keine Pflicht
FR	Keine Pflicht	TG	Datenlieferungspflicht
GE	Keine Pflicht	TI	Keine Pflicht
GL	Datenlieferungspflicht	UR	Keine Beteiligung an der RoKo
GR	Datenlieferungspflicht	VD	Keine Pflicht
JU	Keine Beteiligung an der RoKo	VS	Keine Pflicht
LU	Keine Beteiligung an der RoKo	ZG	Datenlieferungspflicht
NE	Keine Pflicht	ZH	Datenlieferungspflicht

Quelle: RoKo-Datenbank; eigene Darstellung.

¹³ Der Praxissektor (gem. Definition der FMH und des Bundesamts für Statistik) ist wesentlich breiter gefasst als die Gruppe der selbständig praktizierenden Ärzt/innen. Zum Praxissektor gehören (siehe Glossar in Hostettler und Kraft, 2021): Konsultations- und Behandlungstätigkeiten von Ärztinnen und Ärzten in Einzel- oder Gruppenpraxen. Inbegriffen sind auch Konsultationstätigkeiten von Privatärzten, die Krankenhäusern angeschlossen sind, sowie Tätigkeiten in Kliniken, Unternehmen, Schulen, Altersheimen, Gewerkschaften und Wohltätigkeitsvereinen.

¹⁴ FHM (2021). Ärztestatistik, Berufstätige Ärzteschaft nach Kanton und Sektor, Stand per 31.12.2020.

3.2.3 Stichprobenziehung (sampling)

Anmerkung. Die RoKo ist keine Stichprobenerhebung, sondern eine Zensusauswahl. Deshalb entfällt dieses Kapitel.

3.2.4 Teilnehmende (respondents)

Wie bereits in Kapitel 3.2.2 erläutert, ist die Teilnahme in 13 Kantonen obligatorisch und in 8 Kantonen freiwillig. Aus der Forschung zur Nonresponse ist bekannt, dass Item- und Unit-Nonresponse (Teilnahme- oder Antwortverweigerung) nicht zufällig, sondern selektiv sind. Mit anderen Worten, Nicht-Antwortende besitzen häufig andere Eigenschaften, als die Antwortenden. Aus den selektiven Antwortausfällen können im Allgemeinen verzerrte Schätzungen resultieren (siehe bspw. Särndal und Lundström, 2005, Kapitel 9).

Es wäre nun zu kurz gedacht, das Bestehen einer Datenlieferungspflicht dahingehend zu interpretieren, dass es keine nennenswerte Unit-Nonresponse geben kann. Wir wollen dies am Beispiel der MAS-Erhebung des Bundesamts für Statistik illustrieren. Obwohl die Statistikerhebungsverordnung des Bundes eine Auskunftspflicht für Arztpraxen, Einrichtungen der ambulanten Krankenpflege verordnet, lag die Teilnehmerate der MAS im Jahr 2019 nur bei 64 %; siehe Exkurs. Im nächsten Abschnitt gehen wir auf die Unit-Nonresponse bei der RoKo genauer ein.

Exkurs. Auskunftspflicht für die MAS-Erhebung und Teilnehmerate

Arztpraxen und Einrichtungen der ambulanten Krankenpflege sind bei der MAS-Erhebung des Bundesamts für Statistik (Strukturdaten der Arztpraxen und ambulanten Zentren) – de jure – zur Auskunft verpflichtet (Obligatorium)¹. Bei der MAS 2019 (Erhebungszeitpunkt 2020/11 bis 2021/2) haben nur 64 % der Befragten teilgenommen. In den Jahren zuvor lag die Teilnehmerate zwischen 52 % und 67 %. Die tiefe Teilnehmerate im Jahr 2020 ist also kein Artefakt der Corona-Pandemie.

¹ Verordnung über die Durchführung von statistischen Erhebungen des Bundes, Anhang 193 (SR 431.012.1).

² BFS (2022). Rapport sur la qualité des données financières. Relevé des données structurelles des cabinets médicaux et des centres ambulatoires MAS, Office fédéral de la statistique, Neuchâtel (Numéro OFS be-f-14.04.05-03).

Analysen zur Unit-Nonresponse

Das Ausmass und die Eigenheiten der Unit-Nonresponse sind kantonale unterschiedlich. Es wäre nicht zielführend, eine Analyse für die gesamte RoKo (ohne kantonale Differenzierung) durchzuführen. Unsere Besprechung beruht auf der Auswahl der Kantone SO, SG, TG und VS. In den ersten drei Kantonen besteht eine Teilnahmepflicht, im Kanton VS ist die Teilnahme freiwillig.

Die Mitgliederverzeichnisse (MV) der kantonalen Ärztesgesellschaften bilden die Rahmengesamtheiten (je Kanton). Die Gruppe der Über-70-Jährigen wird aus den MV ausge-

schlossen.¹⁵ Durch Vergleiche der MV mit den realisierten RoKo-Erhebungsdaten lassen sich die Teilnehmeraten berechnen und allfällige Diskrepanzen infolge von Unit-Nonresponse feststellen.¹⁶ Im Folgenden werden vier Variablen betrachtet, um Vergleiche anstellen zu können. Die Ausprägungen der Variablen sind in Klammer angegeben. Es handelt sich um folgende Variablen: Alterskohorte (31-40, 41-50, 51-60, 61-70, über 70), Spezialisierung (Grund: Grundversorger/innen, Spez: Spezialisten/innen), Geschlecht (M: Männer; F: Frauen) und Urbanität (Stadt oder Land).

Die Teilnehmeraten sind in Tabelle 6 dargestellt. Die niedrige Teilnahmequote im Kanton VS ist (vermutlich) darauf zurückzuführen, dass die Teilnahme nicht obligatorisch war. In den Jahren 2017–2019 liegen die Teilnahmequoten in der gleichen Grössenordnung.

Tabelle 6: Teilnehmerate bei der RoKo nach Kanton

	Kanton SO	Kanton SG	Kanton TG	Kanton VS
Teilnehmerate	74.5 %	65.3 %	67.0 %	27.4 %

Quelle: RoKo-Datenbank (Jahr 2020); eigene Berechnung.

Anmerkung: Die Teilnehmerate ist für das Mitgliederverzeichnis der kantonalen Ärztesgesellschaft berechnet (ohne die Altersgruppe der Über-70-Jährigen).

Die Diskrepanzen zwischen dem MV und den RoKo-Erhebungsdaten (nach Alterskohorte, Spezialisierung, Geschlecht und Urbanität) sind für die Kantone SO, SG und TG in Tabelle 7 ausgewiesen. Die mittlere absolute Abweichung zwischen den Anteilswerten nach MV und den Erhebungsdaten belaufen sich auf 0.91 Prozentpunkte.¹⁷ Mit anderen Worten, die Abweichung bei den Kantonen SO, SG und TG sind gering.

Für den Kantons VS sind die Abweichungen zwischen dem MV und den RoKo-Erhebungsdaten (nach Alterskohorte, Spezialisierung, Geschlecht und Urbanität) in Tabelle 8 ausgewiesen. Es fällt auf, dass es mitunter grosse Diskrepanzen gibt. Die Altersgruppe der 41-50-Jährigen, die Grundversorger/innen, die Frauen und die ländlichen Gebiete sind in der RoKo überrepräsentiert. Eine Über- oder Unterrepräsentation bedeutet aber nicht zwangsläufig, dass die Schätzungen für andere Merkmale verzerrt sind.¹⁸ Darüber hinaus ist eine univariate

¹⁵ Die Mehrheit der Ärztinnen und Ärzte in der Alterskohorte der über 70-Jährigen ist nur noch sporadisch berufstätig. Aus diesem Grund werden sie von den weiteren Untersuchungen ausgeschlossen; siehe dazu auch die Ausführungen in Kapitel 3.2.1.

¹⁶ Die Diskrepanzen könnten auch auf Coverage Errors zurückzuführen sein. Für die anschliessende Kalibrierung ist es unerheblich, ob die festgestellten Abweichungen den Coverage oder Nonresponse Errors zugerechnet werden.

¹⁷ Die mittlere absolute Abweichung wurde als arithmetisches Mittel der absoluten Abweichungen berechnet.

¹⁸ Zum Beispiel: Falls in einer Stichprobe die Anzahl der Linkshänder/innen systematisch überrepräsentiert ist, so ist anzunehmen, dass auch die geschätzte Nachfrage nach Linkshänder-Produkten eine Überschätzung darstellt. Die Schätzungen zu Merkmalen, die keinen direkten Bezug zur

Untersuchung von Einzelvariablen ungeeignet, um Diskrepanzen bei Wechselwirkungen zwischen den Variablen zu entdecken (z. B. Kombination oder Interaktion von Geschlecht und Spezialisierung). Aus diesen Gründen soll nun eine Kalibrierung der Daten für den Kanton VS erfolgen.

Tabelle 7: Diskrepanzen bei den Anteilswerten nach Alter, Spezialisierung etc. zwischen dem Mitgliederverzeichnis der kantonalen Ärztegesellschaften und den RoKo-Daten (für die Kantone SG, SO und TG)

	Alterskohorten				Spezialisierung		Geschlecht		Urbanität	
	31-40	41-50	51-60	61-70	Grund.	Spez.	Mann	Frau	Land	Stadt
Kanton SG										
Mitgliederverzeichnis	3.6%	24.7%	41.8%	29.9%	44.1%	55.9%	70.6%	29.4%	54.1%	45.9%
RoKo	2.7%	25.9%	42.2%	29.2%	45.8%	54.2%	69.8%	30.2%	55.0%	45.0%
Abweichung (PP)	0.9	-1.2	-0.4	0.7	-1.7	1.7	0.8	-0.8	-0.9	0.9
Kanton SO										
Mitgliederverzeichnis	4.4%	20.7%	39.4%	35.5%	53.0%	47.0%	70.5%	29.5%	43.8%	56.2%
RoKo	4.8%	18.7%	39.6%	36.9%	52.9%	47.1%	72.2%	27.8%	43.9%	56.1%
Abweichung (PP)	-0.4	2.0	-0.2	-1.4	0.1	-0.1	-1.7	1.7	-0.1	0.1
Kanton TG										
Mitgliederverzeichnis	8.5%	22.4%	42.9%	26.2%	52.7%	47.3%	62.9%	37.1%	59.9%	40.1%
RoKo	7.6%	23.4%	42.1%	26.9%	53.8%	46.2%	62.9%	37.1%	57.9%	42.1%
Abweichung (PP)	0.9	-1.0	0.8	-0.7	-1.1	1.1	0.0	0.0	2.0	-2.0

Quellen: RoKo-Datenbank und Mitgliederverzeichnisse der kantonalen Ärztegesellschaften, Jahr 2020 (ohne Alterskohorte der Über-70-Jährigen); eigene Berechnungen

Anmerkung: PP: Prozentpunkte

Die Kalibrierung ist ein Verfahren, um mögliche Verzerrungen infolge von Coverage und Non-response Error zu verringern (allenfalls zu beheben); siehe bspw. Särndal und Lundström (2005, Kap. 6). Dabei werden die Erhebungsdaten an die Kennzahlen des Mitgliederverzeichnisses der kantonalen Ärztegesellschaft im Kanton VS angepasst. Es handelt sich um eine mehrdimensionale Anpassung für die fünf Variablen: Alterskohorten (31-40, 41-50, 51-60, 61-70, über 70), Spezialisierung (Grund: Grundversorger/innen, Spez: Spezialisten/innen), Geschlecht (M: Männer; F: Frauen), Urbanität (Stadt oder Land) und Sprache (D: Deutsch, F: Französisch). Für die Kalibrierung wird die Gruppe der Über-70-Jährigen ausgeschlossen.¹⁹

Anmerkung. Die Details zur Kalibrierung sind in Anhang B dokumentiert.

Linkshändigkeit besitzen, sind (ceteris paribus) nicht notwendigerweise verzerrt.

¹⁹ Vgl. dazu die Ausführungen in Kapitel 3.2.1.

Tabelle 8: Diskrepanzen bei den Anteilswerten nach Alter, Spezialisierung etc. zwischen dem Mitgliederverzeichnis der kantonalen Ärztesgesellschaft und den RoKo-Daten für den Kanton Wallis

	Alterskohorten				Spezialisierung		Geschlecht		Urbanität	
	31-40	41-50	51-60	61-70	Grund.	Spez.	Mann	Frau	Land	Stadt
Kanton Wallis										
Mitgliederverzeichnis	10.2%	31.0%	31.5%	27.3%	36.6%	63.4%	59.3%	40.7%	67.6%	32.4%
RoKo	9.9%	39.5%	29.7%	20.9%	48.3%	51.7%	53.5%	46.5%	73.3%	26.7%
Abweichung (PP)	0.3	-8.5	1.8	6.4	-11.7	11.7	5.8	-5.8	-5.7	5.7

Quellen: RoKo-Datenbank und Mitgliederverzeichnis der Walliser Ärztesgesellschaft, Jahr 2020 (ohne Alterskohorte der Über-70-Jährigen); eigene Berechnungen

Anmerkung: PP: Prozentpunkte

Wie bereits eingangs dieses Kapitels erläutert, dient die RoKo als Erhebungsinstrument, um Kenngrößen zu schätzen, die dann für die Parametrisierung des Kostenmodells KOREG verwendet werden (siehe relevante Variablen in Tabelle 3). Der Einfluss der Kalibrierung soll nun anhand der, für die Parametrisierung zentralen Variablen, illustriert werden. In Tabelle 9 sind die geschätzten Mittelwerte mit und ohne Kalibrierung festgehalten. Überdies ist die relative Abweichung in Prozent ausgewiesen (und die Abweichung in Anzahl Standardfehlern).

Tabelle 9: Effekt der Kalibrierung auf den geschätzten Mittelwert der wichtigsten KOREG-Variablen (Kanton VS)

Variable	Mittelwert ¹⁾ (ohne Kalibrierung)	Gew. Mittelwert ²⁾ (mit Kalibrierung)	Relative Abweichung	Abweichung in Anz. Std.-Fehler ³⁾
A1T	Die Einzelwerte sind für ats-tms AG, FMH, curafutura und MTK einsehbar. Sie werden jedoch nicht in der öffentlich zugänglichen Version des Berichts abgedruckt.		4.01%	0.449
A2T			1.15%	0.188
A3T			-0.69%	0.108
A4T			-4.77%	0.301
A5T			2.22%	0.206
B1			2.13%	0.154
E1T			1.41%	0.346
A1T			4.01%	0.449
A2T			1.15%	0.188

Quellen: RoKo-Datenbank Kanton VS, Jahr 2020 (ohne Alterskohorte der Über-70-Jährigen); eigene Berechnungen

Anmerkung: ¹⁾ Arithmetisches Mittel (ohne Vervielfachung der Daten); ²⁾ gewichteter Mittelwert (Gewichtung gem. Kalibrierung); ³⁾ Standardfehler des arithmetischen Mittels.

Aus Tabelle 9 ist ersichtlich, dass die Abweichungen der geschätzten Mittelwerte mit und ohne Kalibrierung in beide Richtungen auftreten. Die mittlere absolute Abweichung beläuft sich auf 2.4 Prozentpunkte.²⁰ Das Ausmass der Diskrepanzen kann auch mithilfe des Standardfehlers des Mittelwertschätzers (zu den Originaldaten) beurteilt werden. Diese Metrik erlaubt es, die den Daten inhärente Variabilität in der Beurteilung der Diskrepanzen zu berücksichtigen. Im Mittel (über alle relevanten Variablen) haben die Diskrepanzen ein Ausmass von 0.27 Standardfehlern. Ruft man sich in Erinnerung, dass das symmetrische 95 %-Vertrauensintervall eine Ausdehnung von knapp ± 2 Standardfehlern besitzt, so erkennen wir, dass die Diskrepanzen zwischen den ungewichteten und gewichteten Schätzern gering sind (und die Differenzen auf dem 5 %-Signifikanzniveau nicht signifikant von null unterschiedlich sind). Wir halten also fest, dass die Kalibrierung zwar einen Einfluss auf die geschätzten Mittelwerte besitzt, dieser jedoch vergleichsweise gering ist.

Analysen zur Item-Nonresponse

Anmerkung. Die detaillierten Untersuchungen zur Item-Nonresponse sind in Anhang B dokumentiert. Fazit: Die Item-Nonresponse tritt bei der RoKo in dem Umfang auf, wie dies auch bei Erhebungen der amtlichen Statistik der Fall ist. Daraus folgt kein weiterer Handlungsbedarf.

Die Ausführungen zu Unit- und Item-Nonresponse haben gezeigt, dass das Ausmass möglicher Verzerrungen begrenzt ist (für die getroffene Auswahl an Kalibrierungsvariablen). Daraus darf aber nicht gefolgert werden, dass die Antwortrate im Kanton VS per se befriedigend ist. Es wäre wünschenswert, eine höhere Antwortrate anzustreben.

Es ist überdies hervorzuheben, dass die möglichen Verzerrungen nur in Bezug auf die getroffene Auswahl an Kalibrierungsvariablen (d. h. alle relevanten Variablen aus dem Mitgliederverzeichnis) beurteilt werden konnte. Es ist nicht möglich zu beurteilen, wie die Verzerrungen ausgefallen wären, wenn weitere Variablen für die Kalibrierung zur Verfügung gestanden hätten.

3.2.5 Abdeckung: Aggregation der kantonalen Grundgesamtheiten

In Kapitel 3.2.2 haben wir abgeschätzt, dass die 21 an der RoKo teilnehmenden Kantone knapp 95 % der Grundgesamtheit abdecken. An dieser Stelle soll die kantonale Zusammensetzung der RoKo detaillierter untersucht werden. Diese Untersuchung beruht auf einem Vergleich zwischen der RoKo und der berufstätigen Ärzteschaft mit Hauptberufstätigkeit im Praxissektor gemäss FMH-Ärztestatistik; letztere wird mit U_{FMH} bezeichnet, resp. mit $U_{\text{FMH}}^{\text{BE}}$, wenn hervorgehoben werden soll, dass nur eine kantonale Teilmenge (hier z. B. Kanton BE) gemeint wird. Der angestrebte Vergleich kann nur eine grobe Annäherung darstellen und unterliegt den bereits in Kapitel 3.2.2 erläuterten Einschränkungen.

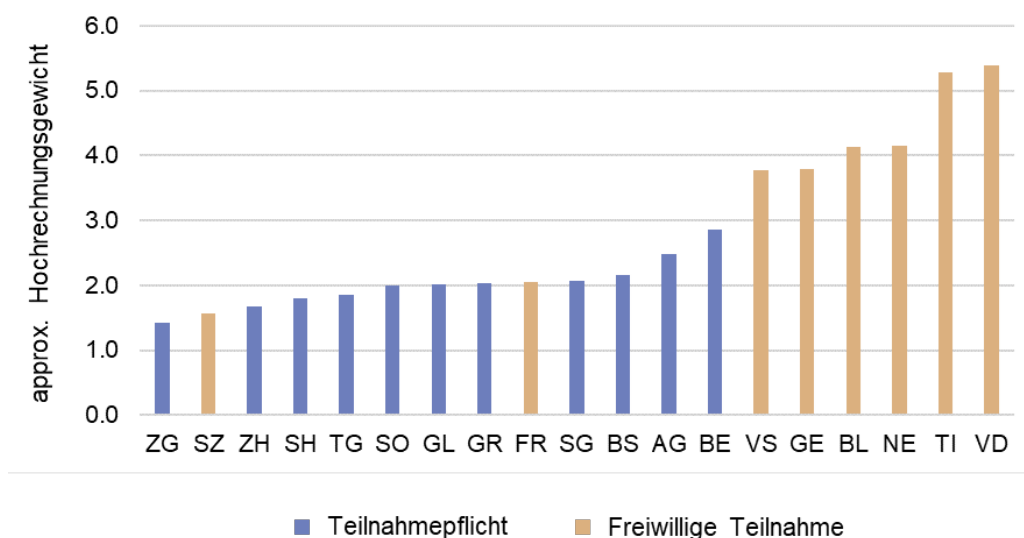
²⁰ Die mittlere absolute Abweichung wurde als arithmetisches Mittel der absoluten Abweichungen berechnet.

Für alle in der RoKo teilnehmenden Kantone wurden sogenannten (approximative) Hochrechnungsgewichte auf der Basis der Grundgesamtheit berechnet. Dies sei am Beispiel des Kantons BE illustriert. Es bezeichne N_{BE} die Anzahl aller Ärztinnen und Ärzte in der FMH-Grundgesamtheit im Kanton BE. Die Anzahl der realisierten Fälle in der kantonalen RoKo-Erhebung wird mit n_{BE} bezeichnet. Das Hochrechnungsgewicht ist durch N_{BE}/n_{BE} gegeben. Es nimmt den Wert 1 an, falls $N_{BE} = n_{BE}$ gilt (d. h. Zensushebung ohne Antwortausfälle). Das Gewicht hat z. B. den Wert 2, wenn n_{BE} nur einem 50 % Anteil von N_{BE} entspricht.

Die approximativen Hochrechnungsgewichte der RoKo-Kantone sind in Abbildung 4 dargestellt. Es wird zwischen Kantonen mit und ohne Teilnahmepflicht unterschieden. Bei Kantonen mit freiwilliger Teilnahme sind die Hochrechnungsgewichte tendenziell höher, weil die Teilnahmebereitschaft geringer ausfiel.

Die Hochrechnungsgewichte können verwendet werden, um kantonale Unterschiede bei den Teilnahmebereitschaften und damit der Abdeckung auszugleichen. Auf der Grundlage der approximativen Hochrechnungsgewichte wurden nun gewichtete Mittelwerte berechnet und mit den nicht gewichteten Mittelwerten verglichen; siehe Tabelle 10. Die Variablen zu den Aufwänden (siehe relevante Variablen in Tabelle 3) wurden zur Variable der gesamthaften Aufwände zusammengefasst und den Erträgen gegenübergestellt. Die Differenz zwischen den Erträgen und den (gesamthaften) Aufwänden ist als Praxisergebnis ausgewiesen.

Abbildung 4: Approximative Hochrechnungsgewichte (je RoKo-Kanton)



Quellen: RoKo-Datenbank, 2017–2019 und FMH-Ärzttestatistik (berufstätige Ärzteschaft im Praxissektor) 2017–2019; eigene Berechnungen.

Der Einbezug der Hochrechnungsgewichte führt (bis auf wenige Ausnahmen) dazu, dass die Mittelwerte der einzelnen Positionen tiefer ausfallen als ohne Gewichtung. Die Abweichungen zwischen der gewichteten und ungewichteten Schätzung belaufen sich in der Grössenordnung von 0.9 %–3.1 % beim Modell X und 5.4 %–7.0 % beim Modell Y (siehe Tabelle 10).

Die Abweichungen sind also klein. Weil die Gewichtung die geschätzten mittleren Aufwände und Erträge in vergleichbarem Ausmass reduziert, fällt der Effekt auf das Praxisergebnis sehr gering aus. Es ist bemerkenswert, dass die Gewichtung beim Modell X zu einer Erhöhung des mittleren Praxisergebnisses führt, beim Modell Y jedoch zu einer Reduktion. Diese Unterschiede dürfen aber nicht überinterpretiert werden, weil die Effektstärke der Gewichtung gering ist; d. h. 1.8 % bzw. -1.2 %.

Tabelle 10: Effekt der Gewichtung (nach Kantonen) auf den Mittelwert der Aufwände, Erträge und des Praxisergebnisses

		Aufwände CHF	Erträge CHF	Praxisergebnis ¹⁾ CHF
Modell X	Mittelwert	114'110	233'319	119'208
	Gewichtetes Mittel	110'613	231'229	120'615
	Abweichung (abs.)	3'497	2'090	-1'407
	Abweichung (in %)	3.1 %	0.9 %	-1.2 %
Modell Y	Mittelwert	471'200	670'824	199'624
	Gewichtetes Mittel	438'425	634'530	196'105
	Abweichung (abs.)	32'775	36'294	3'520
	Abweichung (in %)	7.0 %	5.4 %	1.8 %

Quellen: RoKo-Datenbank, 2017–2018; eigene Berechnung.

Anmerkung: ¹⁾ Das Praxisergebnis berechnet sich als Erträge minus Aufwände.

3.3 Aspekte der Messung

3.3.1 Messung und Messfehler

Messfehler sind bei standardisierten Fragebogenerhebungen vor allem auf Formen der Antwortverzerrung zurückzuführen; vgl. bspw. Schnell et al. (2018, Kapitel 7.1.1.5) und Willis (2020). Derartige Verzerrungen haben vielfältige Ursachen. Bei der Erhebung von sozialstatistischen Merkmalen (z.B. Geschlecht oder Alter) sind die (potenziellen) Verzerrungen durch soziale Erwünschtheit oder Akzeptanz in der Regel geringer als bei Frageitems zu Einstellungen, Bewertungen, Überzeugungen und Verhalten. In der RoKo werden ausschliesslich sozialstatistische Merkmale erhoben, sodass die genannten Ursachen für Verzerrungen nicht vorliegen. Neben merkmalsbezogenen Attributen, können auch Umsetzungsaspekte, wie z. B. Frageeffekte (Formulierung, Logik, Antwortmöglichkeiten und Kontext), zu Verzerrungen führen.

Die im Rahmen der RoKo erhobenen Finanzdaten sind sehr komplex. Die Komplexität rührt daher, dass mit einem Fragebogen versucht wird, Positionen der Erfolgsrechnung und der

Bilanz von Betrieben zu erfassen, die ihrerseits heterogen sind. Wie bei vielen anderen Unternehmensbefragungen (vgl. z. B. Hidiroglou und Lavallée, 2009) zeigt sich auch hier, dass es keine standardisierten Kontenrahmen oder Vorgaben zur Buchführung gibt. Dies kann zu Abgrenzungsproblemen und Verschiebungen in den Teilergebnissen führen, obwohl die Gesamtergebnisse korrekt sind. Eine gezielte Harmonisierung könnte hier vermutlich zu Verbesserungen führen.

3.3.2 Processing

Anmerkung. Die FHNW hat die Datenbearbeitung durch die Ärztekasse vollständig repliziert und überprüft. Die Details dazu (inkl. Sensitivitätsanalysen) sind in Anhang B dokumentiert. Es handelt sich um folgende Processing Steps:

- Datenaufbereitung und -bereinigung; siehe Anhang B.2
- Herausrechnen der Position A215 (tariffremd); siehe Anhang B.3

Es gibt keine nennenswerten Probleme, die an dieser Stelle thematisiert werden müssten.

3.3.3 Editing und Ausreisserentdeckung

Die Verteilungen der relevanten Variablen (siehe Tabelle 3) sind rechtsschief; siehe Anhang B.5. Das arithmetische Mittel ist gegen Ausreisser (an beiden Enden der Verteilung) nicht robust. Die Ärztekasse verwendet darum das symmetrisch 5 %-gestutzte Mittel (engl. trimmed mean). Dadurch werden die Daten unterhalb der 5 %-Perzentile und oberhalb der 95 %-Perzentile gestutzt. Dieses Vorgehen wird separat für die Variablen Gesamtaufwand (AA), totaler Bruttoertrag aus Praxistätigkeiten (E1T) und das Praxisergebnis (Q) angewendet. Für die Berechnung der Mittelwerte zu den relevanten Variablen werden nur Beobachtungen herangezogen, die nicht einer Stutzung der Variablen AA, E1T und QQ unterliegen. Durch die parallele oder gleichzeitige Stutzung (im gleichen Umfang) der Aufwände und Erträge, wird das Verhältnis dieser beiden (für die Analyse zentralen) Grössen nicht beeinträchtigt. Die Wahl des Stutzungsparameters (bei symmetrischer Stutzung) in der Höhe von 5 % ist arbiträr. Im Rahmen einer Sensitivitätsanalyse wurde deshalb der Stutzungsparameter im Wertebereich 2 % – 9 % variiert (siehe Anhang B.5), um dessen Auswirkung auf die Schätzungen zu beobachten. Es zeigte sich, dass die Wahl des Stutzungsparameters innerhalb des Wertebereichs keinen nennenswerten Einfluss auf das geschätzte mittlere Verhältnis der Variablen EE und AA hat.²¹ Das war zu erwarten, weil die Stutzung parallel auf die Aufwände und Erträge angewendet wird.

Durch die symmetrische 5%-Stutzung wird eine vergleichsweise grosse Anzahl von Beobachtungen entfernt. Man könnte dieses Vorgehen als konservativ bezeichnen. Für ein alternatives Verfahren (siehe Anhang B.5) fällt die Anzahl der als Ausreisser deklarierten Beob-

²¹ Sie Anhang B5.2 .

achtungen wesentlich kleiner aus. Weil weniger Beobachtungen ausgeschlossen werden, ist der Umfang der verbliebenen Beobachtungen grösser, wodurch (*ceteris paribus*) eine höhere Effizienz der Schätzer resultiert. Die Effizienzgewinne sind allerdings klein.²²

Anmerkung. Die Details zur Ausreisserentdeckung und -behandlung sind in Anhang B.5 dokumentiert.

3.4 Zwischenfazit

Die RoKo dient – im Kontext der Einführung des TARDOC – als Erhebungsinstrument, um Kenngrössen zu schätzen, die anschliessend für die Parametrisierung des Kostenmodells KOREG verwendet werden. Für die Parametrisierung werden die geschätzten Mittelwerte für die Variablen zu den Aufwänden, Erträgen und Bilanzwerten benötigt.

Die Anwendung des Konzepts der Fitness-for-Use beruht im vorliegenden Kontext darauf, ob die geschätzten Mittelwerte einen möglichst kleinen Mean Square Error (bezgl. des Total Survey Errors) besitzen. Das heisst, ob die Schätzer möglichst unverzerrt und effizient sind. Unsere Besprechung dazu orientierte sich an den in Groves et al. (2009) formulierten Aspekten der Messung (z. B. Processing Errors) und der Repräsentation (z. B. Nonresponse oder Coverage Error).

Die RoKo (2017–2019) deckt die Grundgesamtheit für 21 Kantone ab. Vier Zentralschweizer Kantone (UR, NW, OW und LU) und der Kanton JU sind nicht beteiligt. Die Abdeckung durch die RoKo kann mithilfe der FMH-Ärzttestatistik grob abgeschätzt werden. Auf dieser Basis wird der Anteil der RoKo-Kantone auf 94.9 % aller Ärztinnen und Ärzten der Grundgesamtheit abgeschätzt. In Kantonen mit freiwilliger Teilnahme fiel die Teilnahmebereitschaft in der Tendenz geringer aus als in Kantonen, in welchen die Ärztesgesellschaften ein Teilnahmeobligatorium ausgesprochen hatten. Die freiwillige Teilnahme kann zu Verzerrungen führen – muss es aber nicht. Am Beispiel des Kantons VS haben wir aufgezeigt, dass es mitunter grosse Diskrepanzen bei den Anteilswerten zwischen den RoKo-Erhebungsdaten und dem kantonalen Mitgliederverzeichnis (nach Alter, Spezialisierung, Geschlecht und Urbanität) gibt. Aus dem Vergleich geht hervor, dass die Alterskohorte der 41–50-Jährigen, die Grundversorger/innen, die Frauen und die ländlichen Gebiete in der RoKo des Kantons VS überrepräsentiert sind. Um diese Diskrepanzen genauer abzuschätzen und deren Effekt auf die Schätzung der relevanten Variablen zu beurteilen, wurde eine Kalibrierung für die RoKo im Kanton VS durchgeführt. Die Kalibrierung ist ein Verfahren, um mögliche Verzerrungen infolge von Coverage und Nonresponse Error zu verringern (allenfalls zu beheben). Die geschätzten Mittelwerte der relevanten Variablen (= Parameter für KOREG) wurden ungewichtet (Originaldaten) und mit den kalibrierten Gewichten berechnet. Anschliessend wurden die geschätzten Mittelwerte miteinander verglichen. Die mittlere absolute Abweichung

²² Bei einem Brutto-(Stichproben-)Umfang von 16'000 Beobachtungen (ohne Vervielfachung/ Expansion) fällt es nicht so stark ins Gewicht – was den geschätzten Standardfehler z. B. des Mittelwerts betrifft –, ob 10 % oder 5 % der Beobachtungen ausgeschlossen werden.

zwischen den ungewichteten und gewichteten Mittelwerten beträgt 2.4 Prozentpunkte. Das Ausmass der Diskrepanzen kann auch mithilfe des Standardfehlers des Mittelwertschätzers (zu den Originaldaten) beurteilt werden. Diese Metrik erlaubt es, die den Daten inhärente Variabilität bei der Beurteilung der Diskrepanzen zu berücksichtigen. Im Mittel (über alle relevanten Variablen) haben die Diskrepanzen ein Ausmass von 0.27 Standardfehlern. Ruft man sich in Erinnerung, dass das symmetrische 95%-Vertrauensintervall eine Ausdehnung von knapp ± 2 Standardfehlern besitzt, so erkennen wir, dass die Diskrepanzen zwischen den ungewichteten und gewichteten (d. h. kalibrierten) Schätzern gering sind (und die Differenzen auf dem 5%-Signifikanzniveau nicht signifikant von null unterschiedlich sind). Wir halten also fest, dass die Kalibrierung zwar einen Einfluss auf die geschätzten Mittelwerte besitzt, dieser jedoch vergleichsweise gering ist. Für die RoKo-Daten des Kantons VS erscheint uns – angesichts der geringen Diskrepanzen – eine Kalibrierung im vorliegenden Kontext nicht zwingend. Wir weisen aber darauf hin, dass diese Aussage keine Allgemeingültigkeit besitzt, sondern nur im Kontext der Untersuchungen zu KOREG zutrifft. Für andere Untersuchungen zu den RoKo-Daten des Kantons VS (z. B. Analysen zum Taxpunktwert) gelten andere Anforderungen und die Fitness-for-Use im Zusammenhang mit Coverage und Nonresponse Errors muss separat untersucht werden.

Die bisherige Besprechung fokussierte auf einzelne Kantone. Die RoKo wird jedoch als Erhebungsinstrument für die Schweiz verwendet. Es stellt sich daher die Frage, ob die kantonale Zusammensetzung in der RoKo realitätsgetreu abgebildet ist oder, ob einzelne Kantone bezüglich die Fallzahlen über- oder untervertreten sind. Die Analyse zeigt, dass Kantone, deren Ärztesgesellschaft eine obligatorische Teilnahme beschlossen hatten, im Vergleich zu Kantonen mit freiwilliger Teilnahme überproportional in der RoKo vertreten sind. Um das Ausmass der Überrepräsentation abzuschätzen, wurden approximative Hochrechnungsgewichte berechnet, welche die kantonalen Unterschiede in der Teilnahmebereitschaft und damit in der Abdeckung ausgleichen. Der Einbezug der Hochrechnungsgewichte führt (bis auf wenige Ausnahmen) dazu, dass die geschätzten Mittelwerte tiefer ausfallen als ohne Gewichtung. Die Abweichungen zwischen den gewichteten und ungewichteten Schätzungen sind jedoch überschaubar klein. Weil die Gewichtung die geschätzten mittleren Aufwände und Erträge in vergleichbarem Ausmass reduziert, fällt der Effekt auf das Praxisergebnis sehr gering aus. Diese Erkenntnisse sprechen auch dafür, dass die kantonale Zusammensetzung der RoKo die Schweiz (in Bezug auf die geschätzten Charakteristika) hinreichend gut abbildet.

Die RoKo dient – im Kontext der Einführung des TARDOC – als Erhebungsinstrument, um Kenngrößen zu schätzen, die anschliessend für die Parametrisierung des Kostenmodells KOREG verwendet werden. Für diesen Zweck erachten wir die RoKo – auch im Lichte mangelnder alternativer Datenquellen – als geeignet (im Sinne des Konzepts der Fitness-for-Use bzw. Fitness-for-Purpose).

4 Spartenbetriebsdauern – Datenbank der nationalen Konsolidierungsstelle NAKO

In diesem Kapitel wird das Erhebungsinstrument der Nationalen Konsolidierungsstelle NAKO (NewIndex) charakterisiert und im Hinblick auf die eingangs formulierten Evaluationsfragen untersucht. Die Besprechung orientiert sich an den in Abbildung 2 nach Groves et al. (2009, Abb. 2.4) formulierten Aspekten zur Qualität und zum Total Survey Error.

Metadaten

Contact

Contact organisation	NewIndex
----------------------	----------

Statistical presentation

Statistical unit	Independently practicing physicians
Reference area	Switzerland
Time coverage	2017–2019

Statistical processing

Source data	Data stream of invoices
Frequency of data collection	Continuously
Data collection	Data stream

Accuracy and reliability

Sampling error	Does not apply
Non-sampling error	See below
Coverage error	See below
Measurement error	See below
Non response error	Does not apply
Processing error	See below
Imputation	Does not apply
Seasonal adjustment	Does not apply

Confidentiality

Confidentiality policy	Confidential
------------------------	--------------

4.1 Steckbrief

Die Ärzteneigene Datenbank der Nationalen Konsolidierungsstelle sammelt und aggregiert Informationen zu Leistungsabrechnungen von Arztpraxen. Die Datenbank der NAKO wird von der Firma TrustX Management AG im Auftrag von NewIndex AG betrieben.

Die Ärzteneigene Datenbank der NAKO enthält gegenwärtig rund 330 Millionen TARMED-Abrechnungen der ca. 10'500 teilnehmenden Schweizer Praxen von Ärztinnen und Ärzten. Die einzelnen Abrechnungen nach TARMED, welche die Datenbasis der NAKO darstellen, enthalten Informationen zur Art, dem Zeitpunkt und der Dauer der erbrachten ärztlichen Leistung, dem Abrechnungstarif, der Vergütungsart, dem behandelnden Arzt/ Ärztin, dem Rechnungssteller und zu den Patientinnen und Patienten. Die Abgrenzung der fakturierten Leistungen nach Leistungsdatum erfolgt auf Basis des Sitzungsdatums pro Leistungsposition auf der Rechnung. Dies erlaubt eine eindeutige Zuordnung der fakturierten Leistungen zum Tag der Leistungserbringung unabhängig vom Fakturierungsprozess oder vom Zeitpunkt der Rechnungsanlieferung in die Datensammlung.

Die in die Datenbank der NAKO einflussenden Daten zu den ambulanten ärztlichen Leistungen sind anonymisiert, sodass keine Rückschlüsse auf einzelne Ärzte/-innen oder Patienten/-innen möglich sind.

4.2 Aspekte der Repräsentation

4.2.1 Grundgesamtheit

Definition von Norm-Praxen

Die Auswertung zu den Spartenbetriebsdauern wurde auf die Grundgesamtheit der sogenannten «Normärzte» bzw. «Norm-Praxen» beschränkt. Es handelt sich dabei um Ärztinnen und Ärzte, deren Praxis einen definierten Mindestumsatz an über TARMED abgerechneten OKP-Leistungen überschreitet. Die Anforderungen an den Mindestumsatz sind je nach Modell unterschiedlich, um den Besonderheiten der Leistungserbringer Rechnung zu tragen. Für das Modell X beträgt der jährliche Mindestumsatz 65'000 CHF und für das Modell Y 160'000 CHF. Der Einsatz des Kriteriums zum Mindestumsatz schliesst Artefakte aus, sodass mögliche Verzerrungen vermieden werden. Ausgeschlossen werden damit z. B. Ärztinnen und Ärzte, die ihre Praxis innerhalb des Auswertungszeitraums aufgegeben haben oder die sich im Ruhestand befinden und nur noch sporadisch klinisch tätig sind, sowie Ärztinnen und Ärzte, die erst gegen Ende des Auswertungszeitraums ihre Praxistätigkeit aufgenommen haben.

Einzelne Fachspezialitäten werden aus der Datenbasis ausgeschlossen. Der Ausschluss wird (analog zu RoKo) damit begründet, dass die Leistungen dieser Fachspezialitäten nicht mit dem KOREG-Kostenmodell, sondern mit dem INFRA-Modell vergütet werden.²³ Betroffen sind die Fachspezialitäten Infektiologie, Pathologie und die Radiologie. Zusätzlich wurden

²³ Zur Grössenordnung: Der Ausschluss der besagten Fachspezialitäten führt bei der RoKo zu einer Reduktion der Leistungserbringer um 1.3 %; siehe Anhang B.2. Bei der NAKO liegt der Ausschluss in einer ähnlichen Grössenordnung.

die Praktischen Ärzte²⁴ von der Analyse ausgeschlossen, da sie kein wohldefiniertes Leistungsspektrum besitzen, d. h. nicht nach dem Kostenmodell KOREG gerechnet werden.

Spartenzuordnung und Definition der Spartencluster

Die Zuordnung der Sparten zu den TARMED-Leistungspositionen ist durch die TARMED-Struktur vorgegeben. Jede Leistungsposition ist einer bestimmten Sparte zugeordnet, welche den Ort der Leistungserbringung in der Praxis beschreibt. Die Spartencluster entsprechen den KOREG-Sparten, die für das Kostenmodell TARDOC definiert wurden; siehe Anhang A.

Anzahl Norm-Praxen

Nach Ausschluss der Praxen bzw. Ärztinnen und Ärzten, die nicht der Grundgesamtheit der Norm-Praxen bzw. Norm-Ärzten zugerechnet werden, liegen für die Jahre 2017–2019 Daten zu 4'599 Norm-Praxen vor; siehe Tabelle 11. Dieses Total teilt sich auf in 1'050 Norm-Praxen zum Modell X und 3'549 Norm-Praxen für das Modell Y. Weitere Details können Tabelle 11 entnommen werden.

Tabelle 11: Anzahl Praxen und Norm-Praxen (in der Grundgesamtheit), nach Modell X und Modell Y

		Modelle	
		Modell X	Modell Y
Anzahl Praxen (inkl. Mehrfachzählungen)	2017	1'147	4'457
	2018	1'148	4'434
	2019	1'136	4'308
Praxen in Jahren 2017-2019 (ohne Mehrfachzählungen)		1'232	4'850
Ausgeschlossene Praxen (nicht Teil der Grundgesamtheit)		182	1'301
Norm-Praxen in den Jahren 2017-2019 (Teil der Grundgesamtheit, ohne Mehrfachzählung)		1'050	3'549

Quelle: NAKO/ NewIndex.

Regionale Verteilung

Die an der NAKO-Datensammlung teilnehmenden Praxen sind – mit Ausnahme der Zentralschweiz – relativ gleichmässig über die Schweiz verteilt; siehe Tabelle 12.

²⁴ Inhaber des eidgenössischen Weiterbildungstitels «Praktischer Arzt / Praktische Ärztin» verfügen am Ende ihrer Weiterbildung über die Kompetenz, eigenverantwortlich im Bereich der medizinischen Grundversorgung tätig zu sein. Dieser Titel entspricht jedoch nicht einer fachärztlichen Kompetenz (wie z. B. der Facharzttitel für Allgemeine Innere Medizin).

Tabelle 12: Regionale Verteilung der NAKO-Praxen nach Grossregion (in Bezug auf die Anzahl der im ambulanten Sektor tätigen Ärztinnen und Ärzte)

Espace Mittelland	Genfersee- region	Nordwest- schweiz	Ostschweiz	Tessin	Zentral- schweiz	Zürich
20%	18%	23%	26%	28%	3%	30%

Quelle: NAKO/ NewIndex, 2017–2019, Bundesamt für Statistik (2019): Bestand und Dichte der Ärzt/innen, Zahnarztpraxen und Apotheken nach Kanton

Anmerkung: Die Anteilswerte sind in Bezug auf die Anzahl der Ärztinnen und Ärzte im ambulanten Sektor (gem. BFS) ausgewiesen.

Die ausgewiesenen Anteilswerte sind nur geeignet, um die Grossregionen untereinander zu vergleichen. Die Anteilswerte sind untauglich, um die Abdeckung der NAKO in globo abzuschätzen.

Die Berechnung der Anteilswerte in Tabelle 12 zielt nur darauf ab, die Grossregionen untereinander zu vergleichen. Diese Anteilswerte sind nicht geeignet (bzw. verzerrt), um die Abdeckung der NAKO abzuschätzen. Die Anteilswerte sind in Bezug zur Gesamtheit aller im ambulanten Sektor tätigen Ärztinnen und Ärzte ausgewiesen. Die Definition dieses Sektors ist wesentlich weiter gefasst als die Gesamtheit der selbständig praktizierenden Ärztinnen und Ärzte im ambulanten Bereich (ohne Spitalambulatorien). Daraus resultiert, dass die Anteilswerte in Tabelle 12 eine (deutliche) Unterschätzung der Abdeckung darstellen.

Ermittlung der Betriebsdauern

Die Spartenbetriebsdauern werden entsprechend der TARMED-Tarifstruktur ermittelt. Im TARMED ist für jede Leistungsposition eine Dauer (Minutage) hinterlegt, welche jeder einzelnen Komponente der Leistungserbringung den spezifischen Zeitbedarf (z. B. Raumbellegungszeit) angibt.

In der vorliegenden Analyse wird für jede TARMED-Position die Betriebsdauer als Summe aus Raumbellegungszeit und Wechselzeit berechnet. Für jede Norm-Praxis wird anschliessend die Summe aller im Auswertungszeitraum abgerechneten Leistungspositionen ermittelt. Im nächsten Schritt wird die Betriebsdauer pro Leistungsposition und Praxis als Produkt aus Menge und Summe Betriebsdauer (in Minuten) berechnet und durch die Anzahl der Normpraxen dividiert.

Die Betriebsdauern und die abgeleiteten Kennzahlen (z. B. Mittelwerte) beziehen sich auf ein (Geschäfts-) Jahr.

4.2.2 Stichprobenziehung (sampling)

Anmerkung. Die NAKO ist keine Stichprobenerhebung. Deshalb entfällt dieses Kapitel.

4.2.3 Teilnehmende (respondents)

Anmerkung. Die NAKO kennt keine Unit- und Item-Nonresponse, wie dies bei Befragungen der Fall ist. Deshalb entfällt dieses Kapitel.

4.3 Aspekte der Messung

Die Tarifierung erzeugt in einigen Fällen Artefakte, die sich z. B. in sehr geringen Betriebsdauern in bestimmten Fachgebieten bemerkbar machen. In vielen Fällen betrifft dies Tarifpositionen, die für alle Fachrichtungen buchbar sind und daher insgesamt häufig verwendet, aber nur selten pro Arzt/Ärztin abgerechnet werden.

Diese Artefakte treten nicht in allen Sparten auf, sondern betreffen nur die Sparten in den Spartenclustern Röntgen, UBR 2, UBR 3, UBR 4 und Praxis-Ultraschall. Wir wollen dies anhand von zwei Beispielen erläutern. Tabelle 13 zeigt eine Beispielpraxis, in der die Betriebsdauern in den Spartenclustern Röntgen, UBR_2_Allgemein und UBR_3_Allgemein sehr kleine Werte aufweisen, namentlich 11.93 Min., 5.28 Min. und 2.33 Min. (pro Jahr). Diese Betriebsdauern sind vermutlich Artefakte der Tarifierung. Aus diesem Grund werden sie aus dem Datensatz entfernt. Der Ausschluss wird durch eine Stutzung (engl. trimming) umgesetzt, wie sie auch bei der RoKo angewendet wird.

Tabelle 13: Beispielpraxis aus den Rohdaten zur Illustration der Betriebsdauer pro Spartencluster (Basis: ein Geschäftsjahr)

Modell	Spartencluster	Betriebsdauer (Min.)	Klassen*
ModY	Röntgen	11.93	0050
ModY	Sprechzimmer_Allgemein	1'193.08	1200
ModY	Sprechzimmer_Psychiatrie	0.03	0050
ModY	UBR_1_Allgemein	894.09	0900
ModY	UBR_2_Allgemein	5.28	0050
ModY	UBR_3_Allgemein	2.33	0050
ModY	Ultraschall	74.47	0100

Daten: NAKO/ NewIndex; eigene Darstellung.

Anmerkung: *Klasse = Zuordnung Verteilungsklassen innerhalb Spartencluster

Farbcodes: [Feld] meint den Ausschluss aus dem Spartencluster wegen geringer Betriebsdauer da Tarifierung nicht dem typischen Leistungsspektrum der Praxis entspricht

Das zweite Beispiel ist in Tabelle 14 dokumentiert. Dort treten ebenfalls sehr kleine Betriebsdauern auf. Überdies wird für den Spartencluster UBR_2_Allgemein eine Dauer von 2'328.43

Min. ausgewiesen. Es handelt sich hierbei um einen Ausreisser, weil eine Betriebsdauer von über 2'300 Minuten (mit grösster Wahrscheinlichkeit) nicht durch eine Norm-Ärztin bzw. einen Norm-Arzt (d. h. ein Vollzeit-Äquivalent) geleistet werden kann. Folgerichtig werden derartige Beobachtung aus dem Datensatz entfernt.

Tabelle 14: Beispielpraxis aus den Rohdaten zur Illustration der Betriebsdauer pro Spartencluster (Basis: ein Geschäftsjahr)

Modell	Spartencluster	Betriebsdauer (Min.)	Klassen*
ModY	Sprechzimmer_Allgemein	975.44	1000
ModY	UBR_1_Allgemein	48.02	0050
ModY	UBR_2_Allgemein	2'328.43	grösser
ModY	UBR_3_Allgemein	24.27	0050
ModY	UBR_4_Allgemein	113.89	0150
ModY	Ultraschall	7.09	0050

Daten: NAKO/ NewIndex; eigene Darstellung.

Anmerkung: *Klasse = Zuordnung Verteilungsklassen innerhalb Spartencluster

Farbcodes: [Feld] meint den Ausschluss aus dem Spartencluster wegen geringer Betriebsdauer da Tarifierung nicht dem typischen Leistungsspektrum der Praxis entspricht

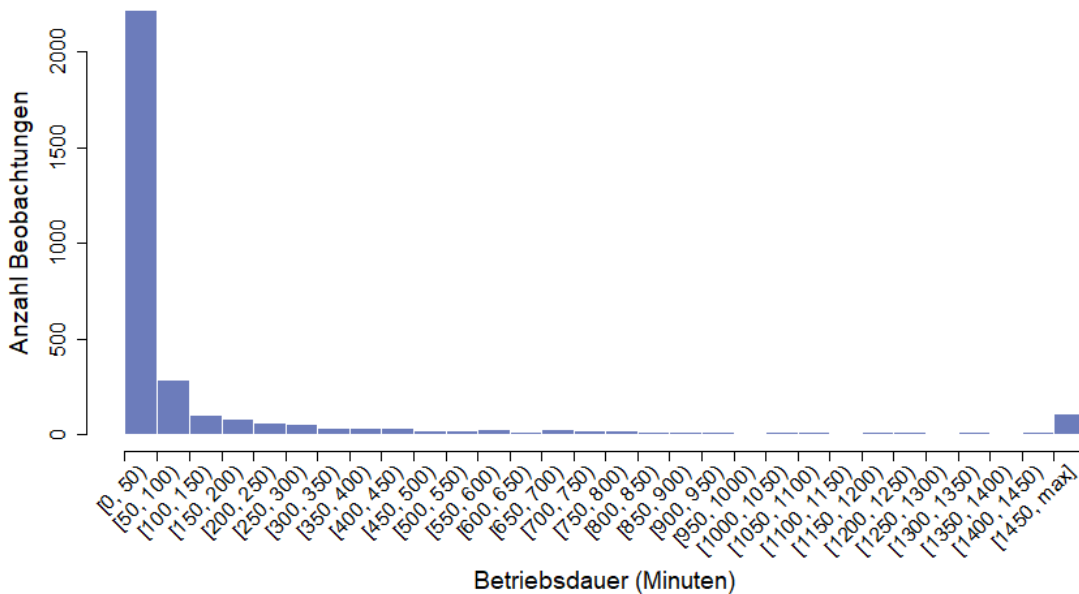
[Feld] meint den Ausschluss aus dem Spartencluster wegen sehr hoher Betriebsdauer die nicht durch ein einzelnes Vollzeit-Äquivalent erbracht werden kann

Artefakte sind auch in der (gesamthaften) Verteilung der Betriebsdauern erkennbar. Dies ist für das Beispiel des Spartenclusters «UBR 2 Allgemein» (Modell Y) in Abbildung 5 dargestellt. Diese Abbildung zeigt zwei Histogramme zu den gleichen Daten, die sich jedoch hinsichtlich der Intervallbildung unterscheiden. Das erste Histogramm (Panel a) beruht auf einer Intervallgrösse von 50 Min. Die Verteilung ist beim Wert 1'450 Min. zensiert und alle Werte oberhalb dieser Schwelle wurden dem Abschlussintervall zugeteilt. Das zweite Histogramm (Panel b) basiert auf Intervallen mit einer Ausdehnung von 200 Minuten, dafür wurde die Verteilung am oberen Ende nicht zensiert.

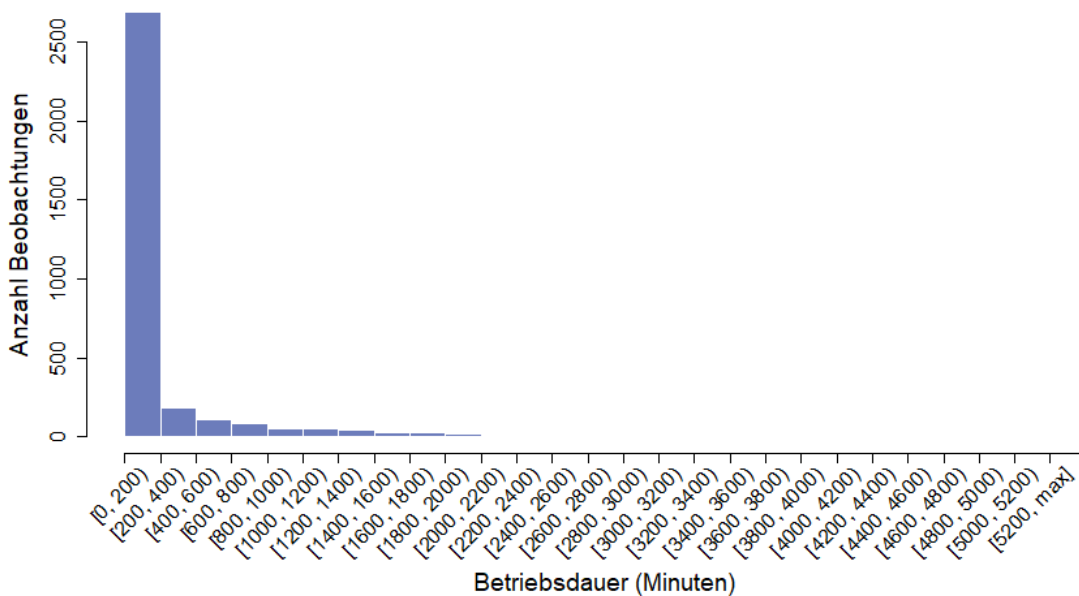
Aus Abbildung 5 ist ersichtlich, dass Betriebsdauern in der Grössenordnung 0–50 Min. sehr häufig abgerechnet wurden. Hierbei handelt es sich zu weiten Teilen um Fälle, die auf Artefakte der Tarifierung zurückzuführen sind. Aus der Verteilung geht auch hervor, dass die Betriebsdauern sehr rechts schief verteilt sind und Ausreisser aufweisen.

Abbildung 5: Verteilung Betriebsdauern für den Spartencluster «UBR 2 Allgemein» (Modell Y) pro Jahr

a) Histogramm



b) Histogramm (gleiche Daten, aber gröbere Intervalle)



Quelle: NAKO/ NewIndex, 2017–2019; eigene Darstellung.

Ermittlung der Betriebsdauern

Die am Beispiel des Spartenclusters «UBR 2 Allgemein» (Modell Y) illustrierte Verteilung der Betriebsdauern (siehe Abbildung 5) tritt in ähnlicher Form auch bei den Spartenclustern

Röntgen, UBR 2, UBR 3, UBR 4 und Praxis-Ultraschall auf. Es ist darum auch bei diesen Spartenclustern erforderlich, die Artefakte und Ausreisser regelbasiert auszuschliessen.

Der Ausschluss von Beobachtungen basiert mehrheitlich auf einer symmetrischen Stutzung im Umfang von 5 % (engl. trimming). Dabei werden die 5 % kleinsten und die 5 % grössten Werte gestutzt. Einige wenige Regeln zur Stutzung basieren auf einer normativen Festlegung, wie z. B. die Festsetzung einer Untergrenze bei 50 Minuten; siehe Tabelle 15.

Tabelle 15: Grenzwerte zur Definition der Norm-Praxen (bzw. Ausschluss von Praxen, die nicht zur Grundgesamtheit gehören)

Spartencluster	Grenzwerte (Zeit: Minuten)		Anmerkung
	5%-Quantil	95%-Quantil	
<i>Modell X</i>			
Sprechzimmer Psychiatrie	330	1'859	-
UBR Nichtärztlich Psychiatrie	0	3'446	(siehe Anm.)
<i>Modell Y</i>			
Röntgen	50	176	fixiert bei 50
Sprechzimmer Allgemein	438	2'517	-
UBR 1 Allgemein	43	1'014	-
UBR 2 Allgemein	50	1'200	fixiert bei 50 und 1'200
UBR 3 Allgemein	50	1'200	fixiert bei 50 und 1'200
UBR 4 Allgemein	50	1'164	fixiert bei 50
UBR Delegierte Psychiatrie Allgemein	0	1'664	-
UBR Nichtärztlich Allgemein	0	1'288	-
Ultraschall	50	684	fixiert bei 50

Anmerkung: Ein Grenzwert wird als «fixiert» bezeichnet, wenn er normativ vorgegeben wurde und nicht durch das entsprechende empirische Perzentil definiert ist.

Die psychologische Psychotherapie ist nicht (mehr) Gegenstand von TARMED/ TARDOC.

Quelle: NewIndex.

Die Grenz- oder Schwellenwerte, die auf einer normativen Festlegung abstellen, sind nur unwesentlich von den jeweiligen 5 %- bzw. 95 %-Quantilen verschieden. Die Verteilungen der Betriebsdauern zu den übrigen Spartenclustern (nach der Stutzung) sind in Anhang C dokumentiert.

4.4 Zwischenfazit

Die Betriebsdauerdaten der NAKO werden – im Kontext der Einführung des TARDOC – verwendet, um die mittleren Betriebsdauern zu schätzen, die anschliessend für die Parametrisierung des Kostenmodells KOREG verwendet werden.

Die NAKO ist ein Data Stream of Invoices. Deshalb gibt es keine potenziellen Verzerrungen infolge von Unit- oder Item-Nonresponse, wie dies bei Befragungen der Fall sein kann. Weil die NAKO überdies keine Stichprobenerhebung nach dem Zufallsprinzip ist, entfallen Untersuchungen zu Stichprobenfehler und Total Survey Error gänzlich. Die Auswertungen zu den Spartenbetriebsdauern wurde auf die Grundgesamtheit der sogenannten «Normärzte/-innen» bzw. «Norm-Praxen» beschränkt, so dass sie mit den Kenngrössen der RoKo kompatibel sind. Der Ausschluss von Beobachtungen bei den Auswertungen zur NAKO ist eng auf die RoKo abgestimmt.

5 Alignment zwischen RoKo und NAKO

Die Erhebungsinstrumente RoKo und NAKO sind nicht miteinander gekoppelt. Es handelt sich um separate Erhebungsinstrumente. Beide Erhebungsinstrumente erfassen nur anonymisierte Daten, sodass es nicht möglich ist, mithilfe eines Identifikators einzelne Subjekte zwischen den Erhebungen zu identifizieren.

Da es sich um unterschiedliche Erhebungsinstrumente handelt, stellt der Prüfbericht des Bundesamts für Gesundheit infrage, ob «Gleiches mit Gleichem» verglichen werde (siehe Prüfbericht, S. 55–56). Die Unterschiedlichkeit der Erhebungsinstrumente stellt an sich noch keine Beeinträchtigung der Vergleichbarkeit dar. Entscheidend ist, ob die Erhebungsinstrumente geeignet sind, möglichst unverzerrte und effiziente Schätzungen von Charakteristika/Parametern zur Grundgesamtheit zu liefern. Mit anderen Worten, es genügt zu prüfen, ob jedes einzelne Erhebungsinstrument diese Anforderungen erfüllt. Diese Untersuchungen wurden in den Kapiteln 3 und 4 geleistet (inkl. Diskussion allfälliger Einschränkungen). Greift man die dort dargestellten Erkenntnisse auf, so kann festgehalten werden, dass die beiden Erhebungsinstrumente (im Kontext der KOREG) ein hinreichend gutes Alignment aufweisen.

Tabelle 16: Relative Differenzen zwischen dem Umsatz (gem. NAKO) und dem Gesamtertrag (gem. RoKo)

Modell	Umsatz (NAKO, CHF)	Gesamtertrag (RoKo, CHF)		Relative Differenz
		E1T	EE	
X	248'612	230'414		7.3 %
X	248'612		238'296	4.1 %
Y	710'719	650'087		8.5 %
Y	710'719		687'505	3.3 %

Quelle: RoKo-Datenbank und NAKO (2017–2019).

Wir wollen den Grad des Alignments an einem zentralen Beispiel illustrieren. Sowohl RoKo als auch NAKO können genutzt werden, um den mittleren Ertrag/ Umsatz einer Norm-Ärztin/ eines Norm-Arztes (= Arbeitsbegriff aus der NAKO) abzuschätzen. Es handelt sich nur um eine grobe Abschätzung, weil die beiden Erhebungsinstrumente unterschiedliche Definitionen unterstellen.²⁵ In Tabelle 16 sind die Umsätze der NAKO und die Erträge der RoKo (es gibt zwei Ausprägungen: EE oder E1T) einander gegenübergestellt, wobei nach den Modellen X und Y unterschieden wird. Trotz definitorischer Unterschiede gibt es bei den Abschätzungen zum mittleren Umsatz bzw. Ertrag keine unerklärbar grossen Diskrepanzen. Dies kann als Indiz aufgefasst werden, dass die beiden Erhebungsinstrumente, trotz unterschied-

²⁵ Die NAKO umfasst nur die Umsätze von Leistungen, die über die OKP abgerechnet wurden. Die RoKo erfasst Erträge/ Umsätze im weiteren Sinn (z. B. Erträge aus konsiliarischer Tätigkeit, Erträge aus Gutachtertätigkeit, Honorare für Vorträge usw.).

licher Basis, zu ähnlichen Abschätzungen kommen (wie dies für den Barwert der Erträge resp. Umsätze exemplarisch gezeigt wurde).

6 Schlussfolgerungen und Empfehlungen

Der Prüfbericht des BAG hat kritische Punkte und Empfehlungen zum Kostenmodell KOREG und die Erhebungsinstrumente (RoKo und Spartenbetriebsdauern) ausgesprochen. Der vorliegende Bericht zur Evaluation greift die kritischen Punkte und Empfehlungen auf und diskutiert diese aus einer methodisch/ statistischen Perspektive.

Die Kritikpunkte und Anmerkungen aus dem Prüfbericht lassen sich in zwei Gruppen aufteilen:

- Forderung nach Herstellung vollständiger Transparenz (inkl. transparente Herleitung sämtlicher Parameter);
- Untersuchungen zu Repräsentativität (einschliesslich möglicher Verzerrungen durch präferenzielle Nonresponse/ Teilnahme an den Erhebungen und mögliche Verzerrungen selbst deklarerter Angaben, sowie durch die Verwendung unterschiedlicher Datenquellen).

Der vorliegende Bericht kommt der Forderung nach Transparenz nach. Er gibt die Datenverarbeitung der RoKo und der NAKO wieder, reproduziert, analysiert und bewertet sie. Die Detailanalysen sind in den Anhängen dokumentiert.

Die Untersuchung zur «Repräsentativität» orientiert sich an einem Quality Framework zum Total Survey Error. Dieses Quality Framework ist in der Forschung zur Stichprobentheorie weitverbreitet und Teil des European Statistical Systems, dem auch die amtliche Statistik in der Schweiz untersteht. Das Framework stellt das Konzept der Fitness-for-Use in den Vordergrund, um zu beurteilen, ob ein Schätzer oder eine Statistik sich in einem bestimmten Kontext für statistische Inferenz eignet. Die Beurteilung der Eignung ist kriterienbasiert. Die Orientierung an einem Quality Framework drängt sich auf, weil der Begriff der «Repräsentativität» – entgegen einer weitverbreiteten Meinung – kein Begriff aus der Stichprobentheorie darstellt und dessen Verwendung, legt man wissenschaftliche Kriterien zugrunde, «ungenau und unnötig» (Schnell et al., 2018, S. 298) ist.

Kostenmodell KOREG: Der Kontext

RoKo und NAKO dienen – im Kontext der Einführung des TARDOC – als Erhebungsinstrumente, um Kenngrössen zu schätzen, die anschliessend für die Parametrisierung des Kostenmodells KOREG verwendet werden. Die Anwendung des Konzepts der Fitness-for-Use beruht im vorliegenden Kontext darauf, ob die geschätzten Mittelwerte zu den relevanten Variablen einen möglichst kleinen Mean Square Error (bezgl. des Total Survey Errors) besitzen. Das heisst, ob die Schätzer möglichst unverzerrt und effizient sind. Die Erhebungsinstrumente RoKo und NAKO wurden hierzu separat untersucht.

Rollende Kostenstudie

Die Grundgesamtheit der RoKo umfasst die berufstätigen, selbständig praktizierenden Ärztinnen und Ärzte im ambulanten Bereich (ohne Spitalambulatorien) und konstituiert sich für jeden Kanton separat (gekoppelt an die Erteilung der Berufsausübungsbewilligungen durch die Kantone). Die Zugehörigkeit zu einer kantonalen Grundgesamtheit wird an zwei Merkmalen des TARMED festgemacht: A1) Die Ärztin/ der Arzt besitzt eine Berufsausübungsbewilligung zur Erbringung von OKP-Leistungen. A2) Sie/ er ist Mitglied der kantonalen Ärztesgesellschaft (inkl. Beitritt zum nationalen Rahmenvertrag TARMED und dem kantonalen Tarifvertrag zwischen der Ärztesgesellschaft und den Versicherern) oder sie/ er ist Nicht-Mitglied, hat sich jedoch dem Rahmen- und Anschlussvertrag des TARMED angeschlossen. Die Definition der (Ziel-) Grundgesamtheit wird an die Mitgliedschaft/ Partizipation in der kantonalen Ärztesgesellschaft geknüpft. Nur Mitglieder der kantonalen Ärztesgesellschaften werden der Zielgrundgesamtheit zugerechnet. Nichtmitglieder der kantonalen Ärztesgesellschaft (und somit nicht Mitglied der FMH) partizipieren nicht an der Aushandlung der kantonalen Tarifverträge. Sie «schweigen» und schliessen sich dem verhandelten Vertrag an. Vieles spricht dafür, dass es sich um ein qualifiziertes Schweigen oder zustimmendes Schweigen handelt.

Die RoKo (2017–2019) deckt die Grundgesamtheit für 21 Kantone ab. Die vier Zentralschweizer Kantone UR, NW, OW und LU und der Kanton JU sind nicht beteiligt. Die Abdeckung durch die RoKo kann mithilfe der FMH-Ärztestatistik grob abgeschätzt werden. Sie beträgt 94.9 % aller Ärztinnen und Ärzten der Grundgesamtheit.

In Kantonen mit freiwilliger Teilnahme an der RoKo fiel die Teilnahmebereitschaft in der Tendenz geringer aus als bei Kantonen mit einer Teilnahmepflicht. Das kann zu Verzerrungen führen – muss es aber nicht. Am Beispiel des Kantons VS haben wir aufgezeigt, dass es mitunter grosse Diskrepanzen bei den Anteilswerten zwischen den RoKo-Erhebungsdaten und dem kantonalen Mitgliederverzeichnis (nach Alter, Spezialisierung, Geschlecht und Urbanität) gibt. Um diese Diskrepanzen genauer abzuschätzen und deren Effekt auf die Schätzung der relevanten Variablen zu beurteilen, wurde eine Kalibrierung für die RoKo im Kanton VS durchgeführt. Die Kalibrierung ist ein Verfahren, um mögliche Verzerrungen infolge von Coverage und Nonresponse Error zu verringern (allenfalls zu beheben). Die geschätzten Mittelwerte der relevanten Variablen (= Parameter für KOREG) wurden ungewichtet (Originaldaten) und mit den kalibrierten Gewichten berechnet. Die mittlere absolute Abweichung zwischen den ungewichteten und gewichteten Mittelwerten beträgt nur 2.4 Prozentpunkte. Das Ausmass der Diskrepanzen kann auch mithilfe des Standardfehlers des Mittelwertschätzers (zu den Originaldaten) beurteilt werden. Diese Metrik erlaubt es, die den Schätzern inhärente Variabilität in der Beurteilung der Diskrepanzen zu berücksichtigen. Im Mittel (über alle relevanten Variablen) haben die Diskrepanzen ein Ausmass von 0.27 Standardfehlern. Wir halten also fest, dass die Kalibrierung zwar einen Einfluss auf die geschätzten Mittelwerte besitzt, dieser jedoch vergleichsweise gering ist. Für die RoKo-Daten des Kantons VS erscheint uns – angesichts der geringen Diskrepanzen – eine Kalibrierung im vorliegenden Kontext nicht zwingend. Wir weisen aber darauf hin, dass diese Aussage keine Allgemeingültigkeit besitzt, sondern nur im Kontext der Untersuchungen zu KOREG zutrifft. Für andere Untersuchungen zu den RoKo-Daten des Kantons VS (z. B. Analysen zum Taxtpunktwert)

gelten andere Anforderungen und die Fitness-for-Use im Zusammenhang mit Coverage und Nonresponse Errors muss separat untersucht werden.

Die RoKo wird im Kontext der Untersuchungen zu KOREG als Erhebungsinstrument für die Schweiz verwendet. Es stellt sich also die Frage, ob die kantonale Zusammensetzung in der RoKo getreu abgebildet ist. Die Analyse zeigte, dass Kantone, deren Ärztesgesellschaft eine Teilnahmepflicht ausgesprochen hat, in der RoKo überproportional vertreten sind. Um das Ausmass der Übervertretung abzuschätzen, wurden approximative Hochrechnungsgewichte berechnet, welche die kantonalen Unterschiede in der Teilnahmebereitschaft und damit in der Abdeckung ausgleichen. Die Abweichungen zwischen gewichteter und ungewichteter Schätzung sind jedoch gering und liegen in der Grössenordnung von 0.9 %–3.1 % beim Modell X bzw. 5.4 %–7.0 % beim Modell Y. Weil die Gewichtung die geschätzten mittleren Aufwände und Erträge in vergleichbarem Ausmass reduziert, fällt der Effekt auf das Praxisergebnis gering aus. Diese Erkenntnisse sprechen dafür, dass die kantonale Zusammensetzung der RoKo die Schweiz (in Bezug auf die geschätzten Charakteristika) hinreichend gut abbildet.

Die RoKo dient – im Kontext der Einführung des TARDOC – als Erhebungsinstrument, um Kenngrössen zu schätzen, die anschliessend für die Parametrisierung des Kostenmodells KOREG verwendet werden. Für diesen Zweck erachten wir die RoKo – auch im Lichte mangelnder alternativer Datenquellen – als geeignet (im Sinne des Konzepts der Fitness-for-Use bzw. Fitness-for-Purpose).

Spartenbetriebsdauern NAKO/ NewIndex

Die Betriebsdauerdaten der NAKO werden – im Kontext der Einführung des TARDOC – verwendet, um die Betriebsdauern zu schätzen, die anschliessend für die Parametrisierung des Kostenmodells KOREG verwendet werden. Die NAKO ist ein Data Stream of Invoices. Deshalb gibt es keine potenziellen Verzerrungen infolge von Unit- oder Item-Nonresponse, wie dies bei Befragungen der Fall sein kann. Weil die NAKO überdies keine Stichprobenerhebung nach dem Zufallsprinzip ist, entfallen Untersuchungen zu Stichprobenfehler und Total Survey Error gänzlich.

Alignment von RoKo und NAKO

Die Erhebungsinstrumente RoKo und NAKO sind nicht direkt miteinander gekoppelt. Es handelt sich um separate Erhebungsinstrumente. Beide Erhebungsinstrumente erfassen nur anonymisierte Daten, sodass es nicht möglich ist, mithilfe eines Identifikators einzelne Subjekte zwischen den Erhebungen zu identifizieren.

Die Unterschiedlichkeit der Erhebungsinstrumente stellt an sich noch keine Beeinträchtigung der Vergleichbarkeit dar. Entscheidend ist, ob die Erhebungsinstrumente geeignet sind, möglichst unverzerrte und effiziente Schätzungen von Charakteristika/ Parametern zur Grundgesamtheit zu liefern. Mit anderen Worten, es ist ausreichend zu prüfen, ob jedes Erhebungsinstrument diesen Anforderungen genügt. Diese Untersuchungen wurden getrennt durchgeführt (inkl. Diskussion allfälliger Einschränkungen). Aufgrund der gewonne-

nen Erkenntnisse kann festgehalten werden, dass die beiden Erhebungsinstrumente (im Rahmen von KOREG) ein ausreichend gutes Alignment aufweisen.

Die Erhebungsinstrumente RoKo und NAKO sind je für sich gesehen von guter Qualität. Die Abweichungen der Instrumente sind vor dem Hintergrund der Komplexität der zu eruierten Phänomene klein. Beide Instrumente zielen im Wesentlichen auf dieselbe Grundgesamtheit ab und ergeben für Schlüsselindikatoren vergleichsweise geringe Abweichungen. Damit wird auch die Kombination der beiden Instrumente zu einem hinreichend validen Instrument. Verbesserungsmöglichkeiten bestehen immer. Die Antwortrate der Kantone ohne Teilnahmepflicht könnte durch eine Pflicht verbessert werden. Durch die Beteiligung der Zentralschweizer Kantone und des Kantons JU, die nicht an der RoKo partizipieren, könnte die schweizweite Abdeckung verbessert werden. Wünschenswert wäre auch eine breitere Abstützung der NAKO-Daten durch eine stärkere Beteiligung der Ärztinnen und Ärzte.

Literaturverzeichnis

- Agresti, A. (2002). *Categorical Data Analysis*. Hoboken (NJ): John Wiley & Sons, 2. Aufl.
- BAG – Bundesamt für Gesundheit (2020). *Tarifstruktur TARDOC 1.0 und 1.1: Ergebnisse der formellen und materiellen Prüfung*. Bericht des BAG vom 19. November 2020. Bern.
- BAG – Bundesamt für Gesundheit (2021). *Ärztinnen und Ärzte*. Bundesamt für Gesundheit (Direktionsbereich für Gesundheitsberufe und Verbraucherschutz). Bern.
- BFS – Bundesamt für Statistik (2018). *Erste Erhebung «Strukturdaten der Arztpraxen und ambulanten Zentren» (MAS 2015): Analyse von Teilnahme und Grundgesamtheit*, Bundesamt für Statistik, Neuchâtel (BFS-Nummer be-d-14.04.05-01).
- BFS – Bundesamt für Statistik (2022). *Rapport sur la qualité des données financières. Relevé des données structurelles des cabinets médicaux et des centres ambulatoires MAS*, Office fédéral de la statistique, Neuchâtel (Numéro OFS be-f-14.04.05-03).
- Biemer, P. P. und Lyberg, L. E. (2003). *Introduction to Survey Quality*. Hoboken (NJ): John Wiley & Sons.
- Billor, N., Hadi, A. S. und Vellemann, P. F. (2000). BACON: Blocked adaptive computationally-efficient outlier nominators. *Computational Statistics and Data Analysis* 34, S. 279–298. [Doi: https://doi.org/10.1016/S0167-9473\(99\)00101-2](https://doi.org/10.1016/S0167-9473(99)00101-2)
- De Waal, T. (2009). *Statistical Data Editing*, in: *Sample Surveys: Theory, Methods and Inference* hrsg. v. Pfeffermann, D. und Rao, C. R., Band 29A: *Handbook of Statistics*, Amsterdam: Elsevier, Kap. 9, S. 187–214. [Doi: https://doi.org/10.1016/S0169-7161\(08\)00009-6](https://doi.org/10.1016/S0169-7161(08)00009-6)
- Deville, J.-C. und Särndal, C.-E. (1992). Calibration Estimators in Survey Sampling, *Journal of the American Statistical Association* 87, S. 376–382. [Doi: https://doi.org/10.1080/01621459.1992.10475217](https://doi.org/10.1080/01621459.1992.10475217)
- Dhar, S. S. und P. Chaudhuri (2012). On the Derivatives of the Trimmed Mean, *Statistica Sinica* 22, S. 655–679. [URL: https://www.jstor.org/stable/24310029](https://www.jstor.org/stable/24310029)
- Diekmann, A. (2009). *Empirische Sozialforschung: Grundlagen, Methoden, Anwendungen*. Reinbek bei Hamburg: Rowohlt-Taschenbuch-Verlag, 20., vollst. überarb. und erw. Aufl.
- Elliott, M. R. und Valliant, R. (2017). Inference for Nonprobability Samples. *Statistical Science* 32, S. 249–264. [Doi: https://doi.org/10.1214/16-STS598](https://doi.org/10.1214/16-STS598)
- Eurostat (2021). *European Statistical System. Handbook for quality and metadata reports, 2021 re-edition*, Luxemburg. [Doi: https://doi.org/10.2785/666412](https://doi.org/10.2785/666412)
- Groves, R. M. und Lyberg, L. (2010). Total Survey Error: Past, Present, and Future. *Public Opinion Quarterly* 74, S. 849–879. [Doi: https://doi.org/10.1093/poq/nfq065](https://doi.org/10.1093/poq/nfq065)
- Groves, R. M., Fowler, F. J., Couper, M. P., Lepkowski, J. M., Singer, E. und Tourangeau, R. (2009). *Survey Methodology*. Hoboken (NJ): John Wiley & Sons.
- Hidiroglou, M. A. und Lavallée, P. (2009). Sampling and Estimation in Business Surveys, in: *Sample Surveys: Theory, Methods and Inference* hrsg. v. Pfeffermann, D. und Rao, C. R., Band 29A: *Handbook of Statistics*, Amsterdam: Elsevier, Kap. 17, S. 441–470. [Doi: https://doi.org/10.1016/S0169-7161\(08\)00017-5](https://doi.org/10.1016/S0169-7161(08)00017-5)

- Hostettler, S. und Kraft, E. (2021). FMH-Ärztstatistik 2020 – die Schweiz im Ländervergleich, Schweizerische Ärztezeitung 102, S. 417–422.
Doi: <https://doi.org/10.4414/saez.2021.19698>
- Kruskal, W. H. und Mosteller, F. (1979a). Representative Sampling, I: Non-Scientific Literature. International Statistical Review 47, S. 13–24.
Doi: <https://doi.org/10.2307/1403202>
- Kruskal, W. H. und Mosteller, F. (1979b). Representative Sampling, II: Scientific Literature, Excluding Statistics. International Statistical Review 47, S. 111–127.
Doi: <https://doi.org/10.2307/1402564>
- Kruskal, W. H. und Mosteller, F. (1979c). Representative Sampling, III: The Current Statistical Literature. International Statistical Review 47, S. 245–265.
Doi: <https://doi.org/10.2307/1402647>
- Loo, Mark P., Van Der, J. und De Jonge, E. (2021). Data Validation Infrastructure for R. Journal of Statistical Software 97, S. 1–31. Doi: <https://doi.org/10.18637/jss.v097.i10>
- Maronna, R. A., Martin, R. D., Yohai, V. J. und Salibián-Barrera, M. (2019). Robust Statistics: Theory and Methods (with R), Hoboken (NJ): John Wiley & Sons, 2. Aufl.
Doi: <https://doi.org/10.1002/9781119214656>.
- Matej, A. (2018). On Some Reweighting Schemes for Nonignorable Unit Nonresponse, Survey Statistician 77, S. 21–33.
- R Development Core Team (2023). R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria.
URL: <http://www.R-project.org>
- Särndal, C.-E. und Lundström, S. (2005). Estimation in Surveys with Nonresponse. Hoboken (NJ): John Wiley & Sons.
- Särndal, C.-E., Swensson, B. und Wretman, J. (1992). Model Assisted Survey Sampling, New York: Springer.
- Schnell, R., Hill, P. B. und Esser, E. (2018). Methoden der empirischen Sozialforschung. Berlin: De Gruyter Oldenbourg, 11., überarb. Aufl.
- Schoch, T. (2021). wbacon: Weighted BACON algorithms for multivariate outlier nomination (detection) and robust linear regression, Journal of Open Source Software 6, 3238.
Doi: <https://doi.org/10.21105/joss.03238>
- Valliant, R. (2020). Comparing Alternatives for Estimation from Nonprobability Samples. Journal of Survey Statistics and Methodology 8, S. 231–263.
Doi: <https://doi.org/10.1093/jssam/smz003>
- Van Zwet, W. R. (1979). Mean, Median, Mode, II. Statistica Neerlandica 33, S. 1–5.
Doi: <https://doi.org/10.1111/j.1467-9574.1979.tb00657.x>
- Weidmann, J.-H. (2004). Unterschiedliche Taxpunktwerte in Arztpraxen und Spitälern: Warum? Schweizerische Ärztezeitung 85, S. 2680.
- Willis, G. B. (2020). Questionnaire Design, Development, Evaluation, and Testing: Where Are We, and Where Are We Headed?, in: Advances in Questionnaire Design, Development, Evaluation, and Testing, hrsg. v. Beatty, P. C., Collins, D., Kaye, L., Padilla, J.-L., Willis, G. G. und Wilmot, A. Hoboken (NJ): John Wiley & Sons, Kap. 1, S. 3–24.

Anhang A: TARDOC

Tabelle 17: Sparten

Modell	Sparten	Anmerkungen
Y	Sprechzimmer Allgemein	
Y	UBR 1 Allgemein	
Y	UBR 1 Allgemein nichtärztliche Tätigkeit	
Y	UBR 2 Allgemein	
Y	UBR 3 Allgemein	
Y	UBR 4 Allgemein	
Y	Röntgen Arztpraxis	
Y	Praxis-Ultraschall	
Y	Labor	Eliminationssparte
Y	Apotheke/ SD	Eliminationssparte
Y	Allgemeinfläche	Indirekt-Sparte
X	Sprechzimmer Psychiatrie	
X	UBR nicht ärztliche Psychiatrie	
X	Apotheke/ SD	

Anmerkung: UBR = Untersuchungs- und Behandlungsraum.

Anhang B: Rollende Kostenstudie RoKo

B.1 Datenmodell und Übersicht

Die Datenaufbereitung zur RoKo erfolgt in mehreren Etappen/ Arbeitsschritten. Zur besseren Verständlichkeit ist die Abfolge der Arbeitsschritte, wie sie durch die Ärztekasse durchgeführt wird, in Abbildung 6 illustriert. Nachfolgend werden die einzelnen Arbeitsschritte erläutert.

Abbildung 6: Prozess der Datenaufbereitung



Quelle: Eigene Darstellung.

B.1.1 Übersicht

Der RoKo-Datensatz, auf dem die Replikation beruht, wurde uns von der Ärztekasse als Microsoft Accessdatenbank zur Verfügung gestellt und umfasst die Daten zur RoKo der Jahre 2017–2019. Die Accessdatenbank umfasst die in Tabelle 18 aufgeführten Tabellen.

Tabelle 18: Auflistung der Tabellen

Tabellen	Tabellen (Forts.)
DATEN_1	AZTAB
DATEN_2	HTNTTAB
DATEN_4	Spez_Modell
DATEN_PAPIER	

Anmerkung: Tabelle DATEN_3 existiert nicht.

Das Erhebungssubjekt der RoKo sind einzelne Ärztinnen und Ärzte. Wir sprechen im Folgenden von den Subjekten. Diese können mithilfe der pseudonymisierten Identifikationsnummer (SEQ) in den Tabellen des RoKo-Datensatzes identifiziert werden. Es ist nicht möglich, die RoKo-Daten mithilfe der SEQ-Nummer (auf der Stufe des Subjekts) mit anderen Datensätzen (z. B. MedReg) zu verknüpfen.

B.1.2 Beschreibung der Tabellen

Die wesentlichen Informationen zum RoKo-Fragebogen sind in den drei folgenden Tabellen enthalten:

- DATEN_1: Die Tabelle DATEN_1 (33'762 Zeilen/ Subjekte) enthält Informationen des RoKo-Fragebogendeckblatts; siehe Tabelle 19.
- DATEN_2: Die Tabelle DATEN_2 (22'944 Zeilen/ Subjekte) enthält Informationen zu den Fragebogenblöcken: Betriebsaufwand und Bruttoertrag, sowie den ersten Teil des Fragebogenblocks zu den Leistungsdimension; siehe Tabelle 21.
- DATEN_4: Die Tabelle DATEN_4 (14'755 Zeilen/ Subjekte) enthält Daten des zweiten Teils des Fragebogenblocks zu den Leistungsdimensionen; siehe Tabelle 20.

Tabelle 19: Merkmale in Tabelle DATEN_1

Variablenname	Bedeutung
SEQ	Identifizier (Arzt/ Ärztin)
Jahr	Erhebungsjahr
canton	Kanton
SEQ_REF	
SD	Selbstdispensation
GruppenP	Gruppenpraxis
AnzP	Wie viele Personen praktizieren ausser Ihnen auf eigene Rechnung?
Institut	Führen / leiten Sie eine Institution / AG (juristische Person)
AnzInstitut	Falls ja, wie viele ärztliche Personen ausser Ihnen sind dort angestellt?
noOfInhabitants	Kategorielle Grösse der Gemeinde nach Personen
Angaben_Finanzdaten	Aus Finanzbuchhaltung oder selbst eingetragen

Quelle: RoKo-Datenbank, eigene Darstellung.

Tabelle 20: Merkmale in Tabelle DATEN_4

Variablenname	Beschrieb
SEQ	Identifizier (Arzt/ Ärztin)
Jahr	Erhebungsjahr
LP3	TOTAL PRAXISGRÖSSE
LP31-LP42	Unterkategoriefragen zur Praxisgrösse

Quelle: RoKo-Datenbank, eigene Darstellung.

Tabelle 21: Merkmale in Tabelle DATEN_2

Variablenname	Bedeutung
SEQ	Identifizier (Arzt/ Ärztin)
Jahr	Erhebungsjahr
canton	Kanton
SEQ_REF	Identifizier (Arzt/ Ärztin)
SD	Selbstdispensation
GruppenP	Gruppenpraxis
AnzP	Wie viele Personen praktizieren ausser Ihnen auf eigene Rechnung?
Institut	Führen / leiten Sie eine Institution / AG (juristische Person)
AnzInstitut	Falls ja, wie viele ärztliche Personen ausser Ihnen sind dort angestellt?
noOfInhabitants	Kategoriale Grösse der Gemeinde nach Personen
Angaben_Finanzdaten	Aus Finanzbuchhaltung oder selbst eingetragen

Quelle: RoKo-Datenbank, eigene Darstellung.

Die folgenden vier Tabellen enthalten Zusatzinformationen zu den Ärztinnen und Ärzten, die den Fragebogen ausgefüllt haben.

DATEN_PAPIER: Die Tabelle DATEN_PAPIER (5'152 Zeilen/ Subjekte) enthält eine Liste der Fragebogen (SEQ und Jahr), die in den Jahren 2017–2019 von Hand ausgefüllt wurden; siehe Tabelle 22.

AZTAB: Die Tabelle AZTAB (45'472 Zeilen/ Subjekte) enthält Informationen über den Fachbereich aller Identifizier (SEQ-Nummern). Dabei werden registerartig alle jemals in der RoKo vorgekommenen Identifizier mit ihrem Fachbereich (als Code), ihrem Geburtsjahr und dem Eröffnungsjahr ihrer Praxis gelistet; siehe Tabelle 23

HTNTTAB: Die Tabelle HTNTTAB (177 Zeilen/ Fachbereiche) listet alle ärztlichen Fachbereiche (inklusive ihres Codes) auf Deutsch, Französisch und Italienisch auf; siehe Tabelle 24.

Spez_Modell: Die Tabelle Spez_Modell (185 Zeilen/ Fachbereiche) klassifiziert die Codes der Fachbereiche in verschiedene Kostenmodelle (INFR, ModX und ModY). Ambulant tätige Ärztinnen und Ärzte werden dem Kostenmodell Y (ModY) zugerechnet; Psychiaterinnen und Psychiater (und Psychotherapeut/innen) sind Teil des Modells X (ModX).

Tabelle 22: Merkmale in Tabelle DATEN_PAPIER

Variablenname	Beschrieb
SEQ	Identifizier (Arzt/ Ärztin)
Jahr	Erhebungsjahr

Quelle: RoKo-Datenbank; eigene Darstellung.

Tabelle 23: Merkmale in Tabelle AZTAB

Variablenname	Beschrieb
SEQ	Identifizier (Arzt/ Ärztin)
HTNT	Code für den Fachbereich
HTNT_BACKUP	Code für den Fachbereich alt
GebJ	Geburtsjahr des Identifiziers (Arzt/ Ärztin)
Eroeff	Eröffnungsjahr der Praxis

Quelle: RoKo-Datenbank; eigene Darstellung.

Tabelle 24: Merkmale in Tabelle HTNTTAB

Variablenname	Beschrieb
SHTNT	Code für den Fachbereich, identisch zu HTNT
STATTEXT	Beschrieb des Fachbereichs
Dtext/Ittext/Ftext	Beschreibung des Fachbereichs in DE/FR/IT
TypTyp	
FMH_ID	
StratGroup	

Quelle: RoKo-Datenbank; eigene Darstellung.

Tabelle 25: Merkmale in Tabelle Spez_Modell

Variablenname	Beschrieb
Spez	Code für den Fachbereich, identisch zu HTNT
Modell	Kostenmodelle

Quelle: RoKo-Datenbank; eigene Darstellung.

B.2 Datenaufbereitung und Datenbereinigung

Im ersten Schritt gilt es, die einzelnen Tabellen zu einem Datensatz zusammenzusetzen.

Die drei Tabellen zu «Grundinformation Fragebogen» (DATEN_1, DATEN_2 und DATEN_4) können mit Hilfe des Schlüssels, der sich aus der SEQ-Nummer und dem Erhebungsjahr zusammensetzt, zu einer Tabelle zusammengefasst werden. Dabei wird der Datensatz DATEN_2 (siehe Tabelle 21) als Ausgangsbasis verwendet,²⁶ die Informationen aus Tabelle 19 und Tabelle 20 werden dann hinzugefügt. Über den gleichen Schlüssel kann auch die erste Tabelle der «Zusatzinformation» (DATEN_Papier, siehe Tabelle 22) hinzugefügt werden.

Mithilfe des Schlüssels «Code für den Fachbereich» (HTNT, SHTNT, Spez) können die übrigen «Zusatzinformation» (d. h. die Information aus Tabelle 23, Tabelle 24 und Tabelle 25) zu einer gemeinsamen Tabelle kombiniert werden. Im nächsten Schritt können die Tabellen zu «Grundinformation Fragebogen» und «Zusatzinformation» fusioniert werden. Die resultierende Tabelle wird als «Mastersheet» bezeichnet und umfasst 22'944 Zeilen/ Subjekte.

Das zusammengesetzte «Mastersheet» bildet die Basis für die weiteren Datenbearbeitungen. Zuerst werden alle Artefakte aus dem «Mastersheet» entfernt. Als Artefakte gelten alle Einträge von Subjekten, welche den folgenden Bedingungen genügen:

- Die totalen Aufwände (Frageitem AA) sind negativ oder null.
- Die subtotalen Aufwände (Frageitems A1T – A7T) sind negativ.
- Alle Einträge, die einen (oder mehrere) fehlenden Eintrag bei einem der folgenden Frageitems haben: AA, A1T – A7T oder den totalen Erträgen (EE).

Durch die Anwendung der Bedingungen reduziert sich die Anzahl der Zeilen/ Subjekte im «Mastersheet» um 377 Zeilen/ Subjekte, so dass das ein Total von 22'567 Zeilen/ Subjekte resultiert.

B.2.1 Entfernung Tarif-fremder Leistungen: INFRA

Leistungen, deren Kosten nach dem INFRA-Kostenmodell bestimmt sind (z. B. Leistungen zu MRI oder CT von Radiologen), werden aus dem Datensatz entfernt. Es handelt sich insgesamt um 291 Zeilen/ Subjekte mit Angaben zum INFRA-Modell. Der Datensatz umfasst nach dem Ausschluss noch Angaben zu 22'276 Subjekten.

B.2.2 Herausrechnen A216

Mit dem Frageitem A216 wird der Personalaufwand von delegierten Psychotherapeuten erhoben. Im TARMED war es vorgesehen, dass auch Leistungen der delegierten Psychotherapie über den TARMED abgerechnet werden. Am 1. Juli 2022 trat das Anordnungsmodell in

²⁶ Durch das Hinzufügen der Deckblattdaten aus der Tabelle DATEN_1 zu den Fragebogendaten DATEN_2 gehen 10'818 Zeilen/ Subjekte verloren, die als angefangene Deckblätter zu werten sind, die aber nicht weitergeführt wurden.

der psychologischen Psychotherapie in Kraft. Das Anordnungsmodell gilt auch für den TAR-DOC. Psychologische Psychotherapeutinnen und Psychotherapeuten in freier Praxis arbeiten neu auf Anordnung einer Ärztin oder eines Arztes als selbständige Leistungserbringer. Als Folge davon gilt es die Personalkosten für die delegierte Psychotherapie (A216) aus dem Total des Personalaufwands herauszurechnen. Dies geschieht dadurch, dass der Personalaufwand von delegierten Psychotherapeuten und -therapeutinnen (Frageitem A216) gleichermassen vom Total des Personalaufwands (Frageitem A2T) und den totalen Aufwänden (AA) subtrahiert wird. In gleicher Weise wird auch das Praxisergebnis (Q) angepasst.

B.3 Korrektur/ Herausrechnen von A215

Der Personalaufwand kann von den Befragten im Fragebogen entweder als Totalwert erfasst werden (siehe Item A2T) oder durch die Angabe von neun einzelnen Teilpositionen/Items erfolgen (die dann die Summe A2T bilden). Das Frageitem A215 ist eines der neun Teilpositionen. Es beinhaltet den Personalaufwand von angestellten Ärzten bzw. Ärztinnen, Praxisvertretungen und ärztlichen Assistenten/innen. Es wurde von den Befragten nur unvollständig ausgefüllt.

Die Beträge im Frageitem A215 betreffen die ärztlichen Leistungen und sind darum nicht Gegenstand der technischen Leistungen gemäss KOREG. Folgerichtig müssen die Beiträge in A215 aus der übergeordneten Position A2T herausgerechnet oder korrigiert werden. Dabei muss nun berücksichtigt werden, dass der Wert von A215 für alle diejenigen Befragten unbekannt ist, die nur den Totalwert in A2T, jedoch nicht Frageitem A215 angegeben haben. Bei diesen Fällen soll ein hypothetischer Wert (counterfactual value) mithilfe eines statistischen Modells geschätzt und dann für A215 imputiert werden. Nach der erfolgten Imputation kann A2T um den imputierten Wert bei A215 korrigiert werden.

Durch das Herausrechnen von A215 aus A2T verändert sich A2T, und das bringt es mit sich (im Sinne einer Kettenreaktion), dass auch die zu A2T übergeordnete Position AA und Q angepasst werden müssen.

Die Details der Imputation werden im nächsten Abschnitt beschrieben.

B3.1 Imputation durch die Ärztekasse

Für die Schätzung des statistischen Modells und die anschliessende Imputation wird ein statistisches Modell herangezogen. Die Schätzung erfolgt nur für eine Teilmenge der Ärztinnen und Ärzte, und zwar für diejenigen, welche folgenden Bedingungen genügen:

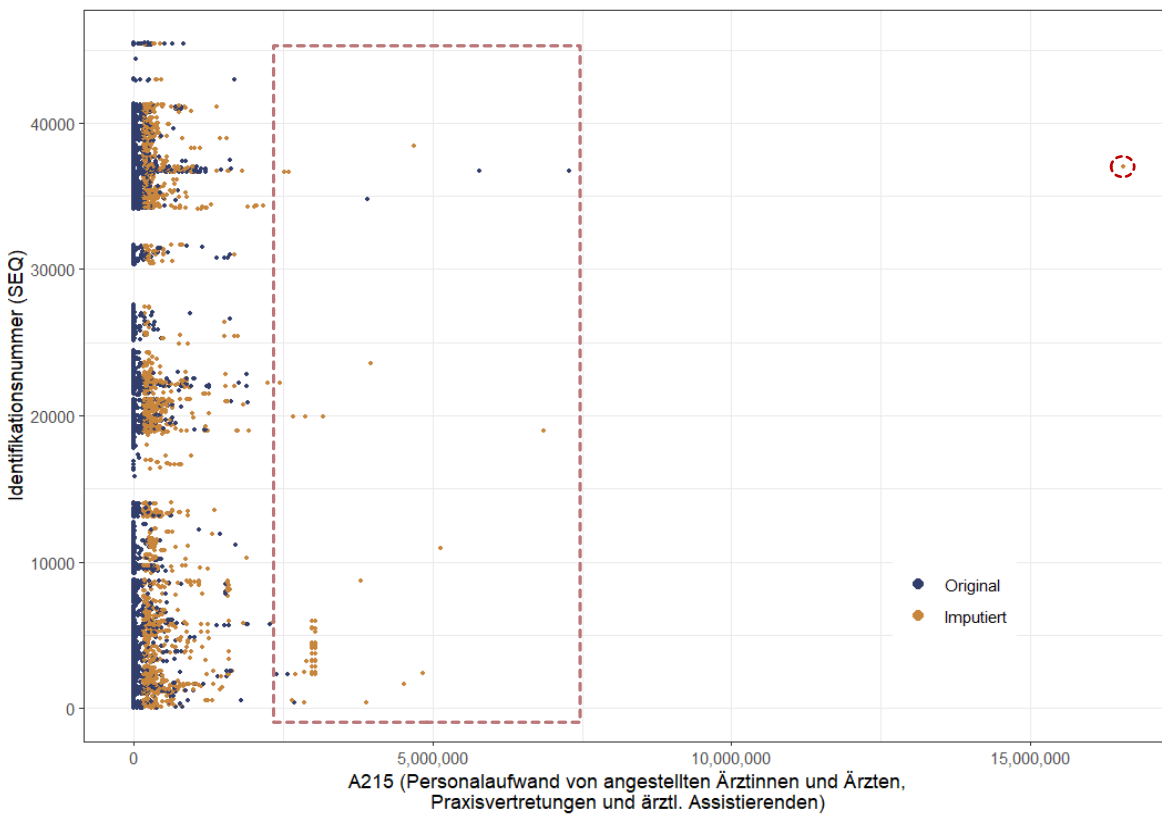
- Der gesamte Personalaufwand (Frageitem A2T) ist grösser als 300'000 CHF
- und der Personalaufwand von angestellten Ärzten/innen und ärztl. Assistenten/innen (Frageitem A215) liegt über 300'000 CHF.

Diesen Bedingungen genügen lediglich eine Teilmenge, T_{A215} , im Umfang von 335 Ärztinnen und Ärzte (in der Referenzperiode), was einem Anteil von knapp 1.5% entspricht (bei einem

Total von 22'567 Subjekten). Dieser kleine Anteilswert ist eine Folge davon, dass das Minimalkriterium (>300'000 CHF) beim Personalaufwand (Frageitem A215) eingefordert wird.

Anmerkung. Das Minimalkriterium in der Höhe von 300'000 CHF wurde von der Ärztekasse gewählt, weil es (nach Abzug der üblichen Personalkosten) ungefähr dem Aufwand eines Ärzteäquivalent entspricht. Die Ärztekasse begründet die Festlegung damit, dass bei Item A215 nicht alle Leistungen aus selbständiger Tätigkeit enthalten sind (es könnte sein, dass das Item auch die Aufwände für Stellvertreter und Assistenzärzte ohne parallele Verrechnung umfasst). Ohne dieses Minimalkriterium ist der Quotient niedriger, der Eliminationsbeitrag kleiner und der Aufwand entsprechend grösser.

Abbildung 7: Verteilung der Personalaufwände von angestellten Ärztinnen und Ärzten, Praxisvertretungen und ärztl. Assistierenden (Variable A215): Originalwerte vs. imputierte Werte



Quelle: RoKo-Datenbank; eigene Darstellung.

Für die Teilmenge T_{A215} der $n=335$ Ärztinnen und Ärzte, die den obenstehenden Bedingungen genügen, wird in einem ersten Schritt das mittlere Verhältnis $A215_i/A2T_i$ über alle Angaben geschätzt. Die gemittelte Verhältniszahl besitzt einen Wert 0.566 (gerundet). Die Werte für die Variable A215 werden dann imputiert, wenn der Wert bei der Variable A2T grösser als 300'000 CHF ist. Diese Bedingung leitet sich daraus ab, dass im RoKo-Fragebogen nur das Ausfüllen von A2T obligatorisch ist (jedoch nicht die Unterpositionen). Bei einem Anteil von

8.8% aller Einträge sind beide Bedingungen erfüllt. Für diese Einträge erfolgt nun die Imputation nach der Formel $\text{Einsetzung} = 0.566 \times \text{A2T}$. Nach erfolgter Einsetzung wird der imputierte Wert bei der Variable A215, wie oben erwähnt, von A2T und auch von AA abgezogen. Das Praxisergebnis (Q) muss nach Abzug des zusätzlichen Aufwands um den gleichen Beitrag erhöht werden.

In Abbildung 7 ist die Verteilung der Personalaufwände von angestellten Ärztinnen und Ärzten, Praxisvertretungen und ärztl. Assistenten/innen (Variable A215) vor und nach der Imputation dargestellt (Originalwerte vs. imputierte Werte). Es fällt auf, dass mitunter hohe Werte (Ausreisser) imputiert wurden. Die hohen imputierten Werte sind jedoch mehrheitlich darauf zurückzuführen, dass bereits die für die Imputation genutzte Variable A2T hohe Werte annimmt. Weil jedoch die finale Schätzung durch das gestutzte Mittel (eng. trimmed mean) erfolgt – siehe unten –, ist der Einfluss einzelner Ausreisser beschränkt. Wir wollen dennoch im nächsten Abschnitt eine Sensitivitätsanalyse durchführen und den Einsatz robuster Methoden prüfen.

B3.2 Sensitivitätsanalyse zur Imputation

Die Imputation der Ärztekasse beruht, was die Schätzung angeht (wie oben beschrieben) auf der Teilmenge T_{A215} , die durch das Minimal Kriterium $>300'000$ CHF beim Personalaufwand bestimmt ist. In diesem Abschnitt wollen wir einerseits robuste Schätzmethoden einsetzen und andererseits untersuchen, wie sich die Resultate verändern, wenn für das Minimal Kriterium ein anderer Betrag als die 300'000 CHF verwendet wird.

Die Einsetzungsmethode auf Grundlage des geschätzten Verhältnisses zwischen den Variablen $A215_i$ und $A2T_i$, wie sie von der Ärztekasse verwendet wurde, kann als lineares Regressionsproblem formuliert werden. Diese Formulierung eröffnet Wege für den Einsatz alternativer Schätzmethoden. Für die Beobachtungen der Teilmenge T_{A215} wird das folgende (heteroskedastische) einfache lineare Regressionsmodell durch den Ursprung formuliert $A215_i = \beta \cdot A2T_i + \epsilon_i$ ($i \in T_{A215}$), wobei ϵ_i ein unabhängiger Zufallsfehler mit Erwartungswert 0 und Varianz $\sigma^2 \cdot A2T_i^2$ ist; $\sigma^2 > 0$ bezeichnet die unbekannte Varianz. Der Kleinstquadrateschätzer von β entspricht – unter diesem Modell – dem arithmetischen Mittel von $A215_i/A2T_i$; also dem Schätzer, den die Ärztekasse für die Imputation verwendet hatte. Anstelle des Kleinstquadrateschätzers soll nun auch ein MM-Regressionsschätzer von β (siehe Maronna et al., 2019, Kapitel 4 und 5) verwendet werden. Diese Klasse von Schätzern ist robust gegenüber Ausreissern in den Daten.

Die Resultate zur robusten Schätzung und der Wahl des Minimal Kriteriums sind in Tabelle 26 aufgeführt. Das Minimal Kriterium wird im Wertebereich 200'000 – 400'000 CHF variiert. Für jede Ausprägung des Minimal Kriteriums wurden Schätzung und Imputation separat durchgeführt. Für die imputierten Werte wurden anschliessend das arithmetische Mittel der Variablen AA, A2T und Q berechnet und mit dem jeweiligen Mittel verglichen, das aus der Referenzimputation der Ärztekasse resultiert (d. h. Mindestkriterium = 300'000 CHF und nicht robuste Schätzung, siehe beige hinterlegte Zellen in Tabelle 26). In der Tabelle sind die relativen Unterschiede (in Prozent) ausgewiesen. Die beige hinterlegten Zellen zeigen die

relativen Unterschiede der Referenzimputation zu sich selbst; sie sind (selbstverständlich) null. Betrachten wir nun den Fall, in welchem das Mindestkriterium dem Wert der Referenzimputation entspricht (300'000 CHF), aber die robuste Schätzmethode zur Anwendung kam. Die relativen Abweichungen des geschätzten Mittelwerts nehmen in diesem Fall für die Variablen A2T, AA und Q folgende Werte an: 0.27%, 0.08% und 0.26%. Es gibt also nur minimale Abweichungen. Als nächstes wollen wir untersuchen, welchen Einfluss die Wahl des Mindestkriteriums besitzt. Wählt man dieses Kriterium als $300'000 \pm 50'000$ CHF, so resultieren relative Differenzen im Vergleich zur Referenzimputation in der Höhe von etwa $\pm 2\%$. Wenn das Mindestkriterium über den gesamten Wertebereich 200'000 – 400'000 CHF variiert wird, resultieren relative Abweichungen in der Höhe von maximal $\pm 5.2\%$ im Vergleich zur Referenzimputation. Wir halten fest, dass die Wahl des Mindestkriteriums (qualitativ) einen Einfluss besitzt, dieser jedoch quantitativ unbedeutend ist. Die Referenzimputation der Ärztekasse bedarf darum keiner Anpassung.

Tabelle 26: Sensitivitätsanalyse: Relative Differenz des Mittelwerts für die Variablen AA, A2T und Q bei

Mindest-Kriterium	(Standard) Schätzung			Robuste Schätzung		
	A2T	AA	Q	A2T	AA	Q
200'000	-5.19%	-1.46%	-5.02%	-4.93%	-1.38%	-4.76%
250'000	-2.14%	-0.60%	-2.06%	-1.88%	-0.53%	-1.81%
300'000	0.00%	0.00%	0.00%	0.27%	0.08%	0.26%
350'000	2.02%	0.57%	1.95%	2.30%	0.64%	2.22%
400'000	4.00%	1.12%	3.86%	4.26%	1.19%	4.11%

Quelle: RoKo-Datenbank; eigene Darstellung.

B.4 Analyse der fehlenden Werte

Für die Frageitems zu den Aufwänden und Erträgen und die darauf basierenden Aggregate (z. B. der Gesamtaufwand AA als Summe aller Teilaufwände) wurden die Häufigkeiten der fehlenden und der gültigen Werte untersucht (Tabelle 27).

Einzelne Frageitems weisen eine grosse Anzahl von fehlenden Werten auf. Dies ist allerdings nicht ungewöhnlich, sondern Ausdruck davon, dass diese Items nur von wenigen Ärztinnen und Ärzten ausgefüllt wurden. Die Aggregate A1T–A7T sind Totalwerte, die sich aus einzelnen Items zusammensetzen. Sie weisen keine fehlenden Werte auf. Das gilt auch für die gesamthaften Aufwände und Erträge.

Tabelle 27: Anteil fehlender Werte und Anzahl gültige (d. h., nicht fehlende) Werte nach Modell

Frageitem	Modell X			Modell Y		
	Anteil fehlende W.	Anz. gültige W.	Mittelwert	Anteil fehlende W.	Anz. gültige W.	Mittelwert
Abgeleitete Variablen						
Erträge	0%	7'501	238'296	0%	29'263	687'505
Aufwände	0%	7'501	113'272	0%	29'263	448'345
Praxisergebnis	0%	7'501	125'024	0%	29'263	239'160
Aufwände						
A1T	0%	7'501	7'770	0%	29'263	140'179
A11	63%	2'806	9'247	52%	13'959	128'788
A122	83%	1'286	85	68%	9'468	11'418
A123	83%	1'273	4	71%	8'516	1'182
A124	79%	1'579	1'683	66%	9'820	15'346
A12	63%	2'749	1'753	53%	13'683	25'881
A2T	0%	7'501	20'549	0%	29'263	127'792
A21	100%	0	NA	100%	0	NA
A22	75%	1'884	5'812	71%	8'466	39'501
A3T	0%	7'501	20'754	0%	29'263	36'912
A31	45%	4'110	19'031	52%	14'061	32'859
A32	49%	3'790	2'017	54%	13'409	4'738
A4T	0%	7'501	1'331	0%	29'263	3'353
A5T	0%	7'501	2'890	0%	29'263	18'379
A6T	0%	7'501	34'438	0%	29'263	61'069
A62	41%	4'407	15'065	49%	14'782	25'798
A63	45%	4'157	21'111	51%	14'350	40'610
A64	45%	4'089	2'112	53%	13'896	2'807
A7T	0%	7'501	25'533	0%	29'263	60'653
A71	39%	4'574	18'560	46%	15'937	33'223
A72	52%	3'616	1'355	56%	12'955	17'823
A73	46%	4'074	1'771	49%	14'943	7'448
A74	44%	4'233	3'687	47%	15'574	6'378
A75	39%	4'584	1'783	44%	16'244	4'264
Erträge						
E1T	0%	7'501	230'414	0%	29'263	650'087
E2T	-	-	-	-	-	-
E3T	-	-	-	-	-	-

Quelle: RoKo-Datenbank; eigene Darstellung; Abkürzung: NA (not available): Wert nicht verfügbar.

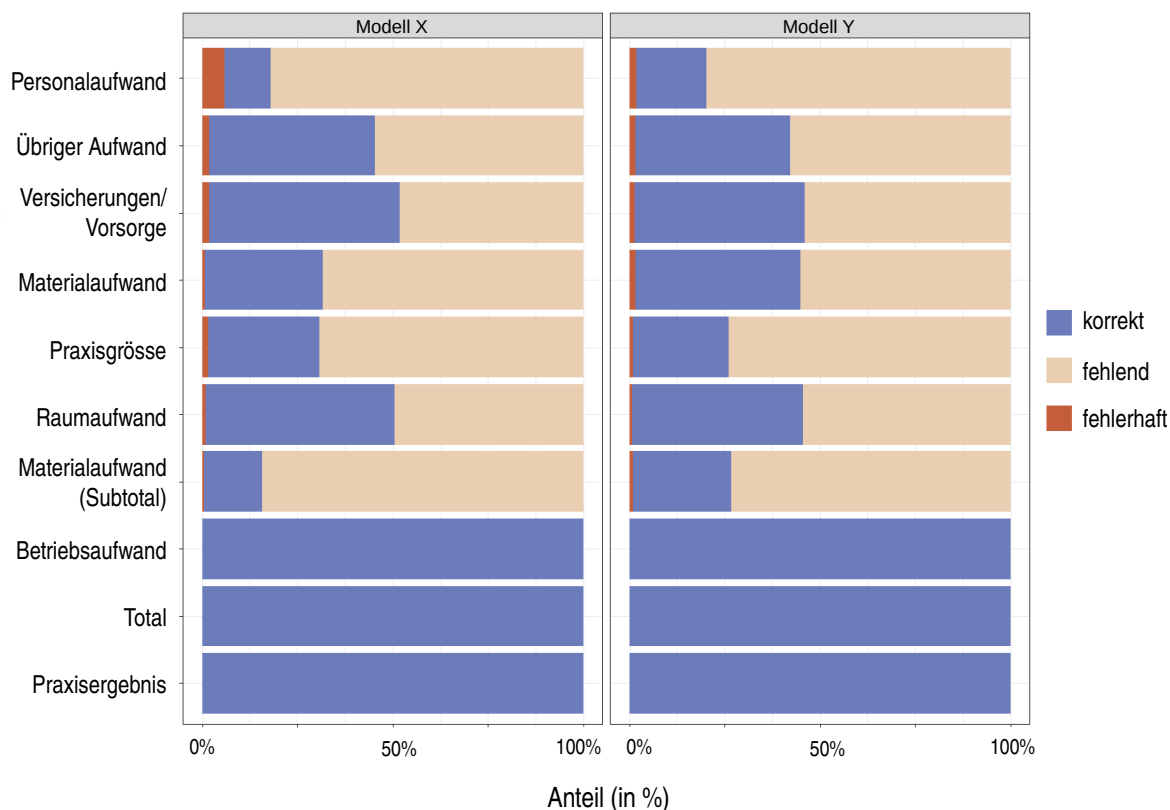
In nächsten Schritt soll die Konsistenz der gebildeten Aggregate oder abgeleiteten Variablen geprüft werden (z. B. Position Materialaufwand setzt sich aus zwei Unterpositionen zusammen). Hierzu wurden die in Tabelle 28 aufgeführten Regeln entwickelt (siehe bspw. auch de Waal, 2009). Im nächsten Schritt werden die Regeln auf die Daten angewendet und geprüft, ob sie eingehalten sind. Dazu wird das Framework von Loo et al. (2021) verwendet. Die Ergebnisse zur Prüfung der Regeln sind in Abbildung 8 visualisiert.

Aus Abbildung 8 erkennen wir, dass nur ganz wenige Fälle auftreten, bei welchen die Regeln verletzt sind. Am meisten Regelverstöße treten bei der Regel Personalaufwand_total auf. Die dominierende Position bei dieser Regel ist (modellübergreifend) die Variable A216 (Personalaufwand Therapeuten). Eine genauere Untersuchung dieser Regelverstöße ist nicht zielführend, weil diese Position seit Inkrafttreten des Anordnungsmodells in der psychologi-

schen Psychotherapie per 1. Juli 2022 nicht mehr relevant ist. Wir haben bereits oben erläutert, dass die Variable A216 für die vorliegende Untersuchung herausgerechnet wurde.

Insgesamt ist die Anzahl der Regelverstösse klein. Angesichts der vergleichsweise kleinen Anzahl an Verstössen, ist auch das Ausmass der daraus resultierenden möglichen Verzerrungen als gering einzuschätzen.

Abbildung 8: Validierung der Regeln nach Modell



Quelle: RoKo-Datenbank (ohne Expandierung); eigene Berechnung.

Tabelle 28: Fragebogenregeln

Kategorie	Name der Regel	Regel
Gesamt	Praxis Ergebnis	$Q = EE - AA$
Materialaufwand	Materialaufwand_subtotal	$A122 + A123 + A124 = A12$
Materialaufwand	Materialaufwand_total	$A11 + A12 = A1T$
Personalaufwand	Personalaufwand_total	$A211 + A212 + A213 + A214 + A215 + A216 + A218 + A219 + A22 = A2T$
Raumaufwand	Raumaufwand_total	$A31 + A32 = A3T$
Versicherungen/Vorsorge	Versicherungen/ Vorsorge_total	$A62 + A63 + A64 = A6T$
Übriger Aufwand	Übriger_Aufwand_total	$A71 + A72 + A73 + A74 + A75 = A7T$
Betriebsaufwand	Betriebsaufwand	$A1T + A2T + A3T + A4T + A5T + A6T + A7T = AA$
Praxisgrösse	Praxisgrösse_total	$LP31 + LP32 + LP33 + LP34 + LP35 + LP36 + LP37 + LP38 + LP39 + LP40 + LP41 + L42 = LP3$
Gesamt	Total_komplett	$is_complete(A1T, A2T, A3T, A4T, A5T, A6T, A7T)$

Eigene Darstellung.

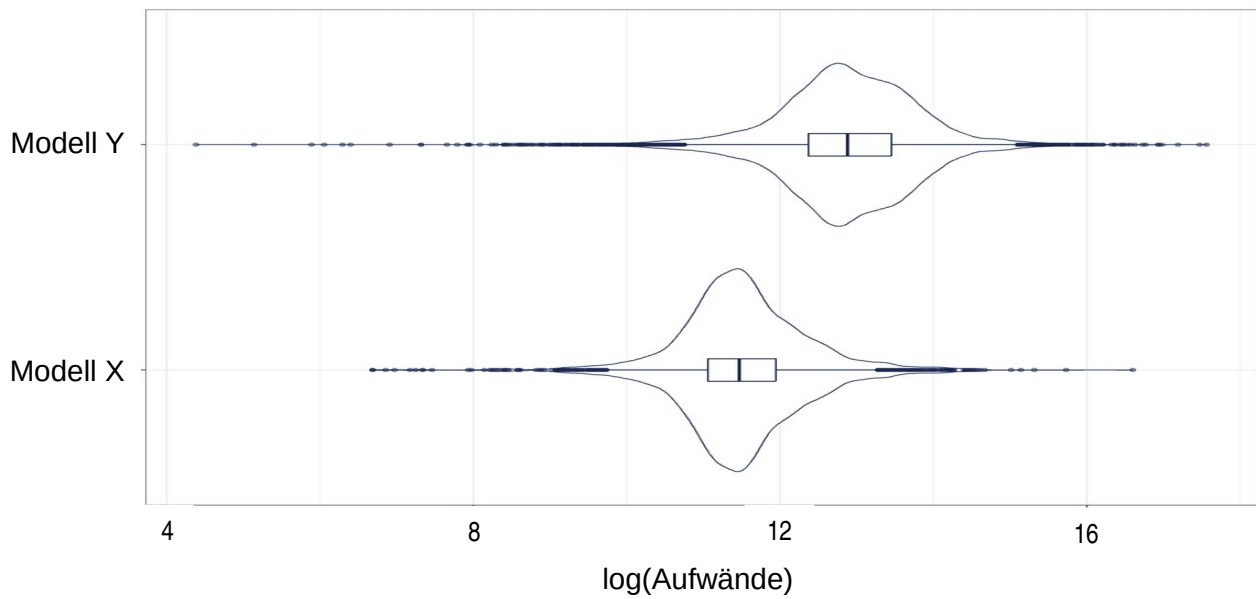
B.5 Ausreisserentdeckung und Ausschluss

B5.1 Untersuchungen zu den univariaten Verteilungen

In diesem Abschnitt werden die Verteilungen der Gesamterträge und -aufwände (Variablen EE und AA) untersucht. Es handelt sich hierbei um Aggregate/ Totalwerte, die sich aus den Teilaufwänden bzw. Teilerträgen herleiten.

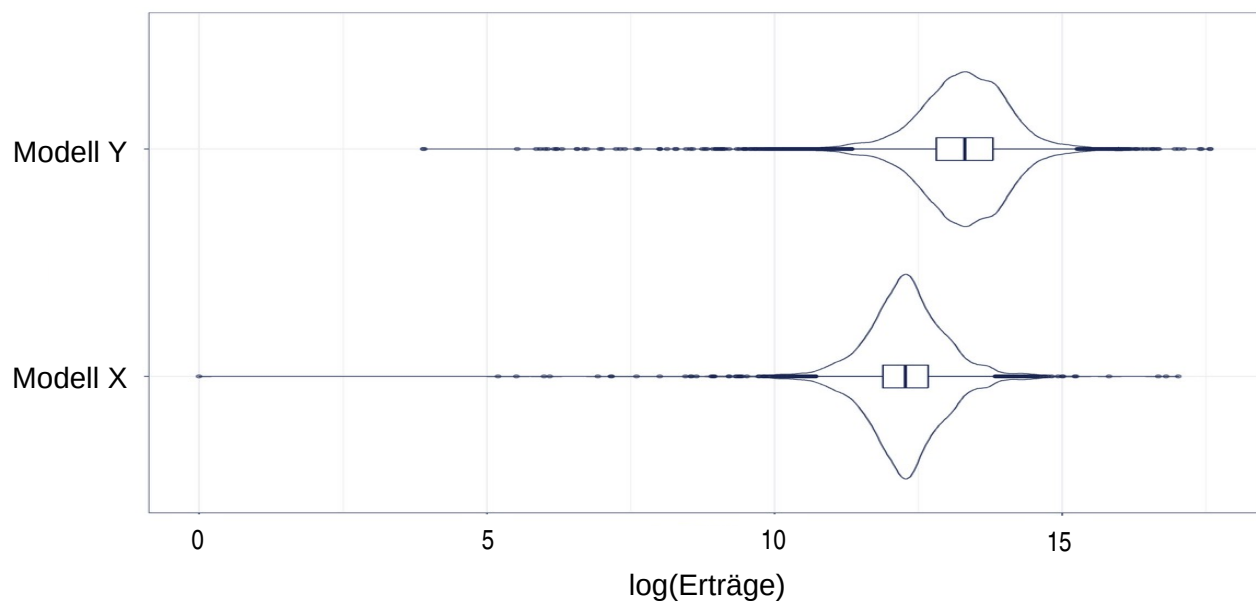
Die geschätzten Dichtefunktionen der gesamthaften Aufwände und Erträge (unter einer Logarithmustransformation) sind für die Modelle X und Y in Abbildung 9 bzw. Abbildung 10 dargestellt. Die Verteilungen der transformierten Daten sind nahezu symmetrisch. Es wird darauf verzichtet, die Dichteschätzungen auch für die nicht transformierten Variablen zu dokumentieren, weil deren Schiefe (wie erwartet) gross ist. Aus den Dichtefunktionen in Abbildung 9 geht klar hervor, dass die Verteilungen – mit Ausnahme einiger Ausreisser – nicht durch statistische Artefakte beeinträchtigt sind. Zur gleichen Erkenntnis gelangt man auch bei der Verteilungsanalyse zu den gesamthaften Erträgen; siehe Abbildung 10.

Abbildung 9: Dichte/ Verteilung + Boxplot der Gesamtaufwände (nach Modell)



Quelle: RoKo-Datenbank (ohne Expandierung); eigene Berechnung.

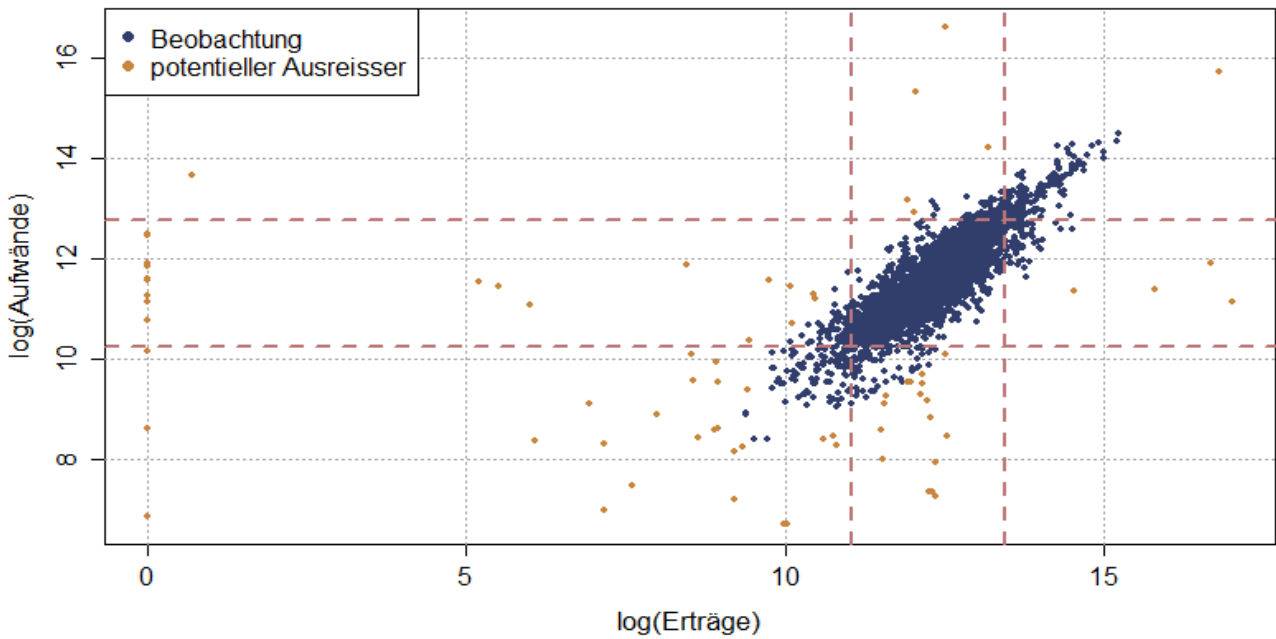
Abbildung 10: Dichte/ Verteilung + Boxplot der Gesamterträge (nach Modell)



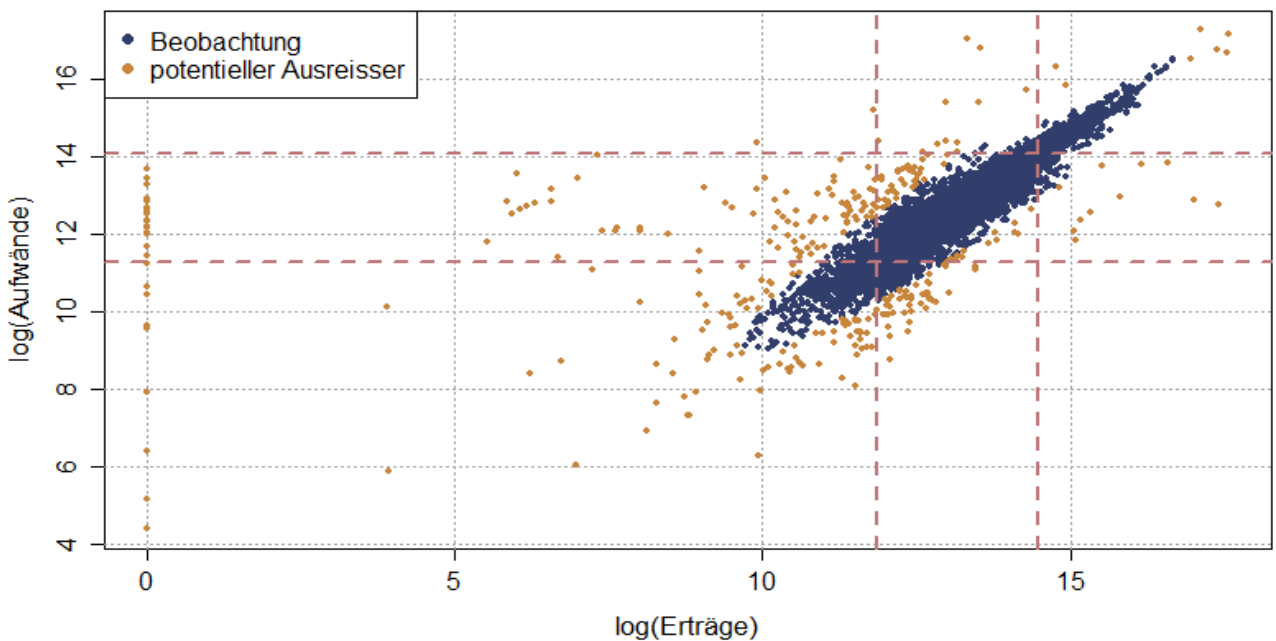
Quelle: RoKo-Datenbank (ohne Expandierung); eigene Berechnung.

Abbildung 11: Streudiagramm der (logarithmierten) Gesamterträge vs. Gesamtaufwände für die Modelle X und Y

a) Modell X



b) Modell Y



Quelle: RoKo-Datenbank (ohne Expandierung); eigene Berechnung.

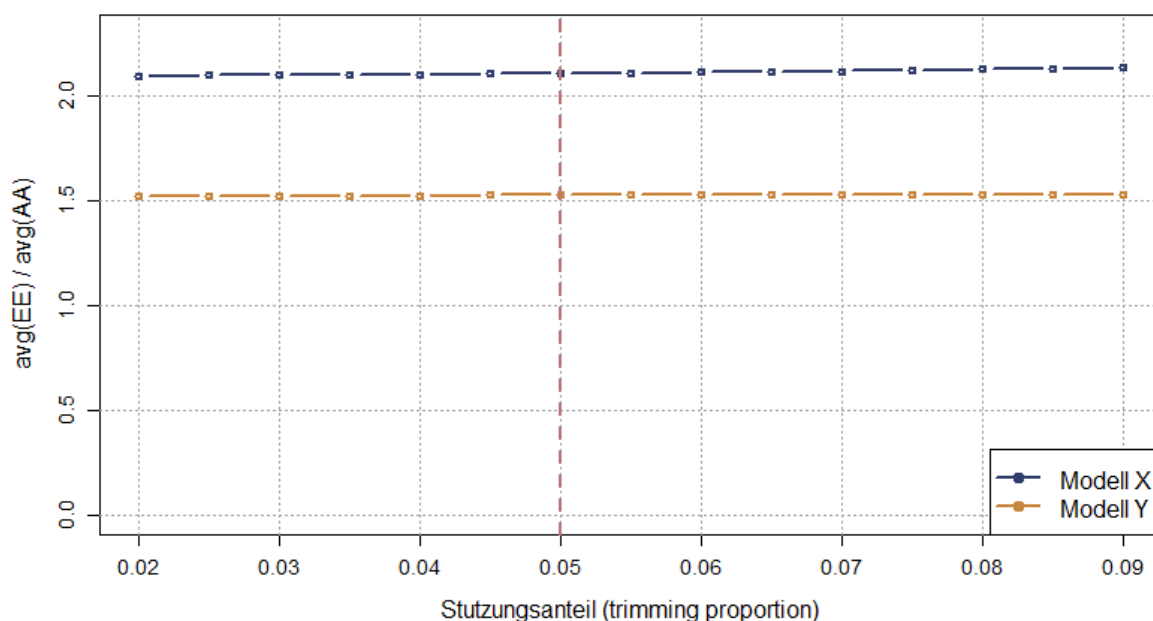
Anmerkung: Die Deklaration potenzieller Ausreisser erfolgte mit dem BACON-Algorithmus von Billor et al. (2000). Die gestrichelten Linien entsprechen den 5%- und 95%-Quantilen der Variablen log(Erträge) bzw. log(Aufwände).

B5.2 Bivariate Analyse

Die logarithmierten Gesamterträge und Gesamtaufwände sind für die Modelle X und Y in Abbildung 11 als Streudiagramme dargestellt. In den Streudiagrammen werden reguläre und potenzielle Ausreisser ausgewiesen. Die Deklaration potenzieller Ausreisser erfolgte mit dem BACON-Algorithmus von Billor et al. (2000); siehe Schoch (2021). Der BACON-Algorithmus wurde konservativ parametrisiert, sodass er eine kleine Anzahl an Beobachtungen als potenzielle Ausreisser deklariert.

Die gestrichelten Linien in Abbildung 11 entsprechen den 5%- und 95%-Quantilen der Variablen $\log(\text{Erträge})$ bzw. $\log(\text{Aufwände})$. Die Ärztekasse verwendet in ihren Berechnungen (siehe unten) das symmetrische 5%-gestutzte Mittel (engl. trimmed mean), um Ausreisser an beiden Enden der Verteilung zu entfernen. Durch die Stutzung im Umfang von je 5% reduziert sich der Umfang der verbleibenden Beobachtungen um jeweils 10% der Beobachtungen. Der BACON-Algorithmus deklariert (in der gewählten Parametrisierung) einen wesentlich kleineren Anteil an potenziellen Ausreissern, nämlich: 1.6% beim Modell X und 1.9% beim Modell Y.

Abbildung 12: Sensitivitätsanalyse: Effekt der Wahl des Stutzungsparameters auf das Verhältnis der gestutzten Mittel von AA zu EE



Quelle: RoKo-Datenbank (ohne Expandierung); eigene Berechnung.

Anmerkung: avg bezeichnet das gestutzte Mittel (average). Die vertikale, gestrichelte Linie bei 0.05 zeigt die (symmetrische) Stutzung bei 5% an, die von der Ärztekasse verwendet wird (Referenz).

Bei unimodalen, rechtsschiefen Verteilungen (die hinreichend regulär sind) – wie sie bspw. bei den Variablen Gesamtaufwände und -erträge (AA und EE) der Modelle X und Y auftreten,

führt eine stärkere Stutzung zu einem kleineren Wert für das gestutzte Mittel.²⁷ Zum Beispiel: Das (symmetrische) 6%-gestutzte Mittel ist kleiner als das 5%-gestutzte Mittel. Die Ärztekasse setzt eine (identische) separate Stutzung für mehrere Variablen an; darauf gehen wir noch detaillierter ein. Durch die parallele Stutzung (im gleichen Umfang) bei bspw. den Gesamtaufwänden (AA) und den Gesamterträgen (EE), wird das Verhältnis dieser beiden (für die Analyse zentralen) Größen nicht verändert. Wir wollen hierfür einen empirischen Nachweis geben. Abbildung 12 zeigt das Verhältnis der gestutzten Mittel für EE und AA. Der Stutzungsparameter wird dabei im Wertebereich 2% – 9% variiert. Die vertikale, gestrichelte Linie bei 0.05 zeigt die (symmetrische) Stutzung bei 5% an, die von der Ärztekasse verwendet wird (Referenz). Wir sehen aus Abbildung 12, dass das Verhältnis der beiden gestutzten Mittel durch die Höhe des Stutzungsparameters (z. B. 5% vs. 7%) nur geringfügig beeinflusst wird.

Die Wahl des Stutzungsparameters (für eine symmetrische Stutzung) in der Höhe von 5% ist arbiträr (wie es jede Festlegung ist). Wichtiger ist, dass die Wahl des Stutzungsparameters (innerhalb des genannten Wertebereichs) keinen nennenswerten Einfluss auf das geschätzte mittlere Verhältnisse der Variablen EE und AA hat. Dies gilt jedoch nur dann, wenn beide Variablen im gleichen Ausmass gestutzt werden.²⁸

B5.3 Umsetzung der symmetrischen 5%-Stutzung

Um ein robusteres Mittel pro Modell und Jahr zu erhalten, werden die Daten unterhalb der 5%-Perzentile und oberhalb der 95%-Perzentile gestutzt. Dieses Vorgehen wird von der Ärztekasse separat für folgenden Variablen angewendet: Gesamtaufwand (AA), den totalen Bruttoertrag aus Praxistätigkeiten (E1T) und das Praxisergebnis (Q). In Tabelle 29 sind die Fallzahlen (pro Jahr) vor und nach dem Ausschluss nach Perzentilen aufgeführt. Die Spalte «Expandiert» weist die Anzahl der Fälle aus, die nach der Expansion/ Vervielfachung der Daten resultiert. Wir gehen im nächsten Abschnitt detaillierter darauf ein.

²⁷ Sei F eine absolut stetige, unimodale Verteilungsfunktion auf den nicht negativen reellen Zahlen. Die Wahrscheinlichkeitsdichte von F wird mit f bezeichnet. Wir wollen überdies annehmen, dass F ein endliches erstes Moment besitzt und rechtsschief ist. Eine rechtsschiefe Verteilung kann bspw. durch die Modus-Median-Mittelwert-Ungleichung (MMM) $\text{Modus} \leq \text{Median} \leq \text{Mittelwert}$ charakterisiert werden. Eine hinreichende Bedingung für die MMM-Ungleichung ist gegeben durch $f\{F^{-1}(q)\} \geq f\{F^{-1}(1 - q)\}$ für alle $q \in (0, 0.5)$; siehe van Zwet (1979). Unter den getroffenen Annahmen folgt nun (Dhar und Chaudhuri, 2012, Proposition 1), dass das symmetrisch α -gestutzte Mittel mit $\alpha \in (0, 0.5)$ für eine feste Verteilungsfunktion F eine abnehmende Funktion von α ist. Mit anderen Worten, je stärker gestutzt wird, desto kleiner ist das gestutzte Mittel.

²⁸ Unsere Analyse beschränkte sich auf eine Untersuchung der gestutzten Mittelwerte, weil die mittleren Aufwände und Erträge für die Tarifgestaltung zentral ist. Die Stutzung kann für andere Charakteristika sehr wohl zu anderen Resultaten führen.

Tabelle 29: Fallzahlen nach Referenzjahr und Modell (vor und nach dem Schneiden nach Quantilen)

Arbeitsschritt	Fallzahlen			Expandiert
	2017	2018	2019	
<i>Modell X</i>				
Ausgangssituation	1'542	1'528	1'494	-
Entfernt durch Schneiden nach Quantilen	271	261	262	-
Realisierte Fallzahl	1'271	1'267	1'232	7'501
<i>Modell Y</i>				
Ausgangssituation	5'842	5'951	5'897	-
Entfernt durch Schneiden nach Quantilen	1'030	1'032	1'026	-
Realisierte Fallzahl	4'812	4'919	4'871	29'263

Quelle: RoKo-Datenbank; eigene Darstellung.

B.6 Gewichtung – Expansion/ Vervielfachung

Die Datensätze zu den drei Referenzjahren 2017-2019 werden nach dem folgenden Schema

$$3 \times \text{Daten 2019} + 2 \times \text{Daten 2018} + 1 \times \text{Daten 2017}$$

«vervielfacht» (expandiert). Der resultierende Datensatz umfasst also 3 Kopien der 2019er-Daten und 2 Kopien der 2018er-Daten. Die Daten für das Jahr 2017 sind nur einmal im resultierenden Datensatz enthalten. Mit dem 3-2-1-Gewichtungsschema will die Ärztekasse den Angaben neueren Datums eine grössere Bedeutung zumessen. Als Folge der Expansion resultieren für die Modelle X und Y eine Anzahl von 7'501 bzw. 29'263 Einträgen; siehe letzte Spalte in Tabelle 29.

Anmerkung. Anstelle der Expansion/ Vervielfachung nach dem 3-2-1-Schema hätte man auch die Jahresdatensätze kombinieren (ohne Vervielfachung) und mit einem Gewicht versehen können, das dem 3-2-1-Schema entspricht. Als Schätzer wäre dann das gewichtete Mittel (engl. weighted mean) zu verwenden.

B.7 Kanton VS: Kalibrierung

In diesem Abschnitt wird die Kalibrierung des RoKo-Teildatensatzes für den Kanton VS (Referenzjahr 2020) beschrieben. Die Kalibrierung ist ein Verfahren, um mögliche Verzerrungen infolge von Coverage und Nonresponse Error zu verringern (allenfalls zu beheben); siehe bspw. Särndal und Lundström (2005, Kap. 6).

Das Mitgliederverzeichnis der kantonalen Ärztegesellschaft umfasst die folgenden fünf Variablen (die Ausprägungen der Variablen sind in Klammer angegeben): Altersgruppen (31-40, 41-50, 51-60, 61-70, über 70), Spezialisierung (Grund: Grundversorger/innen, Spez: Spezialisten/innen), Geschlecht (M: Männer; F: Frauen), Urbanität (Stadt oder Land) und Sprache (D: Deutsch, F: Französisch). Für die Kalibrierung wird die Gruppe der Über-70-Jährigen ausgeschlossen.

Es wäre nicht ausreichend, nur eine Kalibrierung für die fünf einzelnen Variablen durchzuführen. Darum wurden mehrere log-lineare Modelle (siehe bspw. Agresti, 2002, Kapitel 8 und 9) für das Mitgliederverzeichnis geschätzt. Das erklärungsstärkste Modell umfasst sowohl Einzelvariablen als auch Interaktionsterme (z. B. Sprache : Geschlecht steht für den Interaktionsterm der Variablen Sprache und Geschlecht). Diese Interaktionsterme können als zweidimensionale Kreuztabellen aufgefasst werden.

Die Kalibrierung der Ausgangsgewichte w_i (wobei anfänglich $w_i = 1$ für alle $i = 1, \dots, n$ Beobachtungen der Erhebung gilt) erfolgt für die folgenden Merkmalkonfigurationen/ Interaktionsterme:

Spezialisierung : Altersgruppe

Spezialisierung : Urbanität

Sprache : Geschlecht

Geschlecht : Altersgruppe

Urbanität : Sprache

Es handelt sich insgesamt um 28 lineare Kalibrierungsbedingungen.

Anmerkung. Mit der gewählten Anzahl von 28 Kalibrierungsbedingungen bewegen wir uns am oberen Rand der Modellgrösse. Der geschätzte Design Effect des gewichteten Mittels (vor und nach der Kalibrierung) liegt bei 15-20%. Weil der kalibrierte Datensatz zum Kanton VS ins Ensemble aller RoKo-Kantone eingebettet ist, sind Effizienzüberlegungen nicht so zentral, als wenn man kantonsspezifische Auswertungen machen würde. Wenn das Ziel kantonsspezifische Auswertungen sind, dann sollte (nach Möglichkeit) die Anzahl der Kalibrierungsbedingungen reduziert werden. Dadurch werden Effizienzüberlegungen beim trade-off zwischen (potenziellem) Bias und Varianz stärker berücksichtigt.

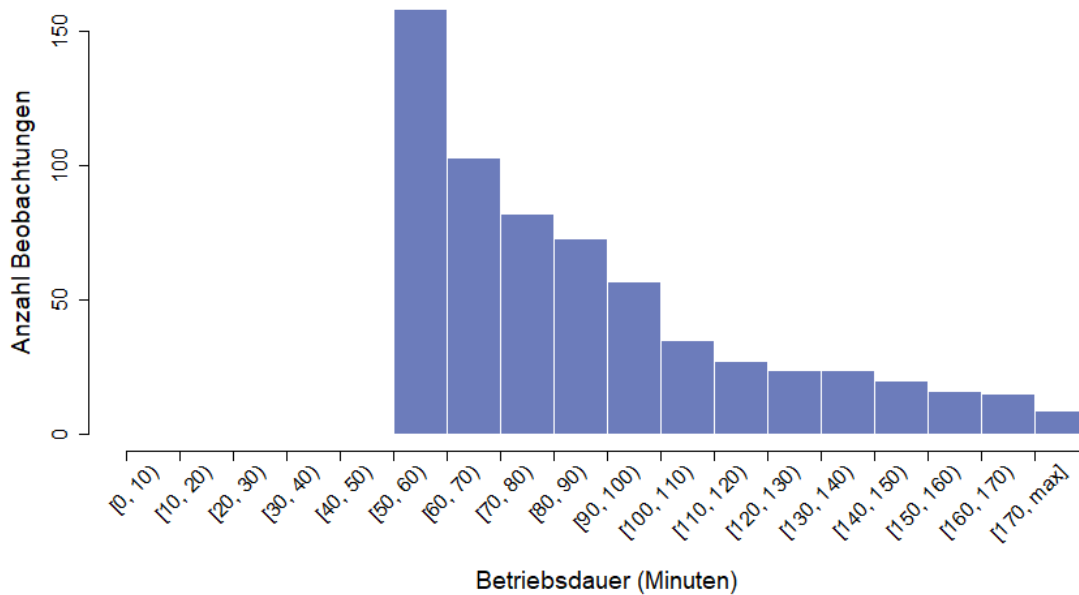
Die Kalibrierung beruht auf den absoluten Häufigkeiten/ Totalwerten zu den Variablen/ Interaktionstermen. Als Distanzfunktion wird die Funktion für das Raking verwendet (siehe Deville und Särndal, 1992, Fall 2 in Tabelle 1).

Das kleinste bzw. grösste kalibrierte Gewicht, g_i , beträgt 1.92 bzw. 9.6. Das Verhältnis zwischen dem grössten und dem kleinsten kalibrierten Gewicht ist 4.9. Es wurden also keine ausnehmend grossen Gewichte erzeugt. Überdies gibt es keine Gewichte, die kleiner 1 sind.

Anhang C: Spartenbetriebsdauern – NAKO

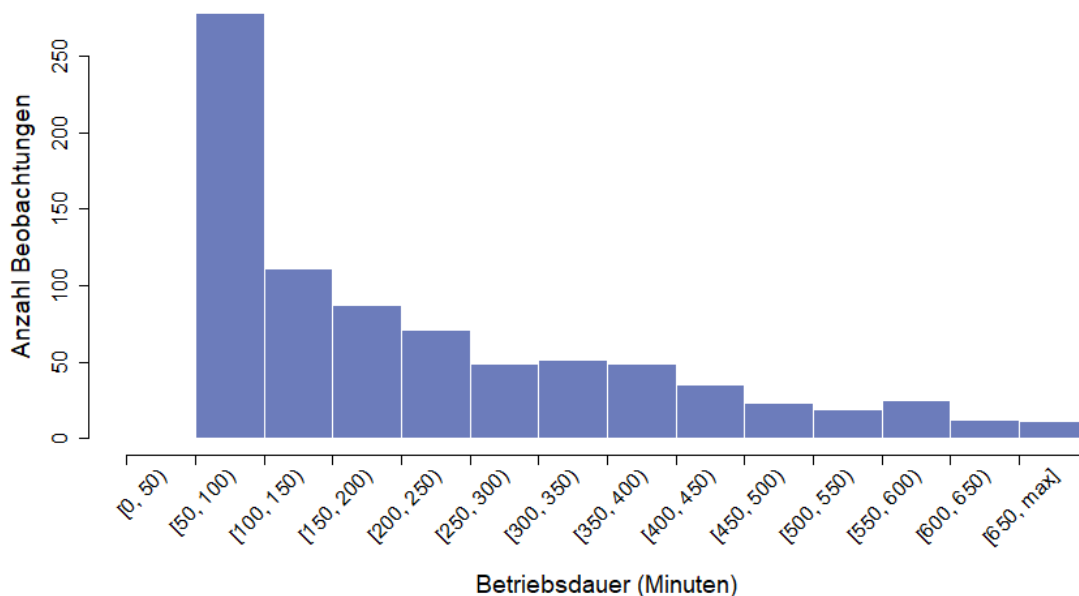
Alle nachfolgenden aufgeführten Histogramme zu den Betriebsdauern zeigen die Verteilung nach der Stützung.

Abbildung 13: Histogramm zu den Betriebsdauern für den Spartencluster Röntgen



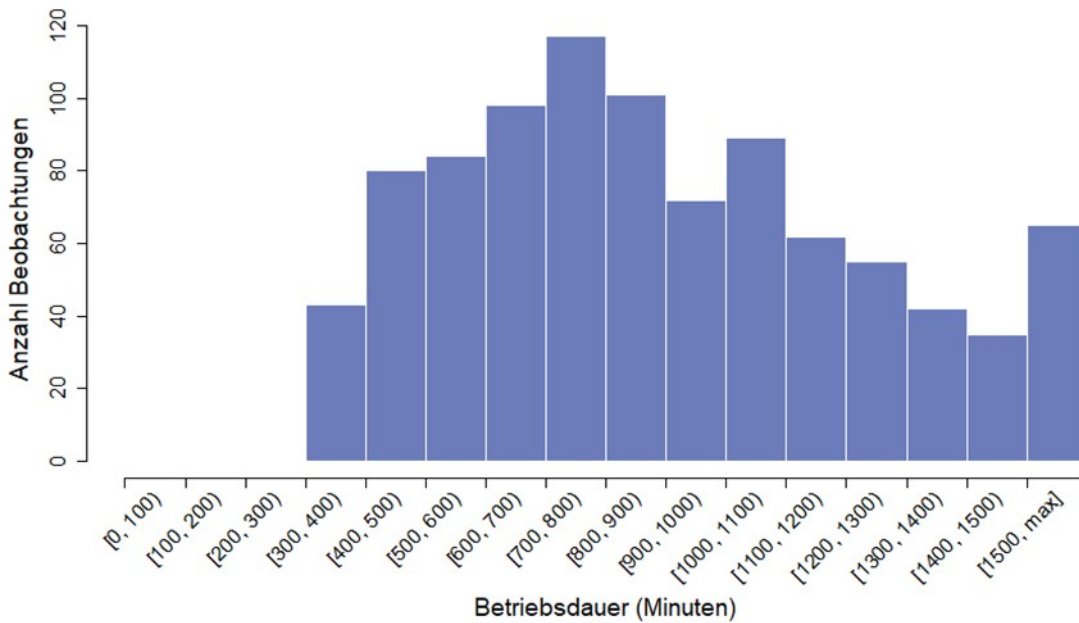
Daten: NAKO/ NewIndex; eigene Darstellung.

Abbildung 3: Histogramm zu den Betriebsdauern für den Spartencluster Ultraschall



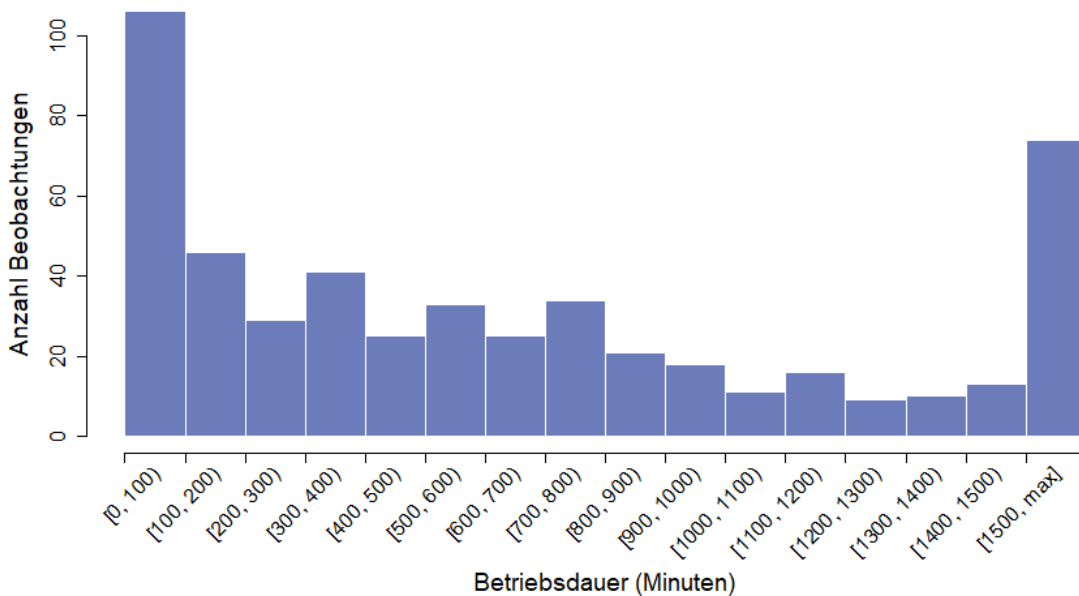
Daten: NAKO/ NewIndex; eigene Darstellung.

Abbildung 4: Histogramm zu den Betriebsdauern für den Spartencluster Sprechzimmer Psychiatrie



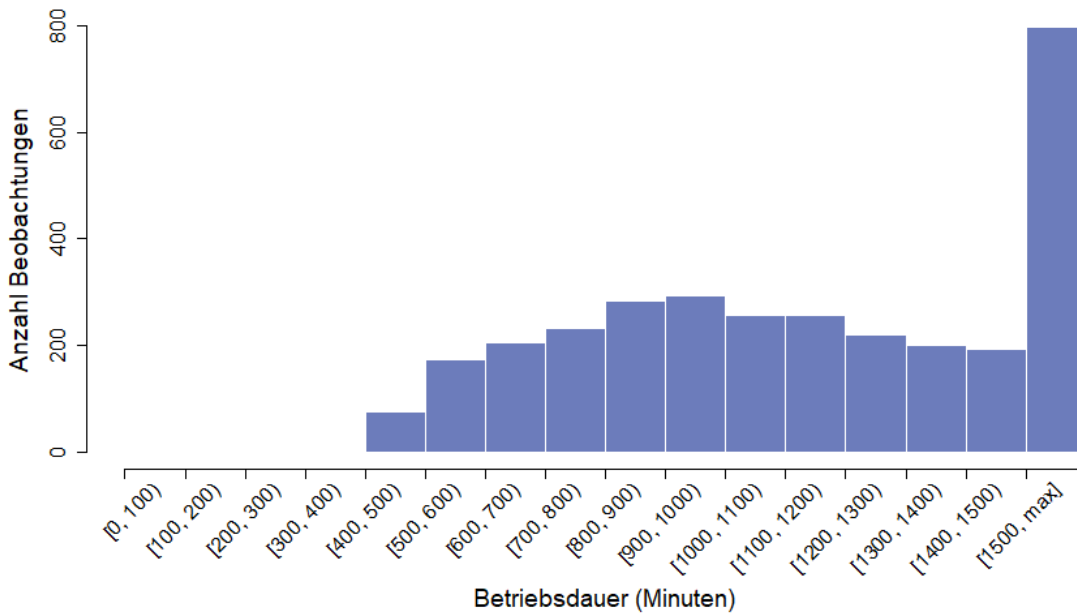
Daten: NAKO/ NewIndex; eigene Darstellung.

Abbildung 5: Histogramm zu den Betriebsdauern für den Spartencluster Sprechzimmer UBR nichtärztlich Psychiatrie



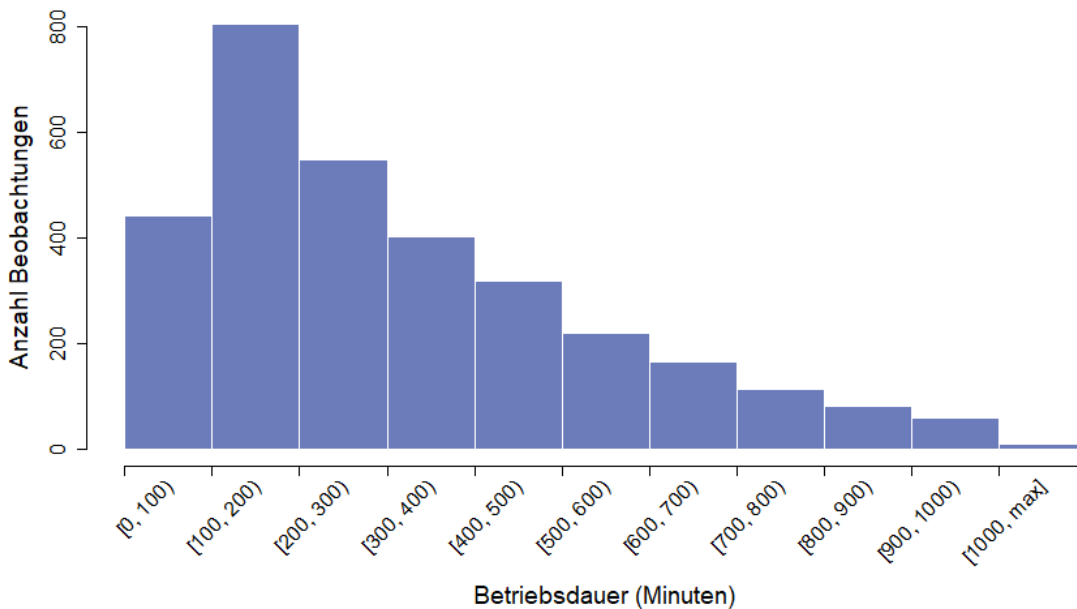
Daten: NAKO/ NewIndex; eigene Darstellung.

Abbildung 6: Histogramm zu den Betriebsdauern für den Spartencluster Sprechzimmer Allgemein



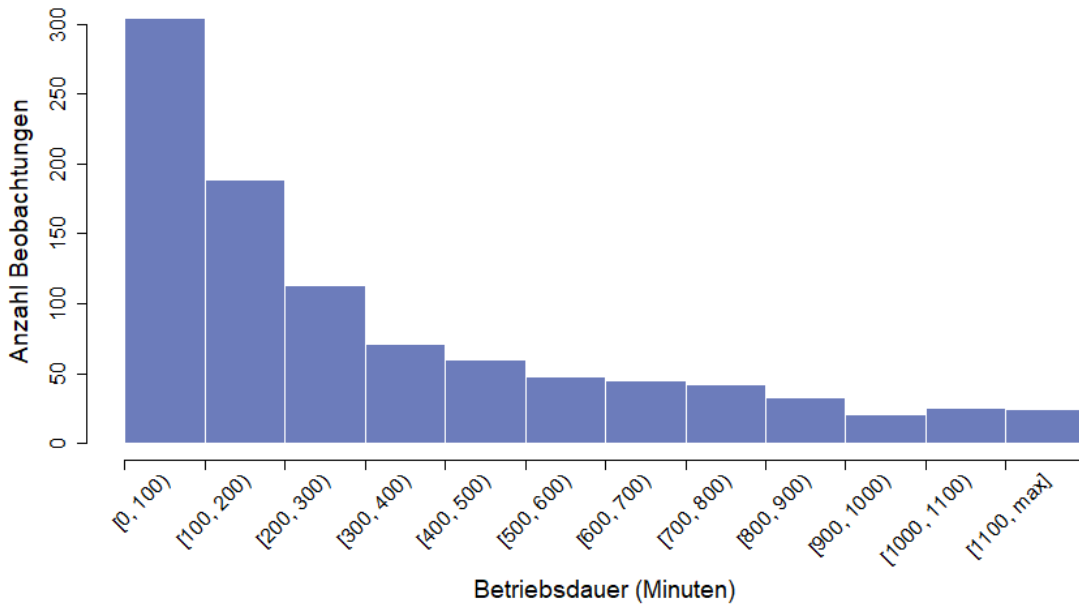
Daten: NAKO/ NewIndex; eigene Darstellung.

Abbildung 7: Histogramm zu den Betriebsdauern für den Spartencluster Sprechzimmer UBR 1 Allgemein



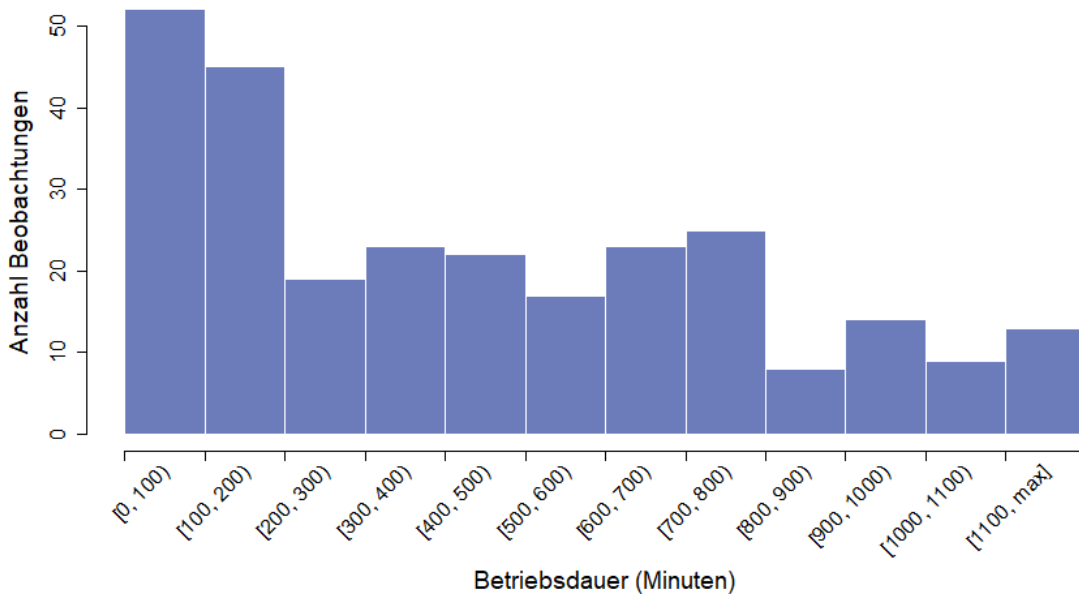
Daten: NAKO/ NewIndex; eigene Darstellung.

Abbildung 8: Histogramm zu den Betriebsdauern für den Spartencluster Sprechzimmer UBR 2 Allgemein



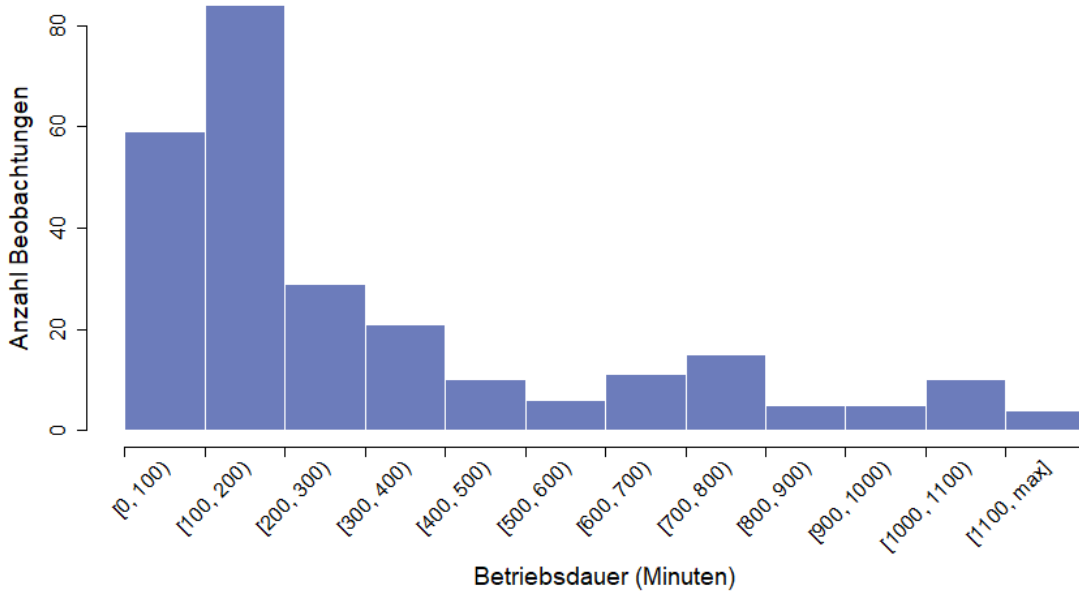
Daten: NAKO/ NewIndex; eigene Darstellung.

Abbildung 9: Histogramm zu den Betriebsdauern für den Spartencluster Sprechzimmer UBR 3 Allgemein



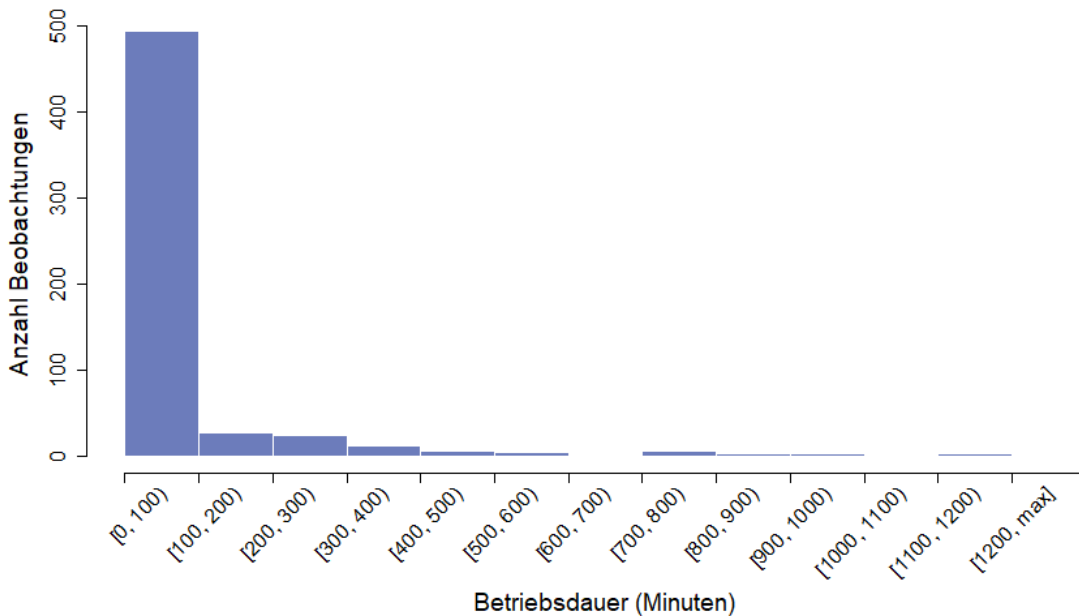
Daten: NAKO/ NewIndex; eigene Darstellung.

Abbildung 10: Histogramm zu den Betriebsdauern für den Spartencluster Sprechzimmer UBR 4 Allgemein



Daten: NAKO/ NewIndex; eigene Darstellung.

Abbildung 11: Histogramm zu den Betriebsdauern für den Spartencluster Sprechzimmer nichtärztlich Allgemein



Daten: NAKO/ NewIndex; eigene Darstellung.