

Forschungsdatenzentrum Gesundheit

Nebenbestimmungen – Anlage 3 Vorgaben für Ergebnismengen

Version 01-01-000

Forschungsdatenzentrum Gesundheit

Inhalt

1	Allgemeine Bemerkungen.....	4
1.1	Verwendung dieses Dokuments.....	4
1.2	Begriffserklärung.....	4
1.3	Prüfansatz des FDZ.....	4
1.4	Mindestfallzahl.....	4
1.5	Absenkung der Mindestfallzahl.....	5
1.6	Analyse von Stichproben.....	5
1.7	Sekundäre statistische Geheimhaltung.....	5
1.8	Export von Ergebnismengen.....	5
1.9	Kommunikation.....	6
1.10	Verantwortlichkeit.....	6
2	Prüfpunkte bei der Ergebnismengenprüfung.....	6
2.1	Skripte.....	6
2.2	Beschreibung der Ergebnismenge.....	6
2.3	Struktur der Ergebnismenge.....	8
2.4	Inhalt der Ergebnismenge.....	8
2.5	Klassifikation von Ergebnistabellen.....	9
2.5.1	Bezug.....	9
2.5.2	Statistische Klassifikation.....	9
3	Grundregeln.....	11
4	Modifikationsmöglichkeiten.....	14
5	Spezielle Regeln zur Prüfung der Ergebnisse nach Klassifikation.....	15
5.1	Häufigkeitstabellen.....	15
5.2	Kennzahltabellen/Größentabellen.....	16
5.3	Extremwerte (Minima, Maxima, Spannweite).....	17
5.4	Perzentile (inkl. Median).....	17
5.5	Modus.....	18
5.6	Lineare Aggregate.....	18
5.7	Überlebenszeitanalyse.....	19
5.8	Standardabweichung, Varianz, Schiefe, Wölbung.....	19
5.9	Korrelationskoeffizienten.....	19
5.10	Konzentrationsmaße.....	20
5.11	Modellkoeffizienten.....	20
5.12	Hypothesentests.....	21
5.13	Residuen.....	21
5.14	Faktorenanalyse.....	21
5.15	Einschluss-/Ausschlusskriterien für Studienpopulationen.....	22
5.16	Grafiken.....	22
6	Literatur.....	23

Tabellenverzeichnis

Tabelle 1: Ergebnistabelle RT_1.....	7
Tabelle 2: Ergebnistabelle RT_2.....	7
Tabelle 3: Data Description Tabelle DD_RESULT_TABLES	7
Tabelle 4: Data Description Tabelle DD_RESULT_COLS	8
Tabelle 5: Arten von Ergebnistabellen	10
Tabelle 6: Grundsätze der Ergebnismengenprüfung.....	11
Tabelle 7: Erlaubte Tabellenstrukturen	15
Tabelle 8: nicht erlaubte Tabellenstrukturen	15
Tabelle 9: nicht erlaubte Tabellenstrukturen (Fortsetzung)	16
Tabelle 10: Beispiel korrekte Formatierung einer Kennzahltable.....	16
Tabelle 11: Beispiel Ergebnistabelle für mehrere Kennzahlen	16
Tabelle 12: Beispiel inkorrekte Formatierung einer Kennzahltable.....	17
Tabelle 13: Spezialregeln der Ergebnismengenprüfung: Extremwerte.....	17
Tabelle 14: Spezialregeln der Ergebnismengenprüfung: Perzentile	17
Tabelle 15: Spezialregeln der Ergebnismengenprüfung: Modus	18
Tabelle 16: Spezialregeln der Ergebnismengenprüfung: Lineare Aggregate.....	18
Tabelle 17: Spezialregeln der Ergebnismengenprüfung: Überlebenszeitanalyse	19
Tabelle 18: Spezialregeln der Ergebnismengenprüfung: Standardabweichung, Varianz, Schiefe, Wölbung	19
Tabelle 19: Spezialregeln der Ergebnismengenprüfung: Korrelationskoeffizienten	19
Tabelle 20: Spezialregeln der Ergebnismengenprüfung: Konzentrationsmaße	20
Tabelle 21: Spezialregeln der Ergebnismengenprüfung: Modellkoeffizienten	20
Tabelle 22: Spezialregeln der Ergebnismengenprüfung: Hypothesentests	21
Tabelle 23: Spezialregeln der Ergebnismengenprüfung: Faktorenanalyse	21
Tabelle 24: Spezialregeln der Ergebnismengenprüfung: Einschluss-/Ausschlusskriterien für Studienpopulationen	22
Tabelle 25: Spezialregeln der Ergebnismengenprüfung: Grafiken	22

1 Allgemeine Bemerkungen

1.1 Verwendung dieses Dokuments

Dieses Dokument beinhaltet die Vorgaben für die Prüfung von Ergebnismengen durch das Forschungsdatenzentrum Gesundheit (FDZ). Diese Vorgaben zur Prüfung von Ergebnismengen sind als Richtlinien für das FDZ und für Forschende gedacht. Das Grundregelwerk in Kapitel 3 Grundregeln ist auf alle Ergebnismengen und die enthaltenen Ergebnistabellen anzuwenden. Sowohl Forschende als auch FDZ haben sich an diese Regeln zu halten, andernfalls kann die Ergebnismenge nicht freigegeben werden. Neben dem Grundregelwerk existieren spezielle Regeln für Statistiken in der Literatur. Diese sind in Kapitel 5 aufgeführt. Diese Regeln stellen entweder direkt eine Übertragung von Grundregeln auf diese spezielle Statistik dar oder gelten zusätzlich zu den Regeln im Grundregelwerk. Wenn für eine Statistik keine Regel aufgeführt ist, so gibt es für diese Statistik keine dem FDZ bekannte Regel in der Literatur. Wenn eine solche Statistik als Ergebnis angefordert wird, so müssen Regeln für diese entwickelt werden. Zulässige Modifikationsmöglichkeiten, welche die Forschenden im Falle einer Regelverletzung anwenden können, sind in Kapitel 4 oder bei den jeweiligen Statistiken in Kapitel 5 genannt.

1.2 Begriffserklärung

Ergebnismenge: Die Gesamtheit aller für einen Antrag zur Freigabe angeforderten Ergebnistabellen (z.B. drei Häufigkeitstabellen und eine ANOVA-Tabelle)

Ergebnistabelle: Ein zur Freigabe aus der virtuellen Analyseumgebung des FDZ-Systems angefordertes Analyseergebnis in Tabellenform, z.B. eine Häufigkeitstabelle, eine Regressionsanalyse etc.

Datenzelle: Eine Zelle, welche aus den Daten berechnete Werte enthält, z.B. Mittelwert, Fallzahl der Versicherten

Beschreibungszelle: Eine Zelle, die den Inhalt von Datenzellen oder Prü fzellen beschreibt. z. B. Geschlecht, Alter oder Diagnosecodes oder Daten aus Schlüsseltabellen

Prü fzelle: Eine Zelle, deren Inhalt ausschließlich der Prüfung der Ergebnismenge durch das FDZ dient, für die Fragestellung des Antrags jedoch irrelevant ist, z.B. Fallzahl der Versicherten, wenn diese nur zu Prü fzwecken ausgegeben wird.

1.3 Prüfansatz des FDZ

In der Literatur zu statistischer Geheimhaltung wird zwischen dem regelbasierten und dem prinzipienbasierten Prüfansatz unterschieden (z.B. Bond et al., 2015). Für das FDZ soll zunächst ein regelbasierter Ansatz verwendet werden. Das bedeutet, dass Forschende und FDZ sich an die in diesem Dokument genannten Regeln zu halten haben, es sei denn, es ist explizit vermerkt, dass Ausnahmen von einer Regel möglich sind.

1.4 Mindestfallzahl

In jedem Antrag ist eine Mindestfallzahl **MFZ** zu beantragen, welche für die Ergebnismenge verwendet werden soll. Die Fallzahl bezieht sich auf die Anzahl der einer Statistik zugrunde liegenden Versicherten, also wie viele Versicherte in die Berechnung des Ergebnisses eingeflossen sind. Der Standardwert liegt bei 30, d.h. in Ergebnissen mit Versichertenbezug tragen zu jedem Zellwert mindestens 30 Versicherte bei. Dies gilt sowohl für Häufigkeiten und Anteile als auch für Kennzahlen (z.B. Mittelwerte). Ergebnisse mit ungewichteten Fallzahlen unter **MFZ** und Statistiken, welche von Gruppen mit Fallzahlen unter **MFZ** abgeleitet wurden, sind zu modifizieren. Die Mindestfallzahl gilt für Versicherte, nicht für Leistungserbringer oder Leistungsträger. Sie ist ebenfalls nicht anzuwenden auf Behandlungsfälle, da einzelne Versicherte mehrere Behandlungsfälle auslösen können. Die Fallzahl der Versicherten muss jedoch immer mit angegeben werden, auch wenn sich die Analyse nur für die Fallzahlen im Sinne von z.B. Anzahl der behandelten Fälle interessiert.

Forschungsdatenzentrum Gesundheit

1.5 Absenkung der Mindestfallzahl

Die Mindestfallzahl kann unter bestimmten Umständen abgesenkt werden, jedoch nicht auf weniger als 5. Der Antragsteller hat dies zu beantragen. Das FDZ prüft eine Absenkung der Mindestfallzahl nach formalen und inhaltlichen Kriterien im Rahmen der inhaltlichen Prüfung. Die Beweislast liegt beim Antragsteller, das FDZ kann die Absenkung der Mindestfallzahl ablehnen.

Die Absenkung der Mindestfallzahl kann möglich sein, wenn der antragstellende Nutzungsberechtigte nachvollziehbar darlegt,

1. dass ein nach § 303 e Abs. 2 SGB V zulässiger Nutzungszweck, insbesondere die Durchführung eines Forschungsvorhabens, die Übermittlung dieser Daten erfordert (§ 303 e Abs. 3 SGB V) und
2. dass eine Reidentifikation von Versicherten trotz der Absenkung der Mindestfallzahl voraussichtlich nicht möglich ist.

Das FDZ lehnt eine Mindestfallzahl-Absenkung in der Regel ab, wenn die geplante Auswertung mehr als eines dieser Kriterien erfüllt:

- Feine regionale Granularität (feiner als PLZ-2-Steller, dabei sind sämtliche Variablen mit Regionalbezug zu berücksichtigen)
- wenn Geschlecht 3 und 4 nicht ausgeschlossen werden
- Feine Granularität des Alters (feiner als 10-Jahres-Schritte)
- Feine Granularität von Erkrankungen (feiner als ICD-3-Steller)
- Krankenkassen-spezifische Auswertung

Wenn nur eines dieser Kriterien erfüllt ist, kann eine Mindestfallzahl-Absenkung genehmigt werden, wenn 1.5.1 und 1.5.2 erfüllt sind.

1.6 Analyse von Stichproben

Sofern Forschende wünschen, lediglich auf einer Stichprobe der Daten zu arbeiten, muss dies im Antrag im Unterpunkt „Methodischer Ansatz“ angegeben werden. Die Stichprobe muss durch die Forschenden selbst in ihrem Analyseskript erzeugt werden. Die Mindestfallzahl gilt in diesem Fall für die Gesamtheit der betrachteten Population und nicht für die Stichprobe.

1.7 Sekundäre statistische Geheimhaltung

Es dürfen durch die Forschenden keine Rückrechnungen vorgenommen werden, um Originalwerte aus modifizierten Werten der Ergebnismenge zu ermitteln. Rückrechnungen dürfen weder innerhalb einer Tabelle noch zwischen Tabellen möglich sein. Differenzmengen und Restgruppen müssen ausgewiesen werden. Differenzmengen und Restgruppen, welche die Mindestfallzahl unterschreiten, sind untersagt.

Sollten Datennutzende auf Möglichkeiten der Rückrechnung von Werten kleiner der Mindestfallzahl aufmerksam werden, sind diese dem FDZ zu melden. (Siehe: § 303e Absatz 5 Sätze 2 und 3 SGB V sowie § 17 Absatz 3 Nummer 1 Verordnung zur Umsetzung der Verfahren beim Forschungsdatenzentrum Gesundheit (FDZGesV)).

1.8 Export von Ergebnismengen

Das FDZ ist darauf ausgelegt, dass der Forschungsprozess bis zur Produktion der Ergebnisse vollständig innerhalb des FDZ-Systems erfolgt. Die Ergebnismengen, welche für den Export aus dem System beantragt werden, sollen in ihrem Format und Umfang publizierbar sein. Es ist nicht vorgesehen, dass Ergebnisse aus dem System exportiert und im Anschluss weitere Analysen auf diesen Ergebnissen durchgeführt werden. Daran orientieren sich auch die Vorgaben für die Prüfung von Ergebnismengen.

Forschungsdatenzentrum Gesundheit

1.9 Kommunikation

Jegliche Kommunikation mit den Forschenden in Bezug auf die Ergebnismenge erfolgt über die Nachrichtenfunktion oder das Hochladen von Dokumenten im Antragsportal. Eine Kommunikation außerhalb des Antragsportals wird bei der Prüfung nicht berücksichtigt. Nebenabsprachen sind nicht zulässig.

1.10 Verantwortlichkeit

Grundsätzlich sind die Forschenden selbst dafür verantwortlich, dass die von ihnen produzierten und aus dem FDZ-System zu exportierenden Ergebnismengen anonym sind und keine Rückschlüsse auf Einzelpersonen, Leistungsträger oder Leistungserbringer mehr möglich sind (vgl. §303e Abs. 5 Satz 2 SGB V). Das FDZ unterstützt diese Verantwortlichkeit durch eine Prüfung der Ergebnismenge vor deren Export.

2 Prüfpunkte bei der Ergebnismengenprüfung

2.1 Skripte

Das FDZ behält sich vor, im Rahmen der Lauffähigkeits- und Ergebnismengenprüfung die von Datennutzenden erstellten Auswertungsskripte zu prüfen. Skripte können auch zu Zwecken der Qualitätssicherung geprüft werden. Datennutzende sind über die Nebenbestimmungen Anlage 2 Programmcoderegeln dazu verpflichtet, sich an diese zu halten. Die Einhaltung der Programmcoderegeln ist Voraussetzung für eine erfolgreiche Prüfung der Lauffähigkeit eines Skripts sowie für die Erstellung einer zu exportierenden Ergebnismenge.

2.2 Beschreibung der Ergebnismenge

Die Beschreibung der Ergebnismenge muss innerhalb von sog. Data Description Tabellen erfolgen. Das von den Datennutzenden eingereichte Skript muss zwei Tabellen mit dem Präfix DD_ erzeugen; eine DD_-Tabelle beschreibt die Tabellen der Ergebnismenge, die andere die jeweils darin enthaltenen Spalten. Beide müssen die unten dargestellte Form haben. Ein Beispielskript zur Erzeugung dieser Tabellen wird vom FDZ im Gitlab in der sicheren Verarbeitungsumgebung bereitgestellt. Die Spalten der Ergebnistabellen sind nach dem folgenden Schema zu klassifizieren:

Beschreibung: Diese Spalte hat rein beschreibenden Charakter und enthält keine statistischen Werte (z.B. betrachtete Altersgruppe)

Kennzahl: Optional, entfällt bei Häufigkeitstabellen. Enthält den Wert der betrachteten Kennzahl (z.B. Mittelwert).

Fallzahl: Enthält die Anzahl der Versicherten, die zu einem Ergebnis beitragen. In Häufigkeitstabellen steht in dieser Spalte die Häufigkeit.

Prüfung: Optional, enthält weitere Werte, welche ausschließlich für die Ergebnismengenprüfung durch das FDZ relevant sind.

Forschungsdatenzentrum Gesundheit

Beispiel für Ergebnismenge und dazugehörige DD_-Tabellen:

Tabelle 1: Ergebnistabelle RT_1

MW_Alter	CNT_D_PSID
40,0	1000

Tabelle 2: Ergebnistabelle RT_2

Altersgruppe	Geschlecht	CNT_D_PSID
0-10	w	50
0-10	m	66
0-10	d	5
11-20	...	...
...	...	...

Tabelle 3: Data Description Tabelle DD_RESULT_TABLES

Ergebnis-tabelle	Klassifikation_Bezug	Klassifikation_statistisch	Beschreibung	Einschluss_Filterkriterien
RT_1	Versichertenbezug	Mittelwert	Kennzahl: Mittelwert des Alters der betrachteten Versicherten	Nur Versicherte aus PLZ- Bereich 4*
RT_2	Versichertenbezug	Häufigkeitstabelle	Anzahl Versicherter mit Eigenschaft X, Stratifizierung nach Altersgruppe und Geschlecht	Nur Versicherte aus PLZ- Bereich 4*

Forschungsdatenzentrum Gesundheit

Tabelle 4: Data Description Tabelle DD_RESULT_COLS

Tabellenname	Klassifikation	Spaltenname	Beschreibung
RT_1	Kennzahl	MW_Alter	Mittelwert des Alters der betrachteten Versicherten
RT_1	Fallzahl	CNT_D_PSID	Fallzahl der Versicherten
RT_2	Beschreibung	Altersgruppe	Altersgruppe der betrachteten Versicherten
RT_2	Beschreibung	Geschlecht	Geschlecht der betrachteten Versicherten
RT_2	Fallzahl	CNT_D_PSID	Fallzahl der Versicherten

2.3 Struktur der Ergebnismenge

Die Anzahl exportierbarer Ergebnistabellen, sind pro Antrag auf fünf Tabellen begrenzt. Dies dient der fristgemäßen Durchführbarkeit der Identifikationsschutzprüfung und sichert antragsübergreifend eine zügige Verfahrensabwicklung. Falls antragstellende Personen mehr als fünf Tabellen anfordern, werden nach Rücksprache maximal fünf Tabellen für den Export freigegeben.

Weiterhin darf die Gesamtzahl der Zellen aller in der Ergebnismenge enthaltenen Ergebnistabellen 2400 Zellen (Klassifikation Datenzeile, vgl. Programmcoderegeln) nicht überschreiten.

Das Forschungsdatenzentrum Gesundheit behält sich vor, in Einzelfällen die Gesamtzahl der Zellen einer Ergebnismenge zu beschränken, wenn dies zur Sicherstellung der Prüfbarkeit erforderlich ist um die Kapazitäten des FDZ nicht unverhältnismäßig zu binden.

2.4 Inhalt der Ergebnismenge

Die beantragte Ergebnismenge wird bei der Prüfung des Antrags inhaltlich geprüft. Bei einem angeforderten Export der Ergebnismenge wird diese mit der beantragten Ergebnismenge abgeglichen. Dabei werden der Antrag, Nachrichten im Antragsverlauf und eventuell erfolgte Korrekturläufe berücksichtigt. Werden Abweichungen zwischen beantragter und zu exportierender Ergebnismenge festgestellt, werden zunächst Rückfragen über die Nachrichtenfunktion des Antragsportals an die antragstellende Person gerichtet. Die inhaltliche Prüfung der zu exportierenden Ergebnismenge muss evtl. wiederholt werden, was die Antragsbearbeitung verzögert. Der Antrag kann negativ beschieden werden, wenn es nicht hinreichend erklärte Abweichungen zwischen der beantragten Ergebnismenge und der zum Export angeforderten Ergebnismenge gibt. Hierbei gilt ein Ermessensspielraum für die FDZ-Mitarbeitenden, wobei geringfügige und nachvollziehbare Abweichungen in der Regel nicht zu einer negativen Bescheidung des Antrags führen.

Die inhaltliche Prüfung der Ergebnismenge muss wiederholt werden, wenn mindestens eines dieser Kriterien erfüllt ist (Änderungen in Umfang, Struktur oder Inhalt):

1. Es gibt Abweichungen in der Anzahl der Ergebnistabellen
2. Es gibt Abweichungen in der Größe (Zeilen und Spalten) der Ergebnistabellen
3. Es gibt inhaltliche Abweichungen (Spaltennamen, Zeilenamen, erwartete Wertebereiche)

Forschungsdatenzentrum Gesundheit

2.5 Klassifikation von Ergebnistabellen

Jede Ergebnistabelle wird nach verschiedenen Gesichtspunkten klassifiziert. Die anzuwendenden Geheimhaltungsregeln unterscheiden sich je nach Klassifikation.

2.5.1 Bezug

Versichertenbezogen: Die Auswertung bezieht sich auf Versicherte (z.B. "Anzahl Versicherte mit Erkrankung X in Region Y")

Leistungsträgerbezogen: Die Auswertung bezieht sich auf Leistungsträger, z.B. bei krankenkassenbezogenen Auswertungen von Daten mehrerer Krankenkassen. (z.B. "Anzahl Versicherte in Krankenkasse A und Krankenkasse B")

Leistungserbringerbezogen: Die Auswertung bezieht sich auf Leistungserbringer, z.B. bei leistungserbringerbezogenen Auswertungen von Daten mehrerer Leistungserbringer. (z.B. "Anzahl behandelte Patienten pro Leistungserbringer in Region X")

Es ist möglich, dass ein Ergebnis mehrere Bezüge aufweist (Mehrfachauswahl möglich).

2.5.2 Statistische Klassifikation

Jedes Ergebnis wird einer bestimmten Kategorie zugeordnet (grundsätzlich ist jedoch auch eine Mehrfachauswahl möglich, vgl.

Forschungsdatenzentrum Gesundheit

Tabelle 5). Diese Klassifikation erfolgt lediglich auf Basis der funktionalen Form des Ergebnisses, nicht auf Basis der verwendeten Daten (Hundepool et al. 2012). Auf Basis dieser Klassifikation kann ein Ergebnis als "niedriger Prüfbedarf" oder "hoher Prüfbedarf" eingestuft werden (in der Literatur "sicher/safe" und "unsicher/unsafe", vgl. Hundepool et al. 2012; auch ONS 2022).

Ergebnisse mit niedrigem Prüfbedarf: Können in der Regel freigegeben werden, wenn sich an gute wissenschaftliche Praxis gehalten wurde. Ausnahme, die Ergebnismengenprüfung des FDZ kommt zu dem Schluss, dass diese nicht freigegeben werden können oder modifiziert werden müssen. Ergebnisse mit niedrigem Prüfbedarf können in der Regel ohne oder mit nur geringfügigen Änderungen freigegeben werden (Beweislast liegt beim FDZ).

Ergebnisse mit hohem Prüfbedarf: Können nicht freigegeben werden, es sei denn, Antragstellende legen nachvollziehbar dar, dass für dieses konkrete Ergebnis mit diesem Inhalt und in diesem Kontext kein Reidentifikationsrisiko besteht (anhand der Regeln zur Ergebnismengenprüfung, eigener Modifikationen der Ergebnisse und ggf. weiterer Informationen). Diese Ergebnisse müssen gründlich geprüft werden und hier muss sichergestellt sein, dass keine Verletzung der statistischen Geheimhaltung erfolgen kann (Beweislast liegt bei den Forschenden). Alle in der Tabelle nicht aufgeführten Formen von Ergebnissen gelten zunächst als "mit hohem Prüfbedarf".

Beide Arten von Ergebnissen müssen gemäß der hier aufgestellten Regeln geprüft werden. Bei begrenzter Prüfzeit sollte sich jedoch vor allem auf als "mit hohem Prüfbedarf" klassifizierte Ergebnisse fokussiert werden.

Forschungsdatenzentrum Gesundheit

Tabelle 5: Arten von Ergebnistabellen

Statistik	Ergebnis	Prüfbedarf
Deskriptive Statistiken	Häufigkeitstabelle	Hoch
	Kennzahltable/Größentabelle	Hoch
	Maxima, Minima	Hoch
	Perzentile, Quartile (inkl. Median)	Hoch
	Modus	Niedrig
	Lineare Aggregate (Mittelwert, Verhältnisse, Indizes, Indikatoren, Anteile)	Hoch
	Konzentrationsmaße (z.B. Gini-Koeffizient)	Niedrig
	Höhere Momente (z.B. Varianz, Kovarianz, Schiefe, Wölbung)	Niedrig
	Grafiken (derzeit nicht exportierbar)	Hoch
Korrelationen, Regressionanalyse	Modellkoeffizienten	Niedrig
	Residuen	Hoch
	Hypothesentests	Niedrig
	Korrelationskoeffizienten	Niedrig

Tabelle leicht angepasst nach Bond et al., 2015.

3 Grundregeln

Folgende Grundsätze der Ergebnismengenprüfung gelten für alle Arten von Ergebnissen:

Tabelle 6: Grundsätze der Ergebnismengenprüfung

ID	Prüfpunkt	Erläuterung
1	Übergreifendes	
1.1	Beschreibung	Eine Ergebnismenge muss selbsterklärend sein oder in den DD_Tabellen genau beschrieben sein: Ergebnistabellen, Spalten, Zeilen, selbst erstellte Variablen (Beispiel: Alter_MW). Eine Ergebnismenge muss ohne projektspezifische Kenntnisse verständlich sein. Dies beinhaltet, dass eventuell vorhandene Abkürzungen erläutert werden müssen.
1.2	Wertebereiche	Die Werte der Ergebnistabelle müssen innerhalb der für diese Art der Ergebnistabelle erwarteten Wertebereiche liegen. Die Datentypen entsprechen den erwarteten Datentypen.
1.3	Pseudonyme	Eine Ergebnismenge darf keine Pseudonyme enthalten. Dies gilt für Pseudonyme von Versicherten, Leistungserbringern und Leistungsträgern. Die einzige Ausnahme bildet eine beantragte Krankenkassen-bezogene Auswertung, bei der die Einwilligung der betroffenen Kassen durch den Antragsteller eingeholt wurde. Hier ist die Verwendung der Betriebsnummer erlaubt.
1.4	Aggregation	Ergebnismengen werden nur aggregiert bereitgestellt. Es werden keine Einzelangaben zur Verfügung gestellt.
1.5	Regionalauswertungen	Regionalauswertungen werden nur in der Detailtiefe bereitgestellt, die im Antrag genehmigt wurde.
2	Versicherte	
2.1	Mindestfallzahl	In Tabellen mit Versichertenbezug tragen zu jedem Zellwert mindestens MFZ Versicherte bei. Dies gilt auch, wenn statt absoluter Werte Anteile berichtet werden (hier müssen zu den Werten des Zählers mindestens MFZ Versicherte beitragen). Zu prüfen anhand der Spalte "CNT_D_PSID", welche jeder Tabelle angefügt werden muss. Statistische Modelle müssen über mindestens MFZ Freiheitsgrade verfügen (Freiheitsgrade = Anzahl Beobachtungen - Anzahl Parameter - Anzahl Restriktionen). Nullwerte von Fallzahlen sind nicht erlaubt außer im Fall einer "strukturellen Null". In diesem Fall ist der einzige logisch mögliche Wert Null. Beispielsweise wird jemand, der nicht von Diabetes betroffen ist, nicht am Disease Management Program für Diabetes teilnehmen. Wenn zur Erkennung einer strukturellen Null forschungsspezifisches Fachwissen erforderlich ist, muss die antragstellende Person es dem FDZ erläutern (z.B. Medikamente, die erst im späteren Untersuchungszeitraum zur Verfügung standen). Strukturelle Nullen bilden demnach die einzige Ausnahme von der Mindestfallzahlregel.

Forschungsdatenzentrum Gesundheit

2.2	Gruppen	Informationen über Gruppen sind zu schützen (class disclosure). Es darf nicht möglich sein, eine informative Aussage über 100% der Personen in einer Gruppe zu treffen. Nullwerte von Fallzahlen sind nicht erlaubt außer im Fall einer "strukturellen Null" (der einzige logisch mögliche Wert ist Null). Dieser Prüfpunkt ist die einzige Ausnahme von einer rein regelbasierten Prüfung, da diese unter Umständen forschungsbehindernd wäre. Wenn ein Ergebnis nach den genannten Regeln nicht freigegeben werden kann, dann prüft das FDZ tiefergehend. Hierfür greifen FDZ-Mitarbeitende auf das bei ihnen vorhandene Domänenwissen zurück ohne tiefergehende zusätzliche Recherche. Wenn auf diesem Weg nicht geklärt werden kann, ob ein geheimzuhaltener Fall vorliegt, dann wird noch tiefer (prinzipienbasiert) geprüft. Hier ist vor Freigabe des Ergebnisses dieses in den Kontext der "5 Safes¹" zu setzen (wer stellt den Antrag, zu welchem Zweck, ...). Im Zweifel kann der Antrag dem Arbeitskreis Versorgungsdaten vorgelegt werden.
2.3	Rückrechnenmöglichkeiten (innerhalb einer Ergebnistabelle sowie tabellenübergreifend für die gesamte Ergebnismenge)	Differenzmengen Differenzmengen: Zwei Ergebnistabellen oder Zellen in einer Ergebnistabelle beinhalten Aussagen über sich überlappende Gruppen. Beispiel: Es wird berechnet, wie viele Versicherte Krankheit A, aber nicht Krankheit B haben. Dann wird berechnet, wie viele Versicherte Krankheit B, aber nicht Krankheit A haben. Wenn dann noch berechnet wird, wie viele Versicherte entweder Krankheit A oder Krankheit B haben, so ist rückrechenbar, wie viele Versicherte sowohl Krankheit A als auch Krankheit B haben. Vorgabe: Forschende haben darauf zu achten, MFZ-unterschreitende Differenzbildung zu vermeiden. Wird festgestellt, dass unbeabsichtigt eine solche Differenz gebildet werden kann, so ist dies dem FDZ unverzüglich zu melden. Nullwerte von Fallzahlen sind nicht erlaubt außer im Fall einer "strukturellen Null" (der einzige logisch mögliche Wert ist Null). Restgruppen Restgruppen: Gruppen, welche für die angestrebte Auswertung nicht interessant sind, aber bei diesen als "Randprodukt" anfallen. Beispiel: In einer Spalte wird eine Auswertung für Männer durchgeführt, in einer anderen für Frauen. Werden gleichzeitig Randsummen angegeben, so sind die Ergebnisse für "divers" rückrechenbar. Vorgabe: Restgruppen müssen ausgewiesen werden (d.h. die Ergebnistabelle muss um die Restgruppe ergänzt werden). Die ausgewiesenen Restgruppen müssen ebenfalls die Mindestfallzahl einhalten. Nullwerte von Fallzahlen sind nicht erlaubt außer im Fall einer "strukturellen Null" (der einzige logisch mögliche Wert ist Null).
3	Leistungsträger (krankenkassenbezogene Auswertung)	
3.1	Gruppen (Krankenkassen)	Informationen über Gruppen sind zu schützen (class disclosure). Es darf nicht möglich sein, eine informative Aussage über 100% der Krankenkassen (oder Gruppen von Krankenkassen) zu treffen. Nullwerte von Fallzahlen sind nicht erlaubt außer im Fall einer "strukturellen Null" (der einzige logisch mögliche Wert ist Null).

¹ Desai et al. (2016)
Nebenbestimmungen – Anlage 3

Forschungsdatenzentrum Gesundheit

3.2	Geschäfts- und Betriebsgeheimnisse von Krankenkassen	Hat eine Krankenkasse ihre Einwilligung nach § 17 Absatz 2 FDZGesV zu einer Krankenkassen-bezogenen Auswertung erteilt, trifft das Forschungsdatenzentrum soweit wie möglich Vorsorge dafür, dass durch diese Auswertungen in Kombination mit weiteren Auswertungen Erkenntnisse über andere Krankenkassen, die keine Einwilligung nach § 17 Absatz 2 FDZGesV zu einer Auswertung erteilt haben, nicht gewonnen werden können.
3.3	Krankenkassenbezogene Rückrechenmöglichkeiten (innerhalb einer Ergebnistabelle sowie tabellenübergreifend für die gesamte Ergebnismenge)	Differenzmengen Differenzmengen: Zwei Ergebnistabellen oder Zellen in einer Ergebnistabelle beinhalten Aussagen über sich überlappende Gruppen. Vorgabe: Forschende haben darauf zu achten, MFZ-unterschreitende Differenzbildung zu vermeiden. Wird festgestellt, dass unbeabsichtigt eine solche Differenz gebildet werden kann, so ist dies dem FDZ unverzüglich zu melden. Nullwerte von Fallzahlen sind nicht erlaubt außer im Fall einer "strukturellen Null" (der einzige logisch mögliche Wert ist Null). Restgruppen Restgruppen: Gruppen, welche für die angestrebte Auswertung nicht interessant sind, aber bei diesen als "Randprodukt" anfallen. Vorgabe: Restgruppen müssen ausgewiesen werden (d.h. die Ergebnistabelle muss um die Restgruppe ergänzt werden). Nullwerte von Fallzahlen sind nicht erlaubt außer im Fall einer "strukturellen Null" (der einzige logisch mögliche Wert ist Null).
4	Leistungserbringer	
4.1	Regionalauswertungen (Leistungserbringer)	Regionalauswertungen, welche sich auf Leistungserbringer beziehen, dürfen nur auf Bundesland- oder auf PLZ-2-Steller-Ebene erfolgen.

4 Modifikationsmöglichkeiten

Wenn Ergebnismengen von den FDZ-Mitarbeitenden nicht freigegeben werden können, so werden die Forschenden über das Antragsportal darüber informiert. Die FDZ-Mitarbeitenden stoßen in diesem Fall den "Korrekturlauf Ergebnismengenprüfung" an. Dabei wird der Analyserraum erneut geöffnet und Datennutzende können das Analyseskript erneut modifizieren und eine weitere Lauffähigkeits- und Ergebnismengenprüfung anstoßen. Zur Unterstützung einer Anpassung des Skripts können die Datennutzenden im Git Repository das Prüfskript einsehen, in welchem auf bestehende Probleme hingewiesen wird und Lösungsvorschläge unterbreitet werden können. Darüber hinaus geben die FDZ-Mitarbeitenden den Forschenden eine strukturierte Rückmeldung zur Prüfung der Ergebnismenge. Die Anzahl möglicher Korrekturläufe wird, um zügige Verfahren in der Summe zu gewährleisten, i.d.R. auf einen Korrekturlauf begrenzt. Für in Anspruch genommene Korrekturläufe können Gebühren erhoben werden.

Zulässige Modifikationsmöglichkeiten, welche die Forschenden im Falle einer Regelverletzung anwenden können, basieren auf den Prinzipien von Umstrukturierung und Generalisierung. Die gesamte Tabelle wird dabei umstrukturiert, z.B. durch Generalisieren, Zusammenfassen von Zeilen, Zusammenfassen von Spalten, Entfernen von Zeilen, Entfernen von Spalten, Intervalle verändern bzw. Ausprägungen zusammenfassen.

Die Modifikationsmöglichkeit der Unterdrückung von Zellwerten, die die **MFZ** nicht einhalten, ist in der Regel nicht erlaubt. Bei einer solchen Unterdrückung müssen Rückrechnmöglichkeiten über andere Zellwerte, Randsummen, andere Ergebnisse etc. ausgeschlossen werden und in der Regel führt dies zu weiteren Unterdrückungen. Diese Prüfung ist sehr komplex und zeitaufwändig und kann durch das FDZ nicht geleistet werden.

Gleiches gilt für andere Modifikationsmöglichkeiten, welche die Daten einzelner Zellen verändern, ohne die dahinter liegende Fallzahl zu verändern. Beispiele hierfür sind Runden, stochastische Überlagerung oder die Imputation von Werten. Diese Modifikationsmöglichkeiten sind ebenfalls in der Regel nicht erlaubt.

Das FDZ behält sich vor, in Ausnahmefällen eine Ergebnismenge selbst zu modifizieren. Eine Modifikation der Ergebnismenge durch FDZ-Mitarbeitende kann nur posttabular erfolgen. D.h. die durch die Forschenden erzeugten RT_ Tabellen werden in Kopie ("RT_X_EXP") modifiziert. Dabei minimiert das FDZ das Reidentifikationsrisiko von Versicherten unter angemessener Wahrung des angestrebten wissenschaftlichen Nutzens durch geeignete Maßnahmen. Antragstellende können bei starker Modifikation neue Anträge einreichen und Ergebnismengen passender aggregieren.

5 Spezielle Regeln zur Prüfung der Ergebnisse nach Klassifikation

5.1 Häufigkeitstabellen

Ergebnistabellen und Zwischenergebnisse werden in einer SQL-Datenbank gespeichert. Inhalte von Spalten müssen atomar sein. Das bedeutet, dass pro Spalte nicht mehrere Werte enthalten sein dürfen (nicht erlaubt: „Weiblich, 0-50 Jahre“, erlaubt: „Weiblich“ & „0-50 Jahre“, s.u.). Nicht atomare Spalten müssen in atomare Spalten aufgeteilt werden. Komplexere Tabellenformate wie z.B. Kreuztabellen müssen für Zwischenergebnis- und Ergebnistabellen so formatiert abgespeichert werden. Forschende können außerhalb des FDZ-Systems beliebig komplexe Tabellen aus den normalisierten Ergebnistabellen mit entsprechender Software selbst automatisiert oder manuell zusammensetzen.

Die folgende Tabellenstruktur ist erlaubt:

Tabelle 7: Erlaubte Tabellenstrukturen

Geschlecht	Altersgruppe	CNT_D_PSID
Weiblich	0-50 Jahre	x ₁
Weiblich	51-100 Jahre	x ₂
Männlich	0-50 Jahre	x ₃
Männlich	51-100 Jahre	x ₄
Divers	0-50 Jahre	x ₅
Divers	51-100 Jahre	x ₆

Die folgende Tabellenstruktur ist nicht erlaubt, da die verwendeten Variablen (Geschlecht, Altersgruppe) nicht getrennt abfragbar atomar in getrennten Spalten gespeichert werden:

Tabelle 8: nicht erlaubte Tabellenstrukturen

Variable	CNT_D_PSID
Weiblich_0_50_Jahre	x ₁
Weiblich_51_100_Jahre	x ₂
Männlich_0_50_Jahre	x ₃
Männlich_51_100_Jahre	x ₄
Divers_0_50_Jahre	x ₅
Divers_51_100_Jahre	x ₆

Forschungsdatenzentrum Gesundheit

Die folgende Tabellenstruktur ist ebenfalls nicht erlaubt. Die letzte Zeile und letzte Spalte darf nicht enthalten sein. Die Summen können nachträglich eigenständig berechnet werden.

Tabelle 9: nicht erlaubte Tabellenstrukturen (Fortsetzung)

	0-50 Jahre	51-100 Jahre	Gesamt (Randsumme)
Weiblich	X ₁	X ₂	X ₁ + X ₂
Männlich	X ₃	X ₄	X ₃ + X ₄
Divers	X ₅	X ₆	X ₅ + X ₆
Gesamt (Randsumme)	X ₁ + X ₃ + X ₅	X ₂ + X ₄ + X ₆	X ₁ + X ₂ + X ₃ + X ₄ + X ₅ + X ₆

5.2 Kennzahltabellen/Größentabellen

Eine Kennzahltable enthält neben den Häufigkeiten/Fallzahlen zusätzlich konkrete Werte von Kennzahlen, z.B. eine Summe, einen Mittelwert etc. Für Kennzahltabellen sind die Grundregeln anzuwenden und zusätzlich spezifische Regeln je nachdem, um welche Kennzahlen es sich handelt.

Beispiel für korrekte Formatierung:

Tabelle 10: Beispiel korrekte Formatierung einer Kennzahltable

Geschlecht	Altersgruppe	Anzahl Krankenhausaufenthalte	CNT_D_PSID
Weiblich	0-50 Jahre	13200	13065
Weiblich	51-100 Jahre	50098	47869
Männlich	0-50 Jahre	10787	10507
Männlich	51-100 Jahre	40987	40565

Beispiel für mehrere Kennzahlen in einer Tabelle:

Tabelle 11: Beispiel Ergebnistable für mehrere Kennzahlen

Kennzahl	Wert	CNT_D_PSID
Mittelwert	2	4
Standardabweichung	15	1

NICHT erlaubt (keine normalisierte Datenbankstruktur):

Tabelle 12: Beispiel inkorrekte Formatierung einer Kennzahltable

	0-50 Jahre	51-100 Jahre
Weiblich	13200	50098
Männlich	10787	40987

5.3 Extremwerte (Minima, Maxima, Spannweite)

Tabelle 13: Spezialregeln der Ergebnismengenprüfung: Extremwerte

Prüfpunkt	Erläuterung
Allgemein	Extremwerte werden nicht bereitgestellt. (Ausnahmen strukturelle Null, z.B. Minimum Alter=0)
Spannweite	Die Spannweite kann nicht bereitgestellt werden, da Maximum und Minimum nicht bereitgestellt werden.

5.4 Perzentile (inkl. Median)

Tabelle 14: Spezialregeln der Ergebnismengenprüfung: Perzentile

Prüfpunkt	Erläuterung
Allgemein	Perzentile werden nicht bereitgestellt. Eine Ausnahme gilt für Quartile (25%, 50% (Median), 75%), wenn sichergestellt ist, dass die Mindestfallzahl in den jeweils gebildeten Gruppen eingehalten wird. Das bedeutet, es müssen mindestens MFZ Personen zu einem Quartil sowie zu der Restgruppe beitragen: • beim Median muss jede Hälfte die MFZ erfüllen. Die Gesamtanzahl muss daher mindestens dem Zweifachen der MFZ entsprechen (bei MFZ 30 wären demnach mindestens 60 Personen in der Gesamtgruppe erforderlich) • beim 25- und 75-Perzentil muss die Anzahl der Personen in der Gruppe mehr als das Vierfache der MFZ betragen, da jedes Quartil einem Viertel der Gruppe entspricht (bei MFZ 30 wären demnach mindestens 120 Personen in der Gesamtgruppe erforderlich) Es ist nicht notwendig zu fordern, dass MFZ Personen genau den Wert des Quartils haben müssen.

5.5 Modus

Tabelle 15: Spezialregeln der Ergebnismengenprüfung: Modus

Prüfpunkt	Erläuterung
Allgemein	Der Modus kann bereitgestellt werden, wenn MFZ Versicherte dazu beitragen. Der Modus wird nicht bereitgestellt, wenn alle Werte dem Modus entsprechen.

5.6 Lineare Aggregate

Beispiele für lineare Aggregate:

- Summe
- Mittelwert
- Verhältnisse
- Prozentuale Anteile
- Indizes
- Indikatoren

Tabelle 16: Spezialregeln der Ergebnismengenprüfung: Lineare Aggregate

Prüfpunkt	Erläuterung
Übergreifendes	
Allgemein	Je nach Komplexität der Berechnung haben Kennzahlen einen hohen oder einen niedrigen Prüfbedarf. Je komplexer die Berechnung, desto schwieriger ist eine Rückrechnung. Beispiele für komplexe Kennzahlen (tendenziell sicher): • Fisher Preisindex • Kombinationen von wenig komplexen Kennzahlen Beispiele für wenig komplexe Kennzahlen (tendenziell unsicher): • Mittelwert • Verhältnisse (Summe X / Summe Y)
Dichotome Variablen	Bei dichotomen Variablen (nur Ausprägungen 0 und 1) können Prozentzahlen und Mittelwerte zu Häufigkeitstabellen umgewandelt werden. Es gelten demnach die Regeln für Häufigkeitstabellen (d.h. z.B. die Mindestfallzahlregel gilt für die einzelnen Kategorien und nicht für die Gesamtzahl Personen, die in die Berechnung eingeflossen sind).
Prozentzahlen	Die Mindestfallzahl-Regel ist hier auf den Zähler anzuwenden.
Verhältnisse	Die Mindestfallzahl-Regel ist hier auf die jeweiligen Zähler anzuwenden. z.B. bei Odds Ratio müsste von beiden Argumenten jeweils der Zähler die Mindestfallzahl einhalten.
Beschreibung	Die verwendete Formel zur Berechnung der Kennzahl muss immer angegeben werden. Dies kann, wenn möglich, in der DD-Tabelle erfolgen, ansonsten im Skript oder in einem über das Antragsportal hochgeladenen Dokument.

Forschungsdatenzentrum Gesundheit

Versicherte	
Mindestfallzahl	Die Kennzahlen können bereitgestellt werden, wenn mindestens MFZ Versicherte zu diesem Wert beitragen. Prozentzahlen sollten dafür u.U. ohne Dezimalstellen dargestellt werden. Nullwerte sind nicht erlaubt außer im Fall einer "strukturellen Null" (der einzige logisch mögliche Wert ist Null).

5.7 Überlebenszeitanalyse

Beispiele für Überlebenszeitanalysen:

- Kaplan-Meier-Schätzer
- Hazard Ratios

Tabelle 17: Spezialregeln der Ergebnismengenprüfung: Überlebenszeitanalyse

Prüfpunkt	Erläuterung
Kaplan-Meier-Schätzer	Für den Kaplan-Meier-Schätzer bezieht sich die Einhaltung der Mindestfallzahl auf die Versicherten unter Risiko. Für die Angabe des Kaplan-Meier-Schätzers gilt, dass mindestens MFZ Versicherte unter Risiko zu diesem Wert beitragen. D.h. die Spalte "CNT_D_PSID" muss die Personen unter Risiko angeben.

5.8 Standardabweichung, Varianz, Schiefe, Wölbung

Tabelle 18: Spezialregeln der Ergebnismengenprüfung: Standardabweichung, Varianz, Schiefe, Wölbung

Prüfpunkt	Erläuterung
Standardabweichung	Die Standardabweichung darf nicht exakt 0 sein.
Modifikation	Es gibt keine sinnvollen Modifikationsmöglichkeiten für diese Kennzahlen, wenn es zu wenige Beobachtungen gibt.

5.9 Korrelationskoeffizienten

Tabelle 19: Spezialregeln der Ergebnismengenprüfung: Korrelationskoeffizienten

Prüfpunkt	Erläuterung
Schutz von Gruppen	Der Korrelationskoeffizient darf nicht genau -1 oder 1 sein (100% Aussage über Gruppen möglich).

Modifikation	Nicht-Bereitstellung oder Recodierung können für diese Kennzahlen (ausnahmsweise) sinnvolle Modifikationsmöglichkeiten sein.
--------------	--

5.10 Konzentrationsmaße

Beispiele für Konzentrationsmaße:

- Gini-Koeffizient
- Herfindahl-Index
- absolute und relative Konzentration
- ...

Tabelle 20: Spezialregeln der Ergebnismengenprüfung: Konzentrationsmaße

Prüfpunkt	Erläuterung
Mindestfallzahl	Die Kennzahlen können bereitgestellt werden, wenn mindestens MFZ Versicherte zu diesem Wert beitragen. Prozentzahlen sollten dafür u.U. ohne Dezimalstellen dargestellt werden. Nullwerte sind nicht erlaubt außer im Fall einer "strukturellen Null" (der einzige logisch mögliche Wert ist Null).

Als sinnvolle Modifikationsmöglichkeit könnte hier auch eine Rangfolge statt der exakten Werte angegeben werden.

5.11 Modellkoeffizienten

Beispiele für Modellkoeffizienten:

- Lineare Regression
- Nichtlineare Regression
- ANOVA
- ANCOVA

Tabelle 21: Spezialregeln der Ergebnismengenprüfung: Modellkoeffizienten

Prüfpunkt	Erläuterung
Mindestfallzahl	Die Koeffizienten der unter "Beispiele" genannten Modelle können bereitgestellt werden, wenn die Modelle über mindestens MFZ Freiheitsgrade (residual degrees of freedom) verfügen.
Saturiertes Modell	Das Modell darf nicht saturiert sein, d.h. wenn ausschließlich kategoriale unabhängige Variablen vorliegen und diese vollständig interagiert werden. Ein solches Modell sollte als Tabelle umstrukturiert werden.
Dichotome Variablen	Das Modell darf nicht nur als einzige unabhängige Variable eine dichotome Variable enthalten. Ein solches Modell sollte als Tabelle umstrukturiert werden.

5.12 Hypothesentests

Beispiele für Hypothesentests:

- T-Statistik
- F-Statistik
- p-Werte
- MSE
- Chi²-Test
- R²
- Standardfehler
- Informationskriterien (AIC, BIC, ...)
- Andere Teststatistiken (Wald, Hausman, ...)

Tabelle 22: Spezialregeln der Ergebnismengenprüfung: Hypothesentests

Prüfpunkt	Erläuterung
Mindestfallzahl	Die unter "Beispiele" genannten Kennzahlen können bereitgestellt werden, wenn die Modelle über mindestens MFZ Freiheitsgrade (residual degrees of freedom) verfügen.

5.13 Residuen

Residuen können nicht bereitgestellt werden, da es sich um Einzelfallangaben handelt bzw. diese ableitbar sein können.

5.14 Faktorenanalyse

Tabelle 23: Spezialregeln der Ergebnismengenprüfung: Faktorenanalyse

Prüfpunkt	Erläuterung
Allgemein	Faktorscores können bereitgestellt werden, solange jeder Faktor mit mindestens zwei Variablen in Verbindung steht. Erkennbar an den Faktorladungen.
Mindestfallzahl	Jede Faktorenanalyse sollte anhand einer hinreichend großen Fallzahl durchgeführt werden, da ansonsten die statistische Aussagekraft in Frage zu stellen ist.

Tabelle 24: Spezialregeln der Ergebnismengenprüfung: Einschluss-/Ausschlusskriterien für Studienpopulationen

Prüfpunkt	Erläuterung
Mindestfallzahl	Die Anzahl Versicherter in jedem Ausschlusskriterium muss mindestens MFZ sein.
Gruppen	Durch Einschluss-/Ausschlusskriterien darf keine 100% Aussage über Gruppen möglich sein. Der Fall tritt ein, wenn nach einem Auswahlsschritt die Fallzahl unverändert bleibt oder die Fallzahl auf 0 sinkt.
Rückrechnemöglichkeiten	Rückrechnemöglichkeiten können sich ergeben, wenn Populationen auf verschiedene Arten gebildet werden und mehrere Möglichkeiten exportiert werden sollen. Die Grundregeln für Differenzmengen und Restgruppen sind anzuwenden.

5.16 Grafiken

Beispiele für Grafiken:

Jegliche Art von Grafiken, z.B.

- Boxplots
- Scatterplots
- Residualplots
- ...

Tabelle 25: Spezialregeln der Ergebnismengenprüfung: Grafiken

Prüfpunkt	Erläuterung
Allgemein	Grafiken werden nicht als Ergebnis zugelassen. Auch technisch ist dies aktuell nicht möglich. Wenn Forschende eine Grafik erstellen wollen, müssen sie die zugrundeliegenden Daten als Ergebnis produzieren und nach den jeweils geltenden Regeln (z.B. für Häufigkeitstabellen, Perzentile etc.) freigeben lassen.

6 Literatur

Bond, S., Brandt, M. & De Wolf, P.-P. (2015) Guidelines for the checking of output based on microdata research. Data Without Boundaries project standalone document.

Desai T, Ritchie F, Welpton R (2016) Five Safes: Designing data access for research. Economics Working Paper Series 1601. University of the West of England.

<https://www2.uwe.ac.uk/faculties/BBS/Documents/1601.pdf>

Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder (Hrsg.). Regelungen zur Auswertung von Mikrodaten in den Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder (FDZ). 2022. <https://www.forschungsdatenzentrum.de/de/geheimhaltung>

Griffiths, E., Greci, C., Kotrotsios, Y., Parker, S., Scott, J., Welpton, R., Wolters, A. & Woods, C. Handbook on statistical disclosure control for outputs. Version 1.0. 2019.

Hochfellner, D., Müller, D., Schmucker, A. & Roß, E. Datenschutz am Forschungsdatenzentrum. FDZ Methodenreport 06/2012.

Hundepool, A., et al.: Statistical Disclosure Control. Wiley, Hoboken (2012)

Office for National Statistics (2022). SRS Output Checking Guidance Document. Work Strand: Statistical Disclosure Control. v.1.1

Ritchie, F., Tilbrook, A., Green, E., White, P., Derrick, B., & Kendall, C. (2023). The SACRO guide to statistical output checking (Version 1). Zenodo. <https://doi.org/10.5281/zenodo.10282526>

Stats NZ. Microdata output guide (fifth edition). 2022. Retrieved from www.stats.govt.nz