

Ergänzung zu Bericht über  
*Medizinische Versorgung  
in Pflegeheimen in Österreich:*  
Rohdatenanalyse

---

Barbara Glock, Melanie Zechmeister  
barbara.glock@dexhelpp.at  
11.03.2018

## Inhaltsverzeichnis

1	Einleitung.....	5
2	Daten .....	5
3	Analyse der Rohdaten .....	6
3.1	Der Datensatz <i>Pfif_Eintritte</i> zu Eintritten und Übergangsgrund in ein Pflegeheim .....	6
3.1.1	Variable <i>Pseudonr_patient</i> : Patienten ID .....	7
3.1.2	Variable <i>Pstu_ext</i> : Pflegestufe der Person .....	8
3.1.3	Variable <i>Uebg</i> : Übergangsgrund in das Pflegeheim.....	9
3.1.4	Variable <i>Plz</i> : Postleitzahl der Person .....	9
3.1.5	Variable <i>Bld</i> : Bundesland der Person .....	10
3.1.6	Variable <i>AAlder</i> : Alter einer Person .....	11
3.1.7	Variable <i>Tdat</i> : Sterbedatum einer Person.....	12
3.1.8	Variable <i>Begi</i> : Beginndatum .....	12
3.2	Die Datensätze <i>Vers_daten</i> zu den Versichertendaten.....	13
3.2.1	Variable <i>ALTDEKA</i> : Altersdekade der Person .....	14
3.2.2	Variable <i>Geschlecht</i> : Geschlecht .....	15
3.2.3	Variable <i>Toddatum</i> : Sterbedatum der Person .....	16
3.2.4	Variable <i>Gebdatum</i> : Geburtsdatum der Person.....	16
3.2.5	Variable <i>PLZ</i> : Postleitzahl der Person .....	17
3.2.6	Variable <i>Von</i> : Beginndatum des Dateneintrages der Versicherung.....	18
3.2.7	Variable <i>Bis</i> : Datum zum Versicherungsende.....	19
3.2.8	Variable <i>Vers_key</i> : Dimensionsschlüssel der Versichertendimension .....	20
3.2.9	Variable <i>Pseudonr_patient</i> : Patienten ID .....	21
3.3	Die Datensätze <i>Med_versorgung</i> zu den Leistungsdaten .....	21
3.3.1	Variable <i>Leistungsart</i> : Leistungsart .....	22
3.3.2	Variable <i>Positionsnummer</i> : Identifikationsnummer der Leistung.....	23
3.3.3	Variable <i>Positionsbezeichnung</i> .....	24
3.3.4	Variable <i>Fachgruppe</i> .....	25
3.3.5	Variable <i>Leistungszeitraum</i> .....	26
3.3.6	Variable <i>Vers_key</i> : Dimensionsschlüssel der Versichertendimension .....	26
3.3.7	Variable <i>Lstg_key</i> : Dimensionsschlüssel der Leistungsdimension .....	27
3.3.8	Variable <i>Anzahl</i> .....	28
3.4	Der Datensatz <i>Pflegeheime</i> .....	28
3.4.1	Variable <i>Name</i> .....	29

3.4.2	Variable <i>Ort</i> .....	29
3.4.3	Variable <i>PLZ: Postleitzahl</i> .....	30
3.4.4	Variable <i>Wohnplätze</i> .....	31
3.4.5	Variable <i>Pflegeplätze</i> .....	31
3.4.6	Variable <i>Wohnplätze mit Pflegemöglichkeit</i> .....	32
3.4.7	Variable <i>betreutes/betreubares Wohnen</i> .....	33
3.4.8	Variable <i>ärztliche Betreuung text</i> .....	34
3.4.9	Variable <i>ärztliche Betreuung code</i> .....	34
4	Zusammenfassung der Ergebnisse .....	35
5	Anhang .....	37
6	Tabellenverzeichnis .....	39
7	Abbildungsverzeichnis .....	40
8	Quellenverzeichnis .....	40

## Abkürzungsverzeichnis

<i>MedVers</i>	Abkürzung für den Bericht <i>Medizinische Versorgung in Pflegeheimen in Österreich</i> des Hauptverbandes des österreichischen Sozialversicherungsträger [1], der die Grundlage für diese Ergänzung darstellt
FOKO	FOlgeKOsten - Abrechnungsdaten der Krankenversicherungsträger
<i>Pfif_Eintritte</i>	Kurzbezeichnung des Datensatz <i>Pfif_uebg_2014_2016</i> aus der Pflegegeld-Datenbank
<i>Vers_daten</i>	Datensatz zu den Versicherten (Versichertendaten)
<i>Med_versorgung</i>	Datensatz zu den Leistungen (Leistungsdaten)
PK	Primary Key, eindeutige Identifikation eines Dateneintrages in der Datenbank
FK	Foreign Key, Schlüssel zu anderer Tabelle in der Datenbank
PLZ	Postleitzahl

## 1 Einleitung

Der hier vorliegende Bericht ist eine Ergänzung zu dem Bericht *Medizinische Versorgung in Pflegeheimen in Österreich* des Hauptverbandes der österreichischen Sozialversicherungsträger [1]. Im Weiteren wird dieser als *MedVers* bezeichnet. In den folgenden Kapiteln werden die Rohdaten, welche zur Beantwortung der Forschungsfragen (siehe Kapitel 5.2 aus *MedVers* [1]) verwendet wurden analysiert. Dazu wird in Kapitel 2 nochmals das Schema der verwendeten Rohdaten beschrieben. In Kapitel 3 werden die einzelnen Datensätze und deren Variablen analysiert.

## 2 Daten

Die Daten wurden einerseits aus der Pflegegeld-Datenbank für Personen 65 Jahre oder älter abgefragt und um weitere Angaben dieser Personen aus FOKO (Folgekosten der einzelnen Träger) ergänzt. Weiters wurde eine Liste der Pflegeheime selbst samt zusätzlicher Angaben erstellt. Die genaue Beschreibung der Datenherkunft ist in Kapitel 6 des Berichtes *MedVers* [1] zu finden.

Die extrahierten Rohdaten, die nun für die Auswertung der Forschungsfragen verwendet wurden, bestehen aus 4 Datensätzen (siehe Schema in **Abbildung 1**):

- Der Datensatz *Pflegeheime* ist eine Liste der österreichischen Pflegeheime (siehe Kapitel 4 aus *MedVers* [1]).
- Der Datensatz *Pfif\_Eintritte\_2014\_2016* (im Weiteren bezeichnet als *Pfif-Eintritte*) kommt aus der Pflegegeld-Datenbank und beschreibt die Eintritte und den Übergangsgrund in ein Pflegeheim der Personen. Es werden nur Eintritte in ein Pflegeheim zwischen 01.01.2014 und 31.12.2016 inkludiert, sofern die Person zum Stichtag 31.12.2016 65 Jahre oder älter war und sich in einem Pflegeheim befand.
- Die Datensätze *Vers\_daten* (Versichertendaten) und *Med\_versorgung* (Leistungsdaten) stammen aus FOKO und machen Angaben zu den zuvor identifizierten Patienten mit Pflegeheimaufenthalt und den erbrachten Leistungen. Hier sind alle Leistungen ab 2011 inkludiert. Diese werden bei der Beantwortung der Forschungsfragen auf den Zeitraum eingeschränkt in welchem die jeweilige Person im Pflegeheim war. Weiters wurden diese Datensätze jeweils um eine Spalte *Herkunft* erweitert, um ggf. den Träger bestimmen zu können.

### Erklärung Datenbankschema – Verknüpfung der Datensätze

In **Abbildung 1** sind jene Variablen **fett** gekennzeichnet und mit dem Zusatz (PK) versehen, welche zusammen jeweils den Schlüssel (Primary Key, eindeutige Identifikation) des Dateneintrages kennzeichnen: in den Versichertendaten (*Vers\_daten*) ist das beispielsweise *Vers\_key* (Dimensionsschlüssel der Versichertendimension) zusammen mit *Herkunft* (Versicherungsträger). Die Verknüpfung der Datensätze erfolgt über jene Variablen, welche als FK (Foreign Key, Fremdschlüssel) gekennzeichnet sind:

- Über die Personen ID (*Pseudonr\_patient*) sind die Einträge der Versichertendaten (*Vers\_daten*) und *Pfif-Eintritte* (*Pfif\_Eintritte\_2014\_2016*) verknüpft.
- Über den Dimensionsschlüssel der Versichertendimension (*Vers\_key*) und den Versicherungsträger (*Herkunft*), beides Primary Key der Versichertendaten, sind die Datensätze der Versicherten (*Vers\_daten*) und Leistungen (*Med\_versorgung*) verknüpft.

- Der Datensatz Pflegeheime ist mit keiner weiteren Tabelle direkt verknüpft. Die Daten sind durch den Namen, die Postleitzahl und die Anzahl der Pflegeplätze eindeutig definiert. Aufgrund der limitierten Möglichkeiten, können die Pflegeheime nicht direkt mit den Personen oder den Leistungen in Verbindung gebracht werden. Eine (nicht eindeutige) Kombination ist nur anhand der Postleitzahl möglich.

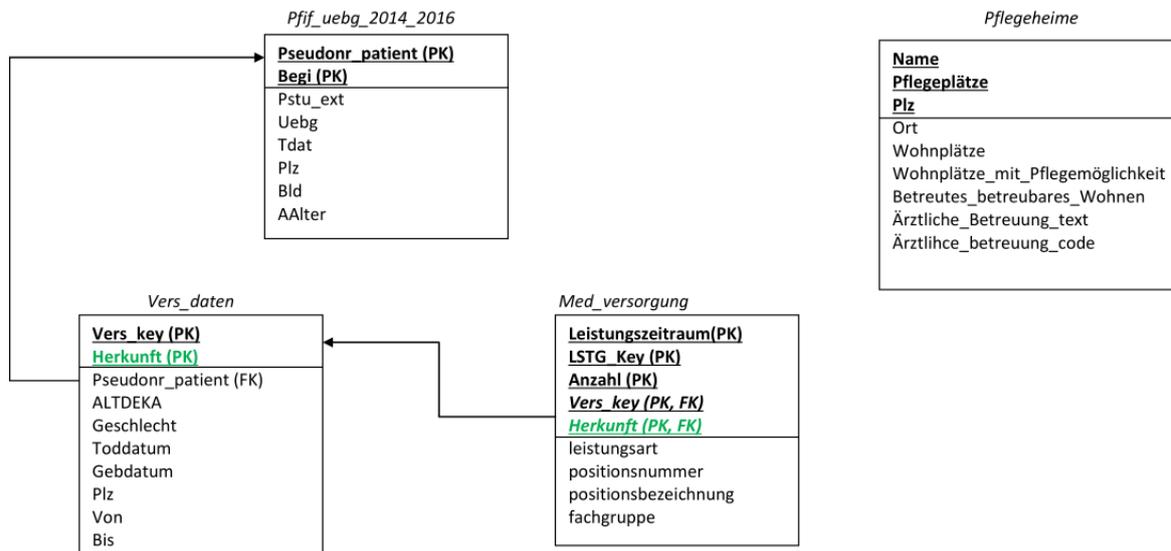


Abbildung 1: Rohdatenschema der vier Datensätze zu Pflegeheimen, Personen in Pflegeheimen sowie Leistungen und Versicherungsdaten dieser Personen.

### 3 Analyse der Rohdaten

Zur Analyse der Daten und um eine erste Übersicht über die Qualität der Daten zu erhalten, wurden sogenannte Profile der Rohdaten erstellt, die für jede Variable folgende Auswertungen beinhalten:

- Maximal- und Minimalwert
- Gesamtanzahl der Werte
- Anzahl der fehlenden Werte
- Anzahl der verschiedenen Werte.
- Histogramm über die Verteilung der Werte

Die Analysen sind im Folgenden nach den jeweiligen Tabellen sortiert aufgelistet.

#### 3.1 Der Datensatz *Pfif\_Eintritte* zu Eintritten und Übergangsgrund in ein Pflegeheim

Der Datensatz *Pfif\_uebg\_2014\_2016* aus der Pflegegeld-Datenbank (im Weiteren als *Pfif\_Eintritte* bezeichnet) beschreibt Daten zu den Personen, welche in ein Pflegeheim eingetreten sind, zusammen mit Datum des Eintrittes sowie Pflegestufe. Eine erste Übersicht der Beschreibung der Variablen ist in **Tabelle 1** zu finden, die genauen Profile sind pro Variable in den folgenden Unterkapiteln zu finden.

Tabelle 1: Übersicht der Variablen der Pflegegeld-Datenbank (*Pfif\_Eintritte*).

Datensatz: <i>Pfif_Eintritte</i>	Pflegeheimeintritte
Abfrage ausgewählter Variablen aus der Pflegegeld-Datenbank zu allen Personen, die zum Stichtag (31.12.2016) 65 Jahre oder älter waren und im Pflegeheim lebten.	
<i>Pseudonr_patient</i>	Pseudonymisierte Personen Identifier (Personen ID)
<i>Pstu_ext</i>	Pflegestufe der Person
<i>Uebg</i>	Übergangsgrund (ist 1, wenn die Person im Pflegeheim lebt)
<i>Tdat</i>	Todesdatum der Person
<i>Plz</i>	Postleitzahl der Person
<i>Bld</i>	Bundesland, in dem die Person lebt
<i>AAlder</i>	Alter der Person
<i>begi</i>	Beginn des Datensatzes

### 3.1.1 Variable *Pseudonr\_patient*: Patienten ID

In **Tabelle 2** ist das Datenprofil der Variable *Pseudonr\_patient* aus den Pflegegelddaten ersichtlich. Diese beschreibt die pseudonymisierte Personen ID eines Patienten. Die Werte befinden sich zwischen 1 und 85271. **Abbildung 2** zeigt ein Histogramm der Werte.

Tabelle 2: Datenprofil der Patienten ID.

Variable	<i>Pseudonr_patient</i>
Datensatz	<i>Pfif_Eintritte</i>
Minimaler Wert	1
Maximaler Wert	85271
Gesamtzahl Werte	189.914
Fehlende Werte	0
Verschiedene Werte	85.271

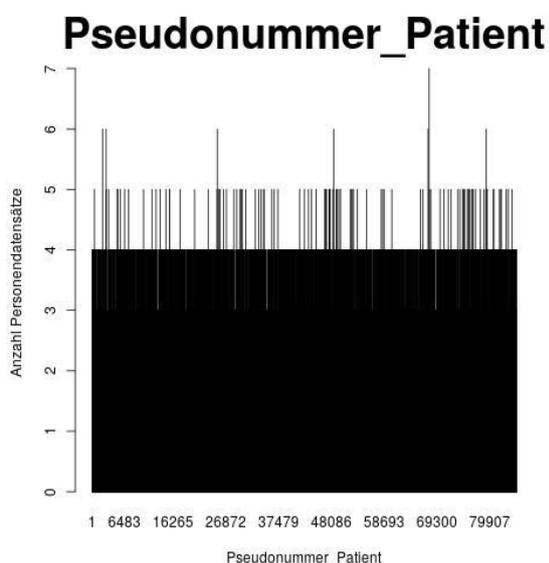


Abbildung 2: Verteilung der Variable *Pseudonr\_patient* (Patienten ID).

### 3.1.2 Variable *Pstu\_ext*: Pflegestufe der Person

In **Tabelle 3** ist das Datenprofil der Pflegestufe aus den Pflegegelddaten ersichtlich. Diese beschreibt die Pflegestufe eines Patienten. Die Werte befinden sich zwischen 1 und 7. **Abbildung 3** zeigt ein Histogramm der Werte. Die am häufigsten vorkommende Pflegestufe ist Stufe 5.

Tabelle 3: Datenprofil der Pflegestufe.

Variable	<i>Pstu_ext</i>
Datensatz	<i>Pfif_Eintritte</i>
Minimaler Wert	1
Maximaler Wert	7
Gesamtzahl Werte	189.914
Fehlende Werte	0
Verschiedene Werte	7

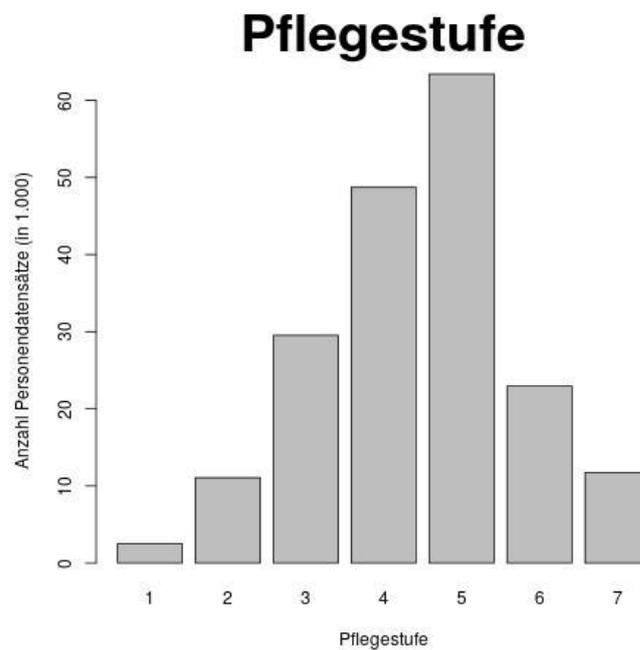


Abbildung 3: Verteilung der Variable *Pstu\_ext* (Pflegestufe).

### 3.1.3 Variable *Uebg*: Übergangsgrund in das Pflegeheim

In **Tabelle 4** ist das Datenprofil des *Übergangsgrundes* in ein Pflegeheim aus *Pfif\_Eintritte* ersichtlich. Diese ist genau dann 1, wenn die Person sich im Pflegeheim befindet. Da dies ein Kriterium für die Abfrage der Personen war, kommen auch keine anderen Werte als 1 vor. **Abbildung 4** zeigt ein Histogramm der Werte.

Tabelle 4: Datenprofil des *Übergangsgrundes*.

Variable	<i>Uebg</i>
Datensatz	<i>Pfif_Eintritte</i>
Minimaler Wert	1
Maximaler Wert	1
Gesamtzahl Werte	189.914
Fehlende Werte	0
Verschiedene Werte	1

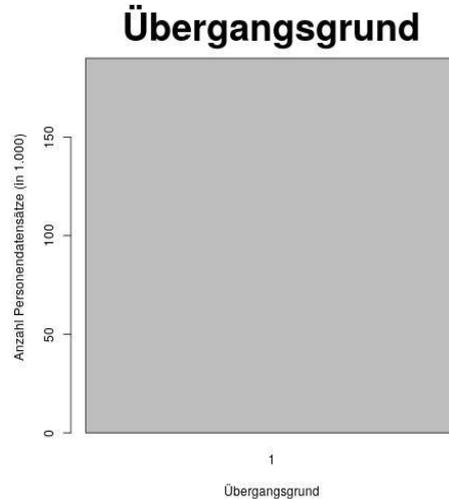


Abbildung 4: Verteilung der Variablen *Uebg* (*Übergangsgrund*).

### 3.1.4 Variable *Plz*: Postleitzahl der Person

In **Tabelle 5** ist das Datenprofil der Postleitzahl aus *Pfif\_Eintritte* ersichtlich. Diese beschreibt die eingetragene Postleitzahl einer Person. Die Variable enthält Werte zwischen 1010 und 9990. Die Zahl der fehlenden Werte liegt dabei weit unter 1%. In **Abbildung 5** ist die Verteilung der Werte zu finden. Die 5 häufigsten Werte sind, beginnend mit dem häufigsten:

- 4020 Linz (4.038 Einträge)
- 1230 Wien (2.913 Einträge)
- 6020 Innsbruck (2.850 Einträge)
- 9020 Klagenfurt (2.754 Einträge)
- 1220 Wien (2.693 Einträge)

Tabelle 5: Datenprofil der Postleitzahl.

Variable	<i>Plz</i>
Datensatz	<i>Pfif_Eintritte</i>
Minimaler Wert	1010
Maximaler Wert	9990
Gesamtzahl Werte	189.894
Fehlende Werte	20
Verschiedene Werte	1.278

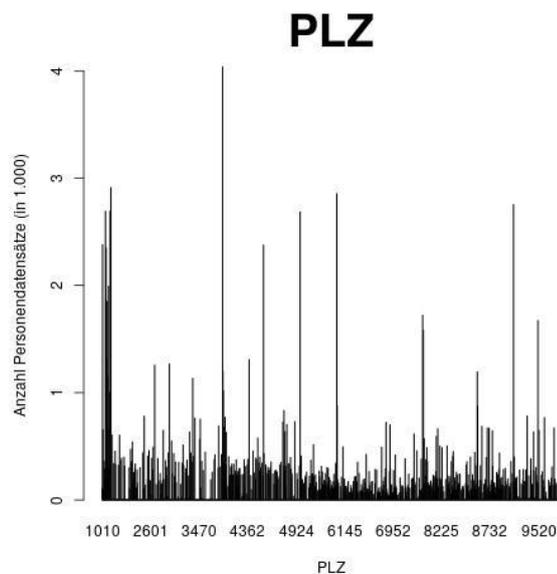


Abbildung 5: Verteilung der Variable *Plz* (*Postleitzahl*).

### 3.1.5 Variable *Bld*: Bundesland der Person

In **Tabelle 6** ist das Datenprofil der Variablen *Bld* zu den Bundesländern ersichtlich. Dieses beschreibt das Bundesland einer Person als Zahl zwischen 0 und 9. Wie die Zahlen von 1 bis 9 gedeutet werden können, ist in **Tabelle 7** zusammengefasst. Beim Wert 0 handelt es sich wohl um einen Fehler bei der Eingabe. **Abbildung 6** zeigt ein Histogramm der Werte.

Tabelle 6: Datenprofil des Bundeslandes.

Variable	<i>Bld</i>
Datensatz	<i>Pfif_Eintritte</i>
Minimaler Wert	0
Maximaler Wert	9
Gesamtzahl Werte	189.914
Fehlende Werte	0
Verschiedene Werte	10

Tabelle 7: Definition der Codes der Bundesländer.

Code	Bundesland
1	Wien
2	Niederösterreich
3	Burgenland
4	Oberösterreich
5	Steiermark
6	Kärnten
7	Salzburg
8	Tirol
9	Vorarlberg

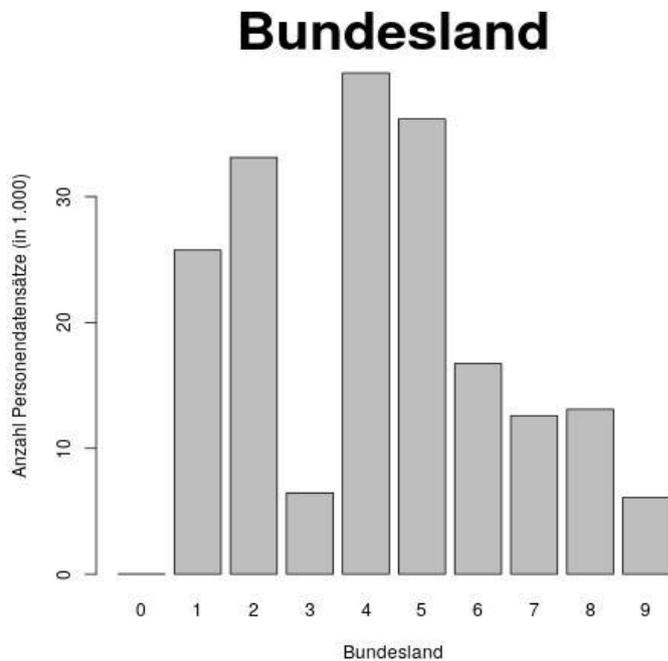


Abbildung 6: Verteilung der Variable *Bld* (Bundesland), Legende siehe **Tabelle 7**.

### 3.1.6 Variable *AAlter*: Alter einer Person

In **Tabelle 8** ist das Datenprofil der Variable zum Alter ersichtlich. Dieses beschreibt das Alter einer Person zum Zeitpunkt der Datensatzerstellung und enthält numerische Werte zwischen 65 und 111. Die meisten Patienten sind ca. zwischen 88 und 92 Jahren alt, wie in **Abbildung 7** ersichtlich.

Tabelle 8: Datenprofil der Variable zu Alter.

Variable	<i>AAlter</i>
Datensatz	<i>Pfif_Eintritte</i>
Minimaler Wert	65
Maximaler Wert	111
Gesamtzahl Werte	189.914
Fehlende Werte	0
Verschiedene Werte	47

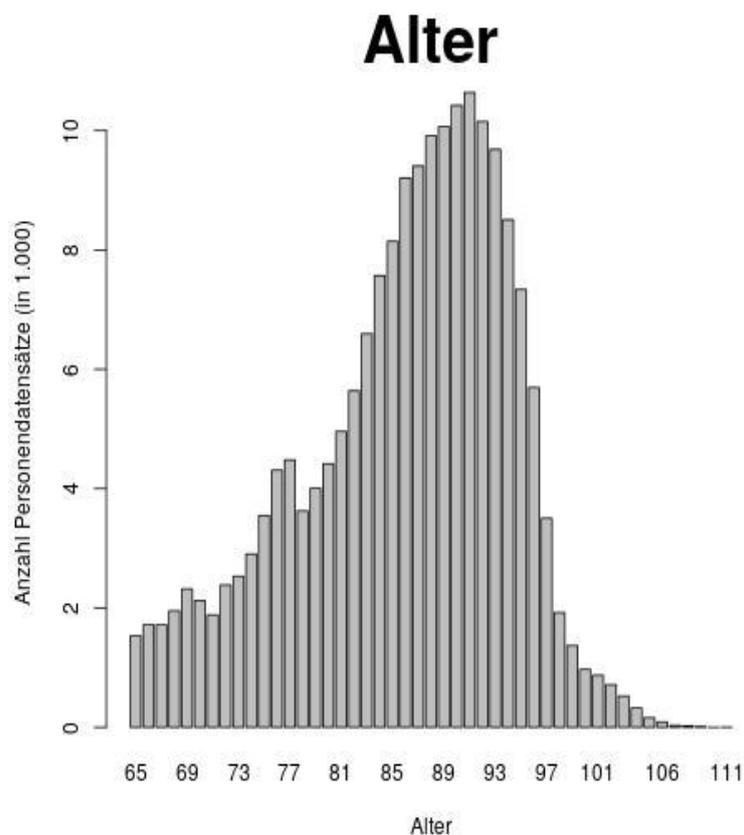


Abbildung 7: Verteilung der Variable *AAlter* (Alter).

### 3.1.7 Variable *Tdat*: Sterbedatum einer Person

In **Tabelle 9** ist das Datenprofil der Variable *Tdat* aus *Pfif\_Eintritte* ersichtlich. Dieses beschreibt das Sterbedatum einer Person, falls es vorhanden ist. Die Variable enthält Werte zwischen 2013-12-19 und 2017-04-08. Da lebende Personen kein Todesdatum besitzen, liegt die Zahl der fehlenden Werte bei 50 % (94.893). Es gibt im Datensatz einen Wert, der vor 1.1.2014 liegt. Dieser ist enthalten, da es Einträge in der Datenbank mit einem Beginndatum nach 1.1.2014 gibt. Auf die Ergebnisse hat diese Person jedoch keinen Einfluss, da für sie keine Leistungen existieren, die nach 1.1.2014 abgerechnet wurden. **Abbildung 8** zeigt ein Histogramm der Werte.

Tabelle 9: Datenprofil der Variable zum Sterbedatum.

Variable	<i>Tdat</i>
Datensatz	<i>Pfif_Eintritte</i>
Minimaler Wert	2013-12-19
Maximaler Wert	2017-04-08
Gesamtzahl Werte	95.021
Fehlende Werte	94.893
Verschiedene Werte	1.195

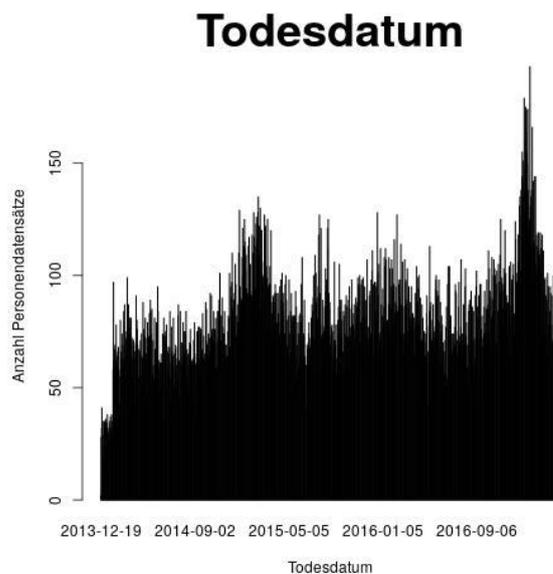


Abbildung 8: Verteilung der Variable *Tdat* (Todesdatum)

### 3.1.8 Variable *Begi*: Beginndatum

In **Tabelle 10** ist das Datenprofil der Variable *Begi* ersichtlich. Diese beschreibt das Beginndatum des jeweiligen Datensatzes und kennzeichnet eine mögliche Änderung der Pflegestufe bzw. den Eintritt in das Pflegeheim. Die Variable enthält Werte zwischen 2014-01-01 und 2016-12-01. **Abbildung 9** zeigt ein Histogramm der Werte. Bei den auffälligen Häufungen am Beginn und am Ende der Skala handelt es sich um (häufig vorkommende) Artefakten am Rand der Zeitreihe. Also alle Personen, die sich am Anfang der Beobachtung bereits in einem Pflegeheim befinden, erhalten als Beginndatum den Beginn der Beobachtung (den ersten Zeitpunkt der Zeitreihe). Alle Personen, die nach Ablauf des

Beobachtungszeitraumes ins Pflegeheim kamen, bekamen das letzte Datum der Zeitreihe als Beginndatum.

Tabelle 10: Datenprofil der Variable zu Beginndatum.

Variable	<i>Begi</i>
Datensatz	<i>Pfif_Eintritte</i>
Minimaler Wert	2014-01-01
Maximaler Wert	2016-12-01
Gesamtzahl Werte	189.914
Fehlende Werte	0
Verschiedene Werte	36

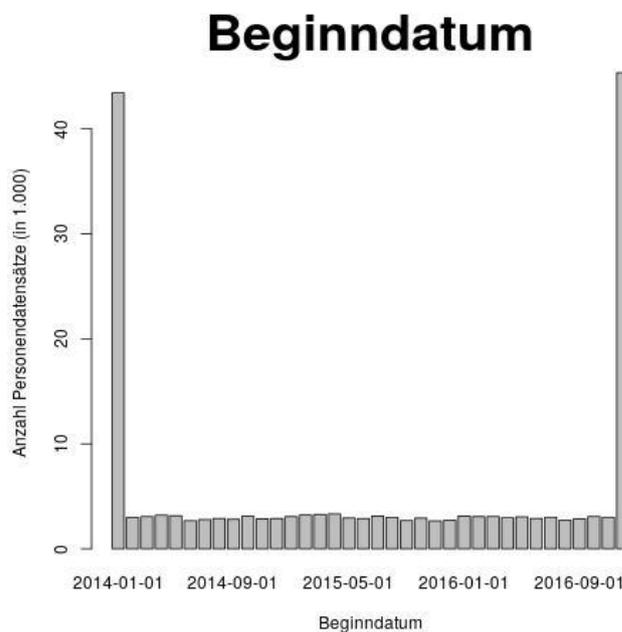


Abbildung 9: Verteilung der Variable *Begi* (Beginndatum)

Die meisten Personen haben mindestens 2 Dateneinträge und somit 2 Beginndaten verzeichnet: 4.502 Personen haben nur einen Eintrag, 118.286 haben 2 Einträge und 67.126 Personen haben 3 bis 7 Einträge.

### 3.2 Die Datensätze *Vers\_daten* zu den *Versichertendaten*

Die Datensätze zu den Versicherten (*Vers\_daten*) sind pro Träger vorhanden (eigenes File) und werden unter Einführung einer zusätzlichen Variablen *Herkunft* in einem Datensatz zusammengefasst. Mittels *Herkunft* ist der Träger weiterhin ersichtlich. Die folgenden Analysen beziehen sich auf alle Träger zusammengefasst und enthalten alle Versicherungsdatensätze, die für das Projekt herangezogen wurden. Eine erste Übersicht der Beschreibung der einzelnen Variablen ist in **Tabelle 11** zu finden. Die genauen Profile sind pro Variable in den folgenden Unterkapiteln ersichtlich.

Tabelle 11: Übersicht der Variablen zu den Versichertendaten (*Vers\_daten*).

Datensatz: <i>Vers_daten</i>	Versichertendaten
Abfrage ausgewählter Variablen aus FOKO zu jenen versicherten Personen, die in <i>Pfifj-Eintritte</i> identifiziert wurden. Es gibt keine weiteren Einschränkungen.	
<i>Vers_key</i>	Dimensionsschlüssel der Versicherten-dimension, welcher zusammen mit <i>Herkunft</i> den Dateneintrag eindeutig identifiziert.
<i>Herkunft</i>	Beschreibt den Träger und identifiziert zusammen mit <i>Vers_key</i> den Dateneintrag eindeutig.
<i>Pseudonr_patient</i>	Pseudonymisierte Personen Identifier (Personen ID)
<i>ALTDEKA</i>	Zugehörige Altersdekade der Person
<i>Geschlecht</i>	Geschlecht der Person
<i>Toddatum</i>	Todesdatum der Person
<i>Gebdatum</i>	Geburtsdatum der Person
<i>PLZ</i>	Postleitzahl der Person
<i>Von</i>	Beginn des Versicherungsdatensatzes
<i>Bis</i>	Ende des Versicherungsdatensatzes

### 3.2.1 Variable *ALTDEKA* : Altersdekade der Person

In **Tabelle 12** ist das Datenprofil der Variable *ALTDEKA* aus den Versichertendaten ersichtlich. Diese beschreibt die Altersdekade einer Person in 10er Schritten, welche mit Werten von 0 bis 9 gekennzeichnet werden. Altersdekade 0 entspricht dem Alter 0 bis 9, etc. Wie auch in **Abbildung 10**, einem Histogramm über die verschiedenen Einträge, ersichtlich ist, gibt es nur die Klassen 0 und 4 bis 9. Besonders stark vertreten ist hier der Wert 8 mit 39 % aller Einträge, das sind 457.459 Datensätze.

Tabelle 12: Datenprofil der Variable zu den Altersdekaden der Personen in den Versichertendaten.

Variable	<i>ALTDEKA</i>
Datensatz	<i>Vers_daten</i>
Minimaler Wert	0
Maximaler Wert	9
Gesamtzahl Werte	1.176.587
Fehlende Werte	0
Verschiedene Werte	7

## Altersdekaden

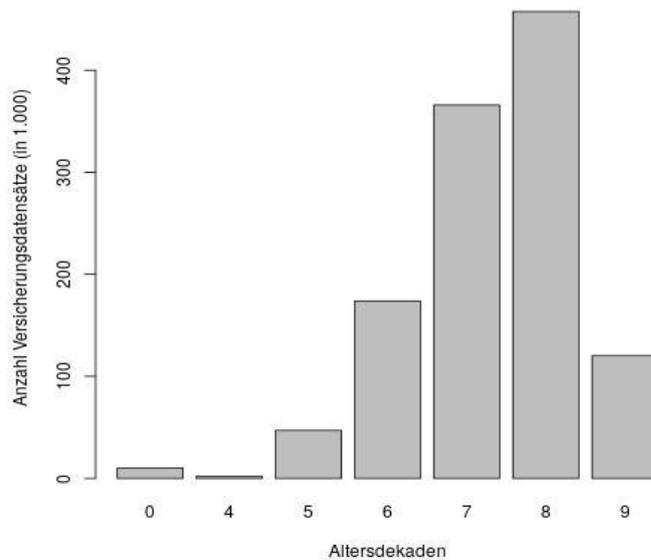


Abbildung 10: Verteilung der Variable *ALTDEKA* (Altersdekaden in 10er Schritten).

Obwohl in den Analysen nur Personen mit 65 Jahren oder älter untersucht werden, scheinen hier jüngere Dekaden ebenfalls auf, da die Einschränkung auf das Alter in *Pfif\_Eintritte* gemacht wurde und danach basierend auf der Personen ID weitere Informationen aus FOKO zu den Versichertendaten abgefragt wurden. Hier wurde keine zusätzliche Einschränkung gemacht.

### 3.2.2 Variable *Geschlecht*: *Geschlecht*

In **Tabelle 13** ist das Datenprofil der Variable *Geschlecht* aus den Versichertendaten (*vers\_daten*) ersichtlich. Diese beschreibt das Geschlecht einer Person und hat die Werte *M* (männlich) und *W* (weiblich). Es zeigt sich, dass es ca. drei Mal mehr Frauen (860.967) gibt als Männer (284.434) und dass 3% aller Datensätze (31.185) keinen Wert eingetragen haben. **Abbildung 11** zeigt die Verteilung.

Tabelle 13: Datenprofil der Variable *Geschlecht*

Variable	<i>Geschlecht</i>
Datensatz	<i>Vers_daten</i>
Gesamtzahl Werte	1.145.402
Fehlende Werte	31.185
Verschiedene Werte	2

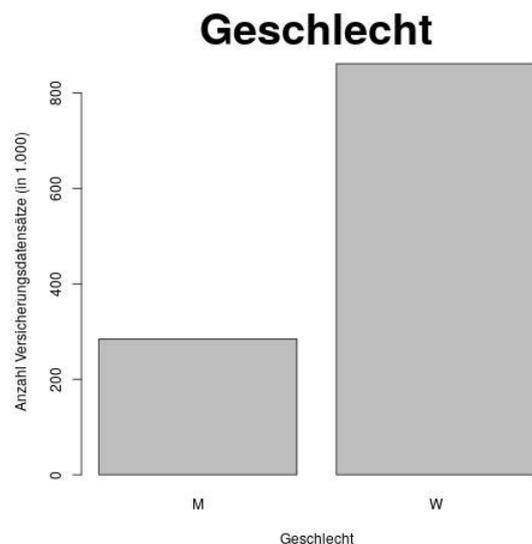


Abbildung 11: Verteilung der Variable *Geschlecht*.

### 3.2.3 Variable *Toddatum*: Sterbedatum der Person

In **Tabelle 14** ist das Datenprofil der Variable *Toddatum*, also des Sterbedatums aus den Versichertendaten (*Vers\_daten*) ersichtlich. Da Personen, die noch leben, kein Todesdatum haben können, ist die Zahl der fehlenden Werte mit 94% (1.109.759 Datensätze) sehr hoch. Die häufigsten Todesdaten sind 2015-01-19, 2015-01-11 und 2015-02-02. In **Abbildung 12** ist die Verteilung der Werte ersichtlich.

Tabelle 14: Datenprofil der Variable zum Sterbedatum in den Versichertendaten.

Variable	<i>Toddatum</i>
Datensatz	<i>Vers_daten</i>
Minimaler Wert	2008-06-12
Maximaler Wert	2017-08-16
Gesamtzahl Werte	66.827
Verschiedene Werte	1.352
Fehlende Werte	1.109.760

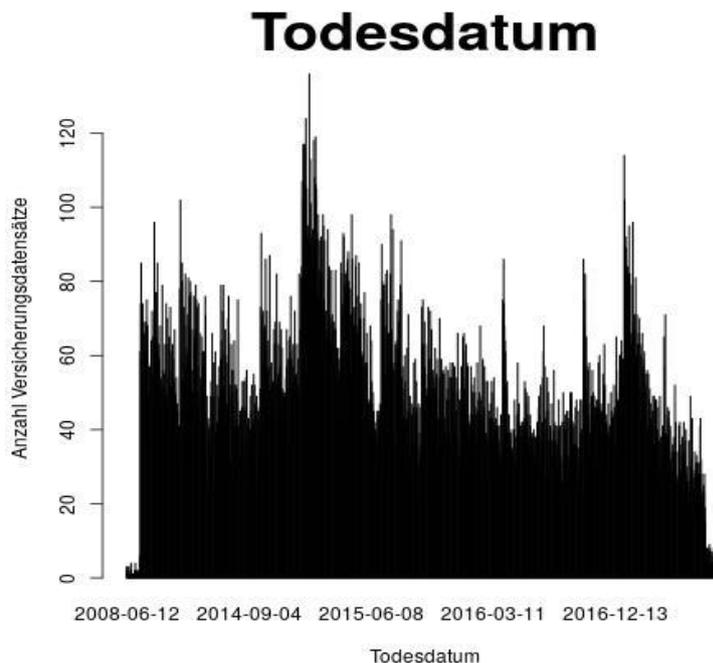


Abbildung 12: Verteilung der Variable *Toddatum* (Todesdatum).

### 3.2.4 Variable *Gebdatum*: Geburtsdatum der Person

In **Tabelle 15** ist das Datenprofil der Variable *Gebdatum* zum Geburtsdatum einer Person aus den Versichertendaten (*Vers\_daten*) ersichtlich. Die Werte befinden sich zwischen 1905-03-18 und 31/12/1951. Es zeigt sich, dass 1 % der Datensätze (9.535) in der Datenbank kein Geburtsdatum aufweisen. **Abbildung 13** zeigt die Verteilung der Werte, wobei hier zu erkennen ist, dass die meisten Personen zum Stichtag 31.12.2016 zwischen 83 und 96 Jahren alt sind.

Tabelle 15: Datenprofil der Variable zum Geburtsdatum in den Versichertendaten.

Variable	<i>Gebdatum</i>
Datensatz	<i>Vers_daten</i>
Minimaler Wert	1905-03-18
Maximaler Wert	1951-12-31
Gesamtzahl Werte	1.167.051
Fehlende Werte	9.536
Verschiedene Werte	13.739

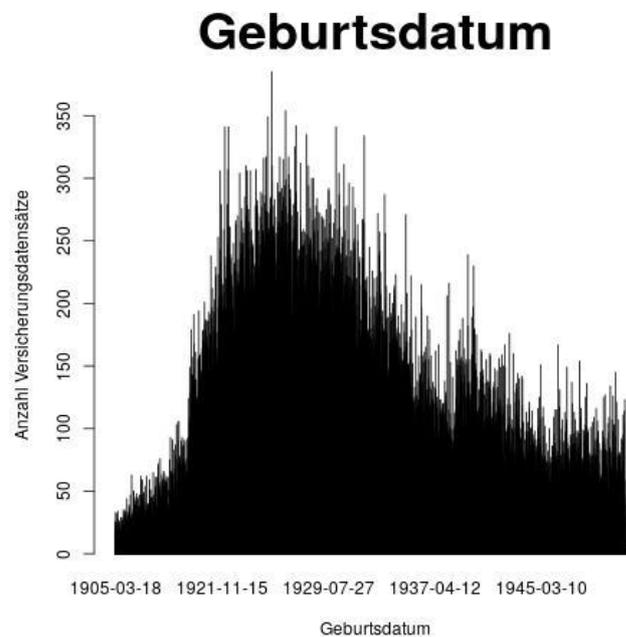


Abbildung 13: Verteilung der Variable *Gebdatum* (Geburtsdatum)

### 3.2.5 Variable *PLZ* : Postleitzahl der Person

In **Tabelle 16** ist das Datenprofil der Variable *PLZ*, also der Postleitzahl aus den Versichertendaten (*Vers\_daten*) ersichtlich. Diese beschreibt die beim jeweiligen Versicherungsträger eingetragene Postleitzahl einer Person. Je nach Zeitraum und Träger kann die Postleitzahl bezüglich einer Person variieren oder sich auch ändern. Die Werte sind einerseits numerische Daten zwischen 1010 und 9971, sowie der Wert 0, aber auch alphanumerische Daten. Die 5 häufigsten Postleitzahlen sind:

- 4020 Linz (20.402 Einträge)
- 1051 Wien, Margareten (19.584 Einträge)
- 1100 Wien, Favoriten (18.177 Einträge)
- 5020 Salzburg (17.307 Einträge)
- 9020 Klagenfurt (16.024 Einträge)

Es zeigt sich, dass 4 % der Datensätze (47.347) in der Datenbank keine Postleitzahl aufweisen. **Abbildung 14** zeigt ein Histogramm der Werte.

Tabelle 16: Datenprofil der Variable zur Postleitzahl in den Versichertendaten.

Variable	PLZ
Datensatz	Vers_daten
Minimaler Wert	0
Maximaler Wert	9971
Gesamtzahl Werte	1.129.240
Fehlende Werte	47.347
Verschiedene Werte	2.455

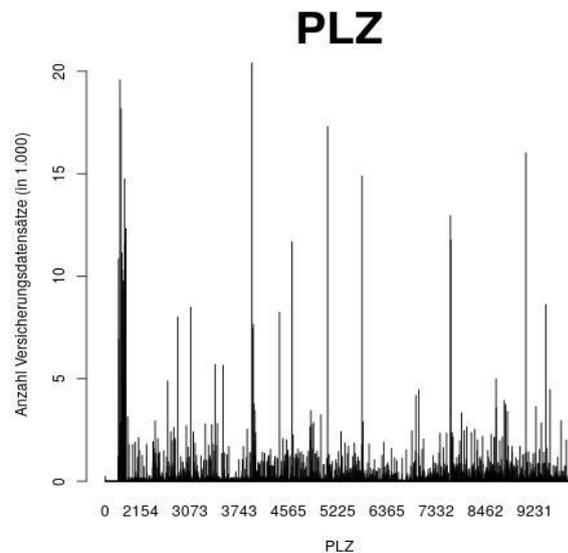


Abbildung 14: Verteilung der Variable PLZ (Postleitzahl)

### 3.2.6 Variable Von: Beginndatum des Dateneintrages der Versicherung

In **Tabelle 17** ist das Datenprofil der Variable *von* aus den Versichertendaten (*Vers\_daten*) ersichtlich. Diese beschreibt den Beginn des aktuellen Versicherungsdatensatzes. In dieser Variable befinden sich Datumswerte zwischen 1998-01-01 und 2017-08-01. Die 5 häufigsten Daten sind, beginnend mit dem Häufigsten sind:

- 2013-01-01 (49.943 Einträge)
- 2009-01-07 (48.598 Einträge)
- 2008-01-01 (37.897 Einträge)
- 2010-01-01 (34.392 Einträge)
- 2014-01-01 (32.332 Einträge)

Es ist ersichtlich, dass dies Einträge zu Jahresbeginn, bzw. Quartalsbeginn sind. **Abbildung 15** zeigt ein Histogramm der Werte.

Tabelle 17: Datenprofil der Variable *Von* zum Beginn der Versicherung in den Versichertendaten.

Variable	von
Datensatz	Vers_daten
Minimaler Wert	1998-01-01
Maximaler Wert	2017-08-01
Gesamtzahl Werte	1.176.587
Fehlende Werte	0
Verschiedene Werte	220

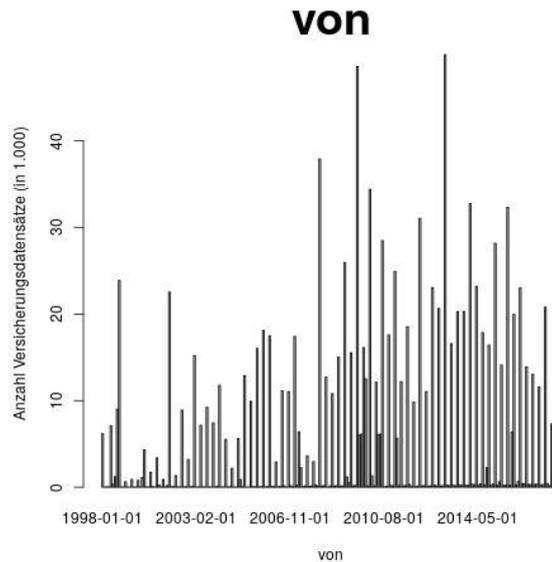


Abbildung 15: Verteilung der Variable *Von* (Beginn der Versicherung).

### 3.2.7 Variable *Bis* : Datum zum Versicherungsende

In **Tabelle 18** ist das Datenprofil der Variable *bis* aus den Versichertendaten (*Vers\_daten*) ersichtlich. Diese beschreibt das Ende des aktuellen Versicherungsdatsatzes. In dieser Variable befinden sich Datumswerte zwischen 1999-03-31 und 9998-12-31. Der maximale Wert wird dann verwendet, wenn das Ende des Versicherungsdatsatzes (noch) nicht bekannt ist. Die 5 häufigsten Daten sind:

- 9998-12-31 (115.814 Einträge)
- 2012-12-31 ( 49.782 Einträge)
- 2009-06-30 ( 48.531 Einträge)
- 2009-12-31 ( 34.237 Einträge)
- 2013-12-31 ( 32.635 Einträge)

Der Wert 9998-12-31 wird dabei in 10 % der Datensätze (115.814) angeführt. **Abbildung 16** zeigt ein Histogramm der Werte.

Tabelle 18: Datenprofil der Variable *Bis* zum Ende der Versicherung in den Versichertendaten.

Variable	<i>Bis</i>
Datensatz	<i>Vers_daten</i>
Minimaler Wert	1999-03-31
Maximaler Wert	9998-12-31
Gesamtzahl Werte	1.176.587
Fehlende Werte	0
Verschiedene Werte	216

## bis

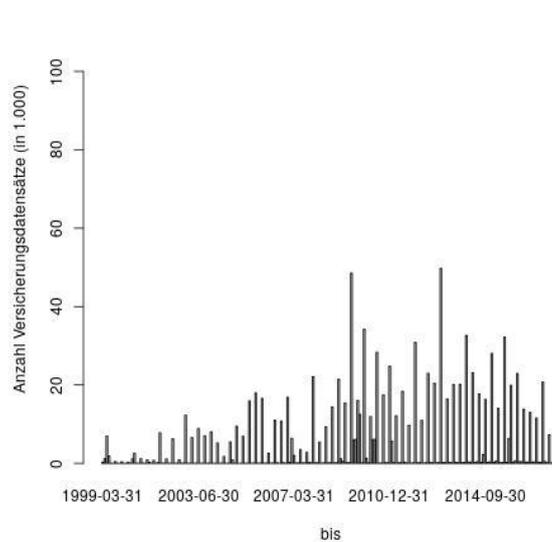


Abbildung 16: Verteilung der Variable *Bis* (Ende der Versicherung einer Person).

### 3.2.8 Variable *Vers\_key*: Dimensionsschlüssel der Versichertendimension

In **Tabelle 19** ist das Datenprofil der Variable *Vers\_key* aus den Versichertendaten ersichtlich. Diese beschreibt den Dimensionsschlüssel der Versichertendimension des aktuellen Versicherungsdatensatzes. Sie ist, gemeinsam mit dem Versicherungsträger (*Herkunft*), eindeutig. In dieser Variablen befinden sich Werte zwischen 26 und 53442398. **Abbildung 17** zeigt ein Histogramm der Werte.

Tabelle 19: Datenprofil des Dimensionsschlüssels.

Variable	<i>Vers_key</i>
Datensatz	<i>Vers_daten</i>
Minimaler Wert	26
Maximaler Wert	53442398
Gesamtzahl Werte	1.176.587
Fehlende Werte	0
Verschiedene Werte	1.157.748

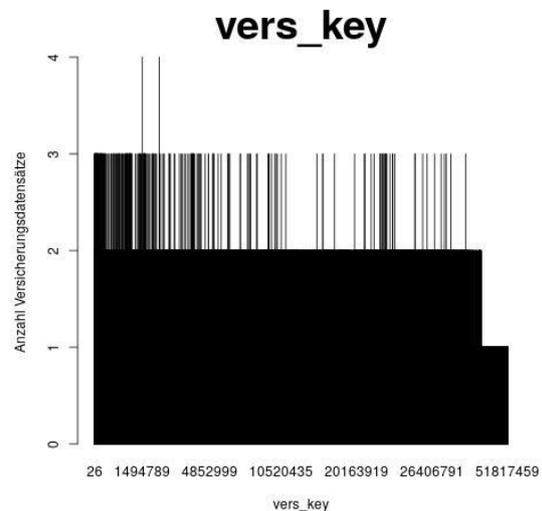


Abbildung 17: Verteilung der Variablen *Vers\_key* (Dimensionsschlüssel der Versichertendimension).

### 3.2.9 Variable *Pseudonr\_patient*: Patienten ID

In **Tabelle 20** ist das Datenprofil der Variablen *Pseudonr\_patient* aus den Versichertendaten (*Vers\_daten*) ersichtlich. Diese beschreibt die pseudonymisierte Identifikationsnummer eines Patienten. Die Werte sind zwischen 1 und 85271. **Abbildung 18** zeigt ein Histogramm der Werte.

Tabelle 20: Datenprofil der Variable der Patienten ID.

Variable	<i>Pseudonr_patient</i>
Datensatz	<i>Vers_daten</i>
Minimaler Wert	1
Maximaler Wert	85271
Gesamtzahl Werte	1.176.587
Fehlende Werte	0
Verschiedene Werte	83.989

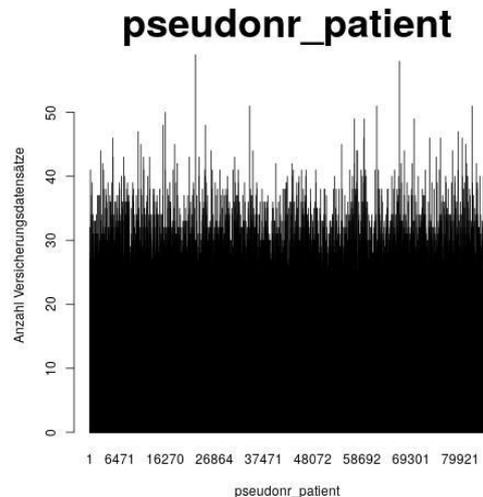


Abbildung 18: Verteilung der Variable *Pseudonr\_patienten* (Patienten ID).

### 3.3 Die Datensätze *Med\_versorgung* zu den Leistungsdaten

Die Datensätze zu den medizinischen Leistungen (*Med\_versorgung*) sind pro Träger vorhanden und werden unter Einführung einer zusätzlichen Variablen *Herkunft* zusammengefasst. Mittels *Herkunft* kann der Träger weiterhin ermittelt werden. Die folgenden Analysen beziehen sich auf alle Träger zusammengefasst und enthalten alle Leistungsdatensätze, die für das Projekt herangezogen wurden, also für jene Personen, die im Datensatz der *Pfif-Eintritte* ermittelt wurden. Eine erste Übersicht der Beschreibung der Variablen ist in **Tabelle 21** zu finden und die genauen Profile pro Variable in den folgenden Unterkapiteln.

Tabelle 21: Übersicht der Variablen zu den Leistungsdaten (*Med\_versorgung*)

Datensatz : <i>Med_versorgung</i>	Leistungsdaten
Abfrage ausgewählter Variablen aus FOKO zu jenen versicherten Personen, die in <i>Pfif-Eintritte</i> identifiziert wurden. Keine weiteren Einschränkungen.	
<i>Lstg_key</i>	Dimensionsschlüssel der Leistungsdimension
<i>Vers_key</i>	Dimensionsschlüssel der Versichertendimension
<i>Herkunft</i>	Träger
<i>Leistungsart</i>	Art der Leistung
<i>Positionsnummer</i>	Nummer der Position
<i>Positionsbezeichnung</i>	Bezeichnung der Position
<i>Fachgruppe</i>	Fachgruppe in der die Leistung abgerechnet wurde
<i>Leistungszeitraum</i>	Datum der Abrechnung der Leistung
<i>Anzahl</i>	Anzahl der abgerechneten Leistungen

### 3.3.1 Variable *Leistungsart*: Leistungsart

In **Tabelle 22** ist das Datenprofil der Variable *Leistungsart* der Leistungsdaten (*Med\_versorgung*) ersichtlich. Diese beschreibt die Art der Leistung als Kurzform. Die einzelnen Leistungsarten sind in **Tabelle 23** ersichtlich. Die häufigste Leistungsart ist mit 55% Eigenkosten (EK) (20.738.124 Datensätze) und die zweithäufigste Leistungsart ist Heilmittel (HM) mit 37 % aller Datensätze (13.905.189). **Abbildung 19** zeigt die Verteilung der Werte.

Tabelle 22: Datenprofil der Variable *Leistungsart*

Variable	<i>Leistungsart</i>
Datensatz	<i>Med_versorgung</i>
Gesamtzahl Werte	37.933.066
Fehlende Werte	0
Verschiedene Werte	9

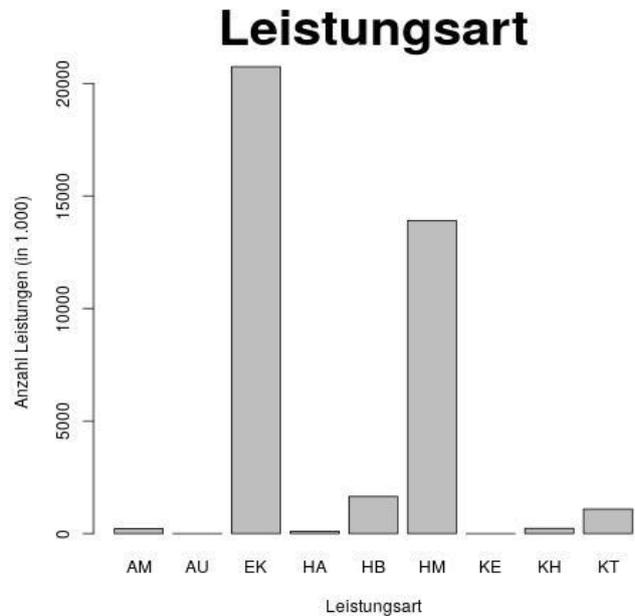


Abbildung 19: Verteilung der Variable *Leistungsart*

Tabelle 23: Definition der Leistungsarten

Leistungsart	Bezeichnung
AM	Krankenhaus ambulant
AU	Arbeitsunfähigkeit
EK	Eigenkosten
HA	Hauseigene Ambulatorien
HB	Heilbehelfe/Hilfsmittel
HM	Heilmittel
KE	Kur und Erholung
KH	Krankenhaus stationär
KT	Krankenhaustransporte

### 3.3.2 Variable *Positionsnummer*: Identifikationsnummer der Leistung

In **Tabelle 24** ist das Datenprofil der Variable *Positionsnummer* der Leistungsdaten ersichtlich. Dies ist eine Identifikationsnummer der erbrachten Leistung. Es gibt hierbei sowohl numerische, als auch textuelle Bezeichnungen. Die Zahl der fehlenden Werte liegt dabei weit unter einem Prozent (0,003%). Die Positionsnummer ist nur zusammen mit der Herkunft eindeutig. Unten genannte Häufigkeiten beziehen sich nur auf die Variable *Positionsnummer*, ohne Berücksichtigung der *Herkunft*. Die fünf häufigsten Werte sind, beginnend mit dem Häufigsten:

- ZAM/INDI/15: (1.179.103 Einträge)
  - Herkunft 15: Ordination
  - Herkunft 19: Ausfertigung eines Rezeptes (Verrechnung einer Ordination nicht zulässig)
- ZLANDHEI/INDI/4911: Allgemeine Pflege Heilmittel (883.832 Einträge)
- ZLANDHEI/INDI/4910: Allgemeine Pflege ärztliche Hilfe (883.831 Einträge)
- ZAM/INDI/3E: BESUCH B. ALTEN-ODER PFLEGEHEIMPATIENTEN (739.569 Einträge)
- ZGRUND/INDI/12: Ordination (620.807 Einträge)

Der häufigste Wert „ZAM/INDI/15“ ist dabei in 3% aller Datensätze (1.179.103) zu finden. **Abbildung 20** zeigt ein Histogramm der Werte.

Tabelle 24: Datenprofil der Variable Positionsnummer

Variable	Positionsnummer
Datensatz	Med_versorgung
Gesamtzahl Werte	37.931.877
Fehlende Werte	1.189
Verschiedene Werte	61.137

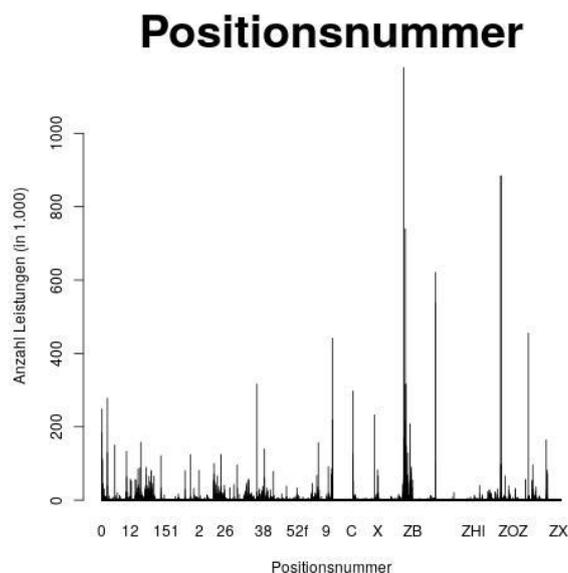


Abbildung 20: Verteilung der Variable *Positionsnummer*

### 3.3.3 Variable *Positionsbezeichnung*

In **Tabelle 25** ist das Datenprofil der Variable *Positionsbezeichnung* aus den Leistungsdaten ersichtlich. Diese beschreibt eine textuelle Beschreibung der Leistung. Die Zahl der fehlenden Werte liegt dabei weit unter einem Prozent (0,00002%). Die 5 häufigsten Werte sind, beginnend mit dem häufigsten:

- Ordination (2.435.359 Einträge)
- Allgemeine Pflege ärztliche Hilfe ( 883.832 Einträge)
- Allgemeine Pflege Heilmittel ( 883.831 Einträge)
- BESUCH B. ALTEN- ODER PFLEGEHEIMPATIENTEN ( 811.317 Einträge)
- Ordination eingeschränkt ( 588.268 Einträge)

Der häufigste Wert „Ordination“ ist dabei in 6% aller Datensätze enthalten. Auch hier ist die Bezeichnung der Position trägerabhängig und nur pro Träger eindeutig. **Abbildung 21** zeigt ein Histogramm der Werte.

Tabelle 25: Datenprofil der Variable *Positionsbezeichnung*

Variable	<i>Positionsbezeichnung</i>
Datensatz	<i>Med_versorgung</i>
Gesamtzahl Werte	37.932.133
Fehlende Werte	933
Verschiedene Werte	42.330

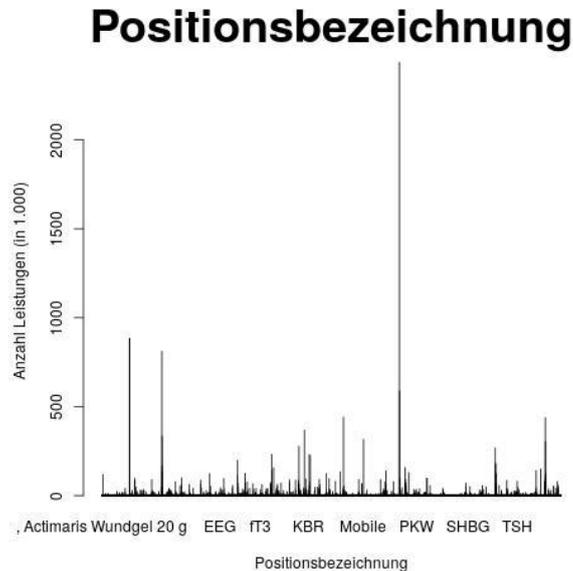


Abbildung 21: Verteilung der Variable *Positionsbezeichnung*

Die Positionsnummer ist nur zusammen mit der Herkunft eindeutig. Positionsbezeichnungen können trägerübergreifend verschieden sein. Die obigen Auswertungen beziehen sich auf jeweils die einzelne Variable, wodurch *Herkunft* nicht berücksichtigt wurde.

### 3.3.4 Variable *Fachgruppe*

In **Tabelle 26** ist das Datenprofil der Variable *Fachgruppe* aus den Leistungsdaten (*Med\_versorgung*) ersichtlich. Diese beschreibt eine Kurzform der Fachgruppe von der die Leistung abgerechnet wurde. In der Variable befinden sich sowohl numerische Werte von 1 bis 99, als auch textuelle Kurzformen (nur bei den Trägern 5, 7, 14, 15, 16, 17, 18). Die Zahl der fehlenden Werte liegt bei 45 % der Datensätze. Die 5 häufigsten Werte sind, beginnend mit dem häufigsten:

- Arzt für Allgemeinmedizin (Fachgruppe 1 mit 12.209.073 Einträgen)
- Pflegeheim für chronisch Kranke (Fachgruppe 87 mit 2.055.143 Einträgen)
- Facharzt für medizinische und chemische Labordiagnosen (Fachgruppe 50 mit 1.747.771 Einträgen)
- Andere Vertragspartner (Fachgruppe 99 mit 855.299 Einträgen)
- selbständiges Ambulatorium, ausgenommen für ZMK (Fachgruppe 85 mit 826.057 Einträgen)

Der häufigste Wert (Fachgruppe 1) ist dabei in 32% aller Fälle (12.209.073) enthalten. **Abbildung 22** zeigt ein Histogramm der Werte. In **Tabelle 44** im Anhang sind die genauen Fachgruppencodes erklärt.

Tabelle 26: Datenprofil der Variable *Fachgruppe*.

Variable	<i>Fachgruppe</i>
Datensatz	<i>Med_versorgung</i>
Gesamtzahl Werte	20.826.754
Fehlende Werte	17.106.312
Verschiedene Werte	81

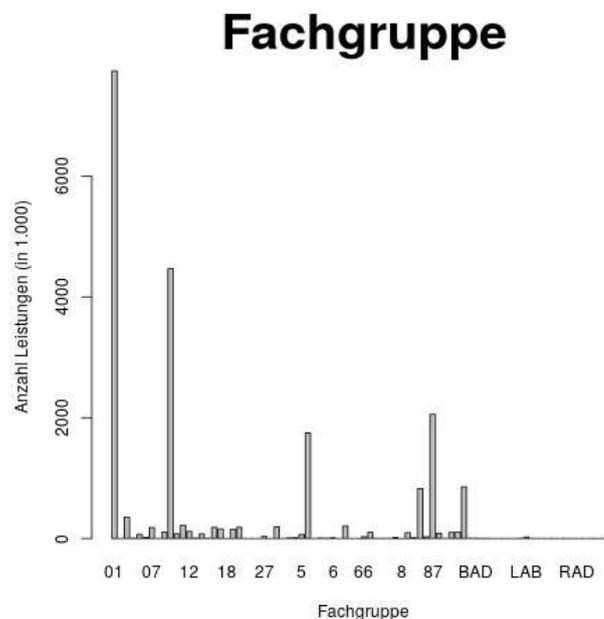


Abbildung 22: Verteilung der Variable *Fachgruppe*

### 3.3.5 Variable *Leistungszeitraum*

In **Tabelle 27** ist das Datenprofil der Variable *Leistungszeitraum* aus den Leistungsdaten (*Med\_versorgung*) ersichtlich. Diese beschreibt das Datum der Leistung. In dieser Variable befinden sich Datumswerte zwischen 01/01/2011 und 01/10/2017. Die 5 häufigsten Werte sind, beginnend mit dem häufigsten:

- 01/04/2014 (116.425 Einträge)
- 01/07/2014 (106.340 Einträge)
- 01/10/2014 (106.158 Einträge)
- 07/01/2014 ( 91.713 Einträge)
- 01/04/2015 ( 84.439 Einträge)

**Abbildung 23** zeigt ein Histogramm der Werte. Die häufigsten Werte bezeichnen den Beginn eines Quartals bzw. auch den Beginn eines Monats.

Tabelle 27: Datenprofil der Variable *Leistungszeitraum*.

Variable	<i>Leistungszeitraum</i>
Datensatz	<i>Med_versorgung</i>
Minimaler Wert	01/01/2011
Maximaler Wert	01/10/2017
Gesamtzahl Werte	37.933.066
Fehlende Werte	0
Verschiedene Werte	2.051

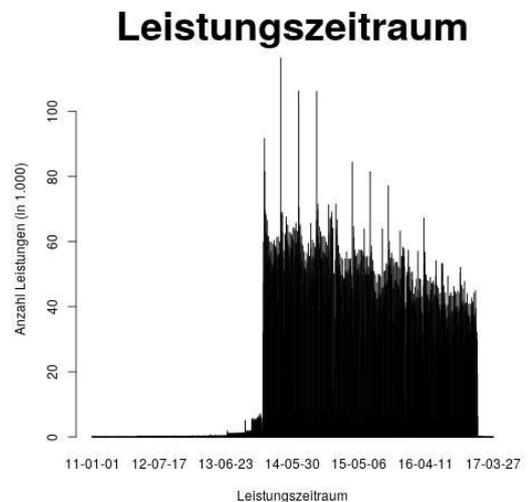


Abbildung 23: Verteilung der Variable *Leistungszeitraum*.

### 3.3.6 Variable *Vers\_key*: Dimensionsschlüssel der Versichertendimension

In **Tabelle 28** ist das Datenprofil des Dimensionsschlüssels zur Versichertendimension (*Vers\_key*) aus den Leistungsdaten (*Med\_versorgung*) ersichtlich. Hier sind numerische Werte zwischen 307361 und 51118597 eingetragen. **Abbildung 24** zeigt ein Histogramm der Werte.

Tabelle 28: Datenprofil des Schlüssels des Versichertendateneintrages.

Variable	<i>Vers_key</i>
Datensatz	<i>Med_versorgung</i>
Minimaler Wert	307361
Maximaler Wert	51118597
Gesamtzahl Werte	37.933.066
Fehlende Werte	0
Verschiedene Werte	289.930

## Dimensionsschlüssel der Versichertendimension

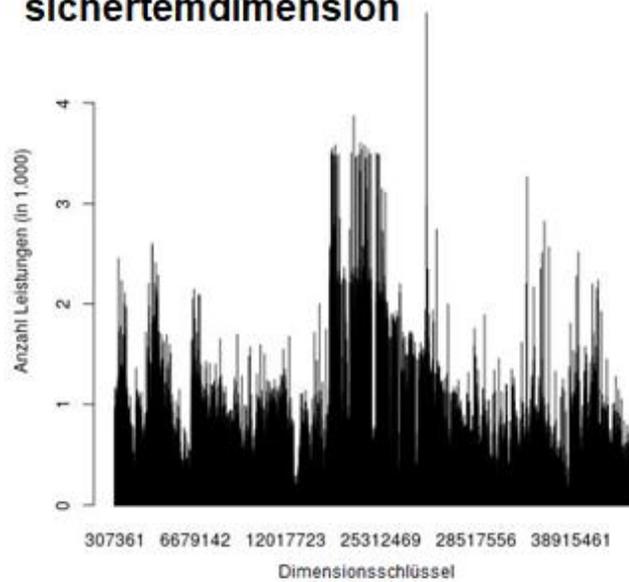


Abbildung 24: Verteilung der Variable *Vers\_key* (Dimensionsschlüssel der Versichertendimension).

### 3.3.7 Variable *Lstg\_key*: Dimensionsschlüssel der Leistungsdimension

In **Tabelle 29** ist das Datenprofil des Dimensionsschlüssels der Leistungsdimension (*Lstg\_key*) aus den Leistungsdaten (*Med\_versorgung*) ersichtlich. Darin sind numerische Werte zwischen 1 und 1898208 gespeichert. **Abbildung 25** zeigt ein Histogramm der Werte.

Tabelle 29: Datenprofil des Dimensionsschlüssels der Leistungsdimension

Variable	<i>Lstg_key</i>
Datensatz	<i>Med_versorgung</i>
Minimaler Wert	1
Maximaler Wert	1898208
Gesamtzahl Werte	37.933.066
Fehlende Werte	0
Verschiedene Werte	158.960

## Dimensionsschlüssel der Leistungsdimension

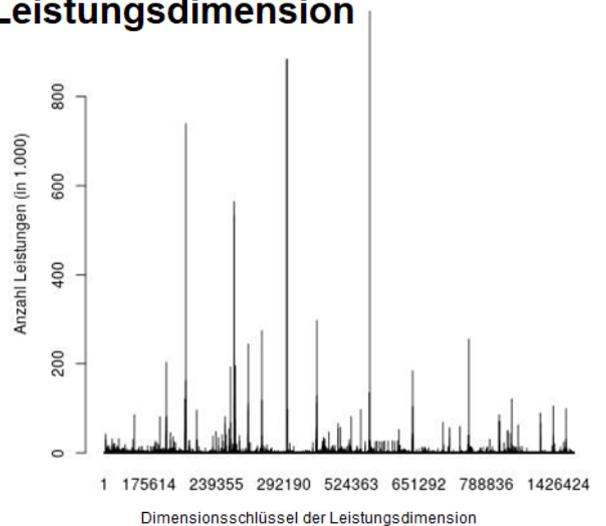


Abbildung 25: Verteilung der Variable *Lstg\_key* (Dimensionsschlüssel der Leistungsdimension)

### 3.3.8 Variable Anzahl

In **Tabelle 30** ist das Datenprofil der Variable *Anzahl* aus den Leistungsdaten (*Med\_versorgung*) ersichtlich. Diese beschreibt wie oft die Leistung abgerechnet wurde. Die numerischen Werte sind zwischen 0 und 2.772. Es fällt auf, dass 34.784.942 Mal eine einzelne Leistung abgerechnet wurde (*Anzahl* = 1), was 91,7% aller Datensätze entspricht und somit mit Abstand der häufigste Wert ist.

**Abbildung 26** zeigt ein Histogramm der Werte.

Tabelle 30: Datenprofil der Variable *Anzahl*.

Variable	Anzahl
Datensatz	<i>Med_versorgung</i>
Minimaler Wert	0
Maximaler Wert	2.772
Gesamtzahl Werte	37.933.066
Fehlende Werte	0
Verschiedene Werte	1.304

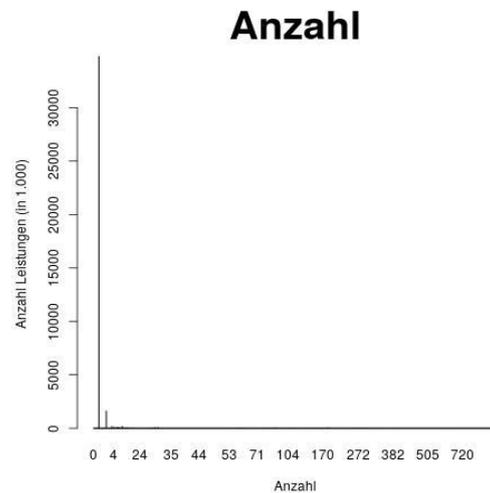


Abbildung 26: Verteilung der Variable *Anzahl*.

### 3.4 Der Datensatz *Pflegeheime*

Der Datensatz *Pflegeheime* beschreibt Daten zu den identifizierten österreichischen Pflegeheimen. Siehe dazu auch Kapitel 6 aus *MedVers* [1]. Eine erste Übersicht der Beschreibung der Variablen ist in **Tabelle 31** zu finden, die genauen Profile sind pro Variable in den folgenden Unterkapiteln zu finden.

Tabelle 31: Übersicht der Variablen zu den Pflegeheimdaten (*Pflegeheime*)

Datensatz: <i>Pflegheime</i>	Pflegeheime
Manuell erstellter Datensatz zu den Pflegeheimen	
<i>Name</i>	Name des Pflegeheimes
<i>Ort</i>	Ort, in dem das Pflegeheim liegt
<i>PLZ</i>	Postleitzahl des Pflegeheimes
<i>Wohnplätze</i>	Anzahl der Wohnplätze
<i>Pflegeplätze</i>	Anzahl der Pflegeplätze
<i>Wohnplätze mit Pflegemöglichkeit</i>	Anzahl der Wohnplätze mit Pflegemöglichkeit
<i>Betreutes/betreubares Wohnen</i>	Anzahl der Plätze für betreutes/betreubares Wohnen
<i>Ärztliche Betreuung text</i>	Art der ärztlichen Betreuung in textueller Form
<i>Ärztliche Betreuung code</i>	Art der ärztlichen Betreuung als Code



**Abbildung 28** zeigt ein Histogramm der Werte.

Tabelle 34: Datenprofil der Variable *Ort*.

Variable	<i>Ort</i>
Datensatz	<i>Pflegeheime</i>
Gesamtzahl Werte	890
Fehlende Werte	0
Verschiedene Werte	638

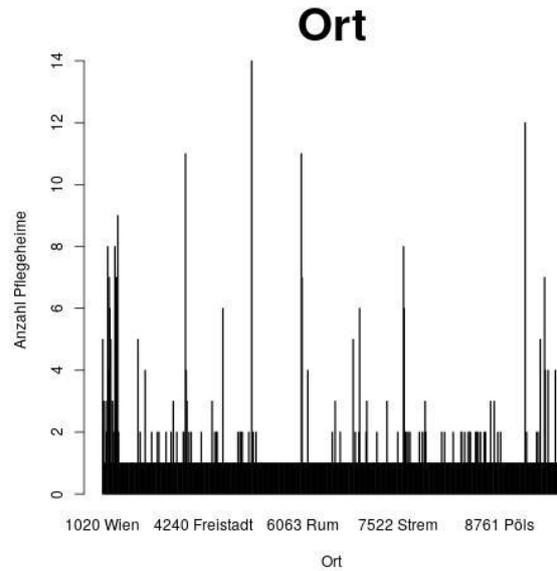


Abbildung 28: Verteilung der Variable *Ort*

### 3.4.3 Variable *PLZ*: Postleitzahl

In **Tabelle 35** ist das Datenprofil der Variable *Plz*, der Postleitzahl des Pflegeheimes aus dem Datensatz *Pflegeheime* ersichtlich. Die Variable enthält numerische Werte zwischen 1020 und 9971.

**Abbildung 29** zeigt ein Histogramm der Werte.

Tabelle 35: Datenprofil der Postleitzahl aus *Pflegeheime*.

Variable	<i>Plz</i>
Datensatz	<i>Pflegeheime</i>
Minimaler Wert	1020
Maximaler Wert	9971
Gesamtzahl Werte	890
Fehlende Werte	0
Verschiedene Werte	621

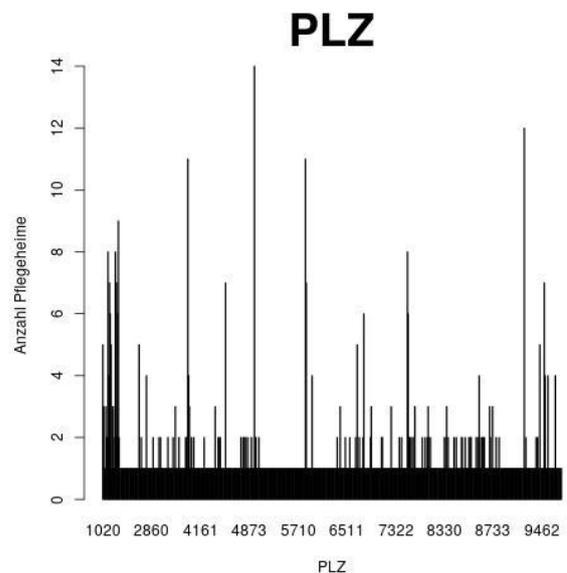


Abbildung 29: Verteilung der Variable *Plz* (Postleitzahl).

### 3.4.4 Variable Wohnplätze

In **Tabelle 36** ist das Datenprofil der Variable *Wohnplätze* aus *Pflegeheime* ersichtlich. Diese beschreibt die Anzahl der Wohnplätze des Pflegeheimes. Die Variable enthält hauptsächlich numerische Werte zwischen 0 und 340, sowie den textuellen Eintrag „Die Einrichtung konnte am Server nicht gefunden werden. Offenbar wurde der Eintrag gelöscht.“, welcher auch mit 12 Datensätzen der Häufigste ist. Die Zahl der fehlenden Werte liegt hier bei 84 % (747 Datensätze).

**Abbildung 30** zeigt ein Histogramm der Werte.

Tabelle 36: Datenprofil der Variable *Wohnplätze*.

Variable	<i>Wohnplätze</i>
Datensatz	<i>Pflegeheime</i>
Minimaler Wert	0
Maximaler Wert	340
Gesamtzahl Werte	142
Fehlende Werte	747
Verschiedene Werte	81

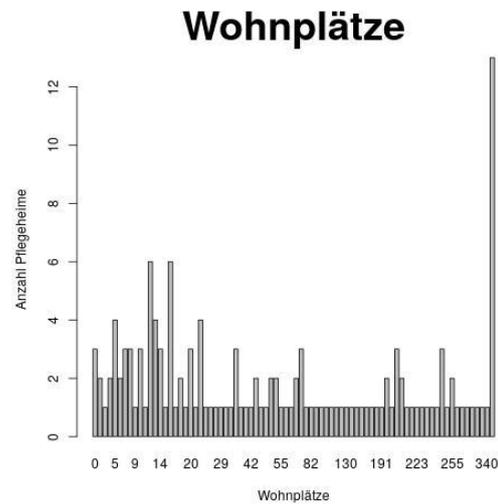


Abbildung 30: Verteilung der Variable *Wohnplätze*.

### 3.4.5 Variable Pflegeplätze

In **Tabelle 37** ist das Datenprofil der Variable *Pflegeplätze* aus *Pflegeheime* ersichtlich. Diese beschreibt die Anzahl der Pflegeplätze des Pflegeheimes. Die Variable enthält numerische Werte zwischen 2 und 392. Die Zahl der fehlenden Werte liegt hier bei 31 % (273 Datensätze). **Abbildung 31** zeigt ein Histogramm der Werte.

Tabelle 37: Datenprofil der Variable *Pflegeplätze*.

Variable	<i>Pflegeplätze</i>
Datensatz	<i>Pflegeheime</i>
Minimaler Wert	2
Maximaler Wert	392
Gesamtzahl Werte	617
Fehlende Werte	273
Verschiedene Werte	163

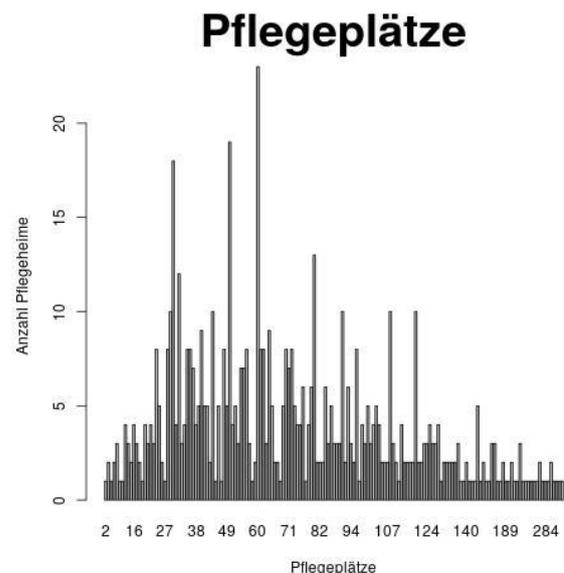


Abbildung 31: Verteilung der Variable *Pflegeplätze*.

### 3.4.6 Variable Wohnplätze mit Pflegemöglichkeit

In **Tabelle 38** ist das Datenprofil der Variable *Wohnplätze mit Pflegemöglichkeit* aus *Pflegeheime* ersichtlich. Diese beschreibt die Anzahl der Wohnplätze mit Pflegemöglichkeit des Pflegeheimes. Die Variable enthält numerische Werte zwischen 0 und 325. Die Zahl der fehlenden Werte liegt hier bei 68 % (601 Datensätze). **Abbildung 32** zeigt ein Histogramm der Werte.

Tabelle 38: Datenprofil der Variable *Wohnplätze mit Pflegemöglichkeit*.

Variable	<i>Wohnplätze mit Pflegemöglichkeit</i>
Datensatz	<i>Pflegeheime</i>
Minimaler Wert	0
Maximaler Wert	325
Gesamtzahl Werte	289
Fehlende Werte	601
Verschiedene Werte	116

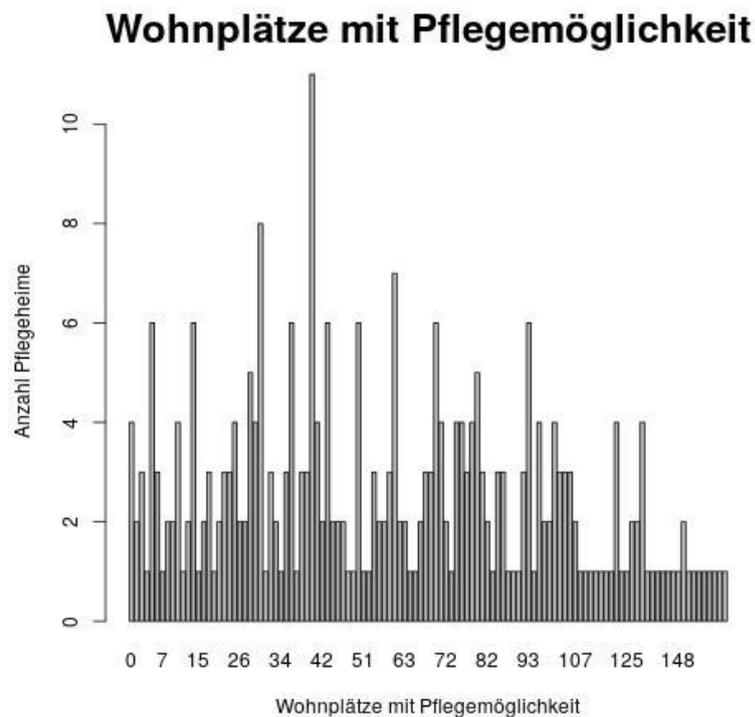


Abbildung 32: Verteilung der Variable *Wohnplätze mit Pflegemöglichkeit*.

### 3.4.7 Variable *betreutes/betreubares Wohnen*

In **Tabelle 39** ist das Datenprofil der Variable *Betreutes/betreubares Wohnen* aus *Pflegeheime* ersichtlich. Diese beschreibt die Anzahl der Plätze für betreutes/betreubares Wohnen des Pflegeheimes. Die Variable enthält numerische Werte zwischen 0 und 231. Die Zahl der fehlenden Werte liegt hier bei 94 % (834 Datensätze). **Abbildung 33** zeigt ein Histogramm der Werte.

Tabelle 39: Datenprofil der Variable *Betreutes/betreubares Wohnen*

Variable	<i>Betreutes/betreubares Wohnen</i>
Datensatz	<i>Pflegeheime</i>
Minimaler Wert	0
Maximaler Wert	231
Gesamtzahl Werte	56
Fehlende Werte	834
Verschiedene Werte	34

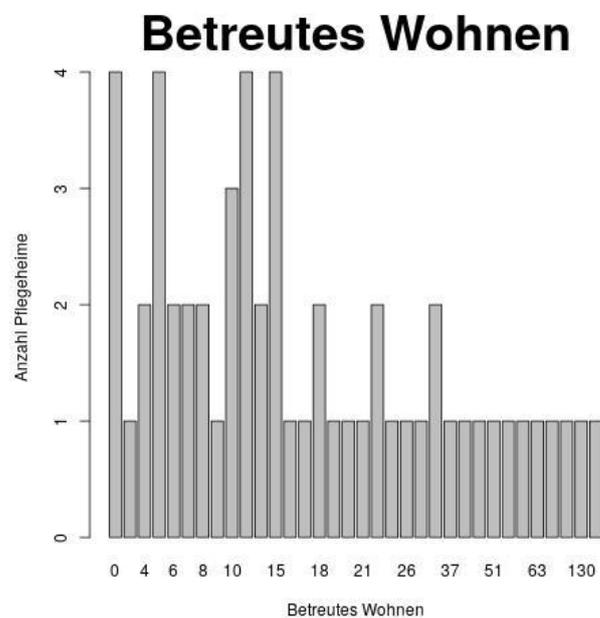


Abbildung 33: Verteilung der Variable *Betreutes Wohnen*

### 3.4.8 Variable *ärztliche Betreuung text*

In **Tabelle 40** ist das Datenprofil der Variable *Ärztliche Betreuung text* aus *Pflegeheime* ersichtlich. Diese beschreibt die Art der ärztlichen Betreuung des Pflegeheimes in Textform. **Abbildung 34** zeigt ein Histogramm der Werte. Mit 140 Datensätzen ist der Text „freie Arztwahl“ der Häufigste.

Tabelle 40: Datenprofil der Variable *Ärztliche Betreuung text*.

Variable	<i>Ärztliche Betreuung text</i>
Datensatz	<i>Pflegeheime</i>
Gesamtzahl Werte	890
Fehlende Werte	0
Verschiedene Werte	485

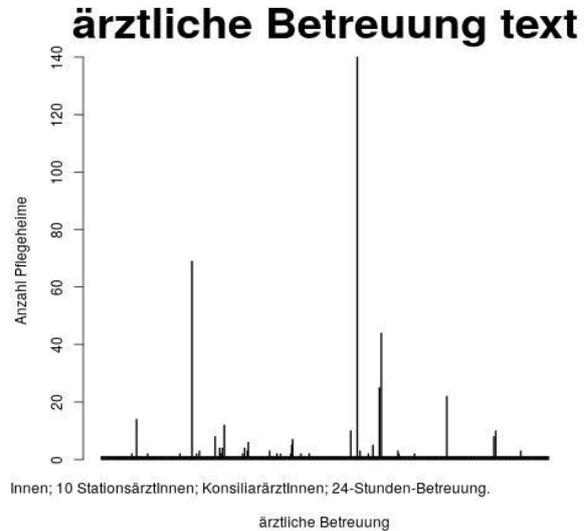


Abbildung 34: Verteilung der Variable *ärztliche Betreuung text*

### 3.4.9 Variable *ärztliche Betreuung code*

In **Tabelle 41** ist das Datenprofil der Variable *Ärztliche Betreuung code* aus *Pflegeheime* ersichtlich. Diese beschreibt die Art der ärztlichen Betreuung des Pflegeheimes als Code. Es sind hier die Werte 1, 2, 3 und 99 vorhanden. Die Bedeutung der Codes wird in **Tabelle 42** dargestellt. Die Zahl der fehlenden Werte liegt hier bei 8 % (69 Datensätze). Mit 571 Datensätzen (64,16%) ist der Code 3 der Häufigste. **Abbildung 35** zeigt ein Histogramm der Werte.

Tabelle 41: Datenprofil der Variable *Ärztliche Betreuung code*.

Variable	<i>Ärztliche Betreuung code</i>
Datensatz	<i>Pflegeheime</i>
Minimaler Wert	1
Maximaler Wert	99
Gesamtzahl Werte	821
Fehlende Werte	69
Verschiedene Werte	4

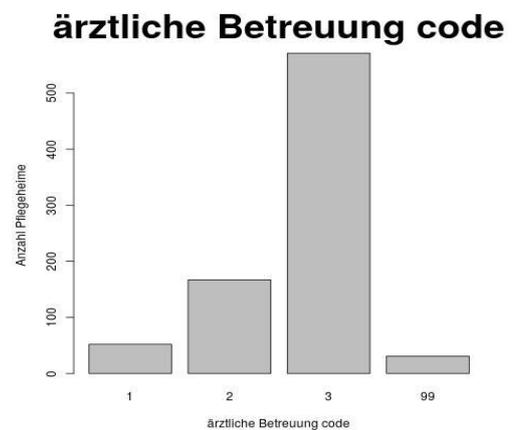


Abbildung 35: Verteilung der Variable *ärztliche Betreuung*.

Tabelle 42: Erklärung der Codes für ärztliche Betreuung.

Code	Erklärung
1	Arzt rund um die Uhr
2	Regelmäßige, organisierte ärztliche Visiten
3	Übliche Hausarztversorgung
99	unklar

## 4 Zusammenfassung der Ergebnisse

Die Daten beziehen sich auf Personen, die im Zeitraum 2014-2016 in einem Pflegeheim waren. In **Tabelle 43** werden die Grunddaten der vier Datensätze nochmals zusammengefasst.

Tabelle 43: Zusammenfassung der Grunddaten

<b>Pfif-Eintritte</b>					
<b>Ausgangslage</b>	<b>Anzahl der Dateneinträge</b>	<b>189.914</b>			
	<b>Anzahl der Personen</b>	<b>85.271</b>			
<i>Anm.: Personen können bei einem Eintrag ein Todesdatum, etc. haben und bei einem anderen nicht/ein anderes.</i>		<b>Anzahl Absolut</b>		<b>Prozent</b>	
		<b>Datensätze</b>	<b>Personen</b>	<b>Datensätze</b>	<b>Personen</b>
Todesdatum	Verstorben	95.021	43.732	50,03%	51,29%
	Lebend	94.893	41.546	49,97%	48,72%
Alter	65-69 Jahre	9.241	4.235	4,87%	4,97%
	70-79 Jahre	31.802	14.332	16,75%	16,81%
	80-89 Jahre	75.913	33.977	39,97%	39,85%
	90-99 Jahre	69.230	31.008	36,45%	36,36%
	100-109 Jahre	3.725	1.717	1,96%	2,01%
	110-119 Jahre	3	2	0,002%	4,97%
Pfleigestufe	Stufe 1	2.513	1.725	1,32%	2,02%
	Stufe 2	11.067	7.412	5,83%	8,69%
	Stufe 3	29.522	18.961	15,54%	22,24%
	Stufe 4	48.722	29.499	25,65%	34,59%
	Stufe 5	63.408	34.548	33,39%	40,52%
	Stufe 6	22.954	12.275	12,09%	14,40%
	Stufe 7	11.728	6.134	6,18%	7,19%
Postleitzahl	Nicht vorhanden	20	11	0,01%	0,01%
<b>Vers_daten</b>					
<b>Ausgangslage</b>	<b>Anzahl der Dateneinträge</b>	<b>1.176.586</b>			
	<b>Anzahl der Personen</b>	<b>83.989</b>			
<i>Anm.: Personen können bei einem Eintrag ein Todesdatum, etc. haben und bei einem anderen nicht/ein anderes.</i>		<b>Anzahl Absolut</b>		<b>Prozent</b>	
		<b>Datensätze</b>	<b>Personen</b>	<b>Datensätze</b>	<b>Personen</b>
Geschlecht	davon männlich	284.434	21.308	24,17%	25,37%
	davon weiblich	860.967	62.756	73,18%	74,72%
	davon nicht vorhanden	31.185	6.632	2,65%	7,90%
Todesdatum	Verstorben	66.827	44.278	5,68%	52,72%
	Lebend	1.109.759	83.981	94,32%	99,99%
Geburtsdatum	Nicht vorhanden	9.535	9.517	0,81%	11,33%
Postleitzahl	Nicht vorhanden	47.347	17.181	4,02%	20,46%
<b>Med_versorgung</b>					
<b>Ausgangslage für Auswertungen: Anzahl an</b>		<b>37.905.996</b>			
Anzahl unterschiedlicher Fachgruppen		74			
Anzahl unterschiedlicher Leistungen		61.122			
<b>Pflegeheime</b>					
Anzahl der Pflegeheime		890			
davon keiner Versorgungsregion		1			

**Fehlende Werte:** Bei der Betrachtung der fehlenden Werte zeigt sich, dass hier zwar fast alle Datensätze betroffen sind. Die große Zahl an fehlenden Todesdaten ist dabei natürlich, da lebende Personen kein Todesdatum haben können. Im Datensatz der Versicherten (*Vers\_daten*) sind Geschlecht, Geburtsdatum und Postleitzahl betroffen, jedoch maximal mit 4 %. Im Datensatz der Leistungen (*Med\_versorgung*) ist vor allem die Fachgruppe betroffen. Diese fehlt in 45% der Datensätze. Im Datensatz Pflegeheime fehlen die Angaben über die Platzzahlen, abhängig von der Art der Plätze, in 30-90 % der Datensätzen. Dies ist dadurch zu erklären, dass die Daten händisch im Internet gesucht wurden und wohl oft auf den Homepages der Heime keine Angaben gefunden werden konnten.

**Identifikation jener den Dateneintrag eindeutig identifizierenden Variablen:** Eine weitere Auffälligkeit im Datensatz *Pflegeheime* ist die Variable *Name*. Da diese nicht einmal zusammen mit der Postleitzahl den Dateneintrag eindeutig identifiziert, fällt es relativ schwer einen Primärschlüssel zu bestimmen. Dieser setzt sich daher aus drei Variablen zusammen: Name, Pflegeplätze und Postleitzahl.

**Verteilungen:** Bei genauer Betrachtung der Variablen sticht die Verteilung der Fachgruppen ins Auge. Es wurden 32% der Datensätze in der Fachgruppe 1, also in der Allgemeinmedizin, abgerechnet. Neben den fehlenden 45%, verteilen sich also nur 23% der Datensätze auf die restlichen 73 Fachgruppen. Auch bei der Leistungsart sind die Einträge sehr konzentriert. Über 90 % der Datensätze verteilen sich auf die zwei stärksten Einträge (Eigenkosten und Heilmittel).

**Verknüpfung der Datensätze:** Auch fällt auf, dass die Pflegeheime keine direkte Verknüpfung zu den anderen Datensätzen aufweisen. Es kann hier lediglich eine Verbindung über die Postleitzahl gezogen werden, die jedoch nicht eindeutig ist.

**Widersprüchliche Angaben:** Dadurch, dass Personen in *Pfif\_Eintritte* als auch in den Versichertendaten (*Vers\_daten*) mehrfach aufscheinen können, kann es in diversen personenbezogenen Variablen zu Unstimmigkeiten kommen. In *Pfif\_Eintritte* wird beispielsweise bei Pflegestufenänderung ein neuer Eintrag angelegt. Bei den Versichertendaten kommt es aufgrund der Tatsache, dass eine Person bei mehreren Trägern versichert sein kann zu Unstimmigkeiten. Darüber hinaus kommt es nicht nur innerhalb der Datensätze zu Unstimmigkeiten, sondern auch zwischen den Datensätzen, da einige Variablen dieselben personenbezogenen Merkmale beschreiben:

- **Alter:** Weiters fällt auf, dass es drei verschiedene Variablen gibt, die sich auf das Alter beziehen. Es gibt die Variablen der Altersdekaden (*ALTDEKA*) und Geburtsdatum (*Gebdatum*) im Datensatz der Versicherten (*vers\_daten*), sowie die Variable *AAlder* im Datensatz *Pfif\_Eintritte*.
- **Sterbedatum:** Ähnlich verhält es sich beim Sterbedatum: Bei den Versichertendaten (*vers\_daten*) gibt es die Variable *Toddatum* und in *Pfif\_Eintritte* die Variable *Tdat*. Innerhalb der Versichertendaten (*vers\_daten*) haben 16 Personen nicht eindeutige, widersprüchliche Angaben zum Todesdatum (Todesdatum vorhanden, aber nicht gleich). In *Pfif\_Eintritte* trifft dies auf 65 Personen zu und zwischen den Datensätzen betrifft es 85 Personen. In Summe (Personen nur ein Mal gezählt) betrifft dies 89 Personen.
- **Geburtsdatum:** Das Geburtsdatum in den Versichertendaten kann aufgrund der Mehrfacheintragung bei unterschiedlichen Versicherungsträgern ebenfalls unterschiedlich sein. Bei 293 Personen ist es daher nicht eindeutig.
- **Postleitzahl:** Die Postleitzahl kommt ebenfalls sowohl in den Versichertendaten als auch in den *Pfif\_Eintritten* vor. Bei 21.882 Personen ist die Postleitzahl nicht übereinstimmend.

Sowohl innerhalb des jeweiligen Datensatzes, als auch nach Verknüpfung der Datensätze kann es aufgrund der Datenstruktur zu mehreren und teilweise unterschiedlichen Werten bezüglich dieser Variablen kommen.

## 5 Anhang

Tabelle 44: Fachgruppencodes Erklärung

Code	Erklärung
01	Arzt für Allgemeinmedizin
02	Facharzt für Anästhesiologie und Intensivmedizin
03	Facharzt für Augenheilkunde und Optometrie
04	Facharzt für Chirurgie
05	Facharzt für Haut- und Geschlechtskrankheiten
06	Facharzt für Frauenheilkunde und Geburtshilfe
07	Facharzt für Innere Medizin
08	Facharzt für Kinderheilkunde
09	Facharzt für Hals-, Nasen- und Ohrenkrankheiten
10	Facharzt für Lungenkrankheiten
11	Facharzt für Neurologie und Psychiatrie
12	Facharzt für Orthopädie und orthopädische Chirurgie
13	Facharzt Physikalische Medizin
14	Facharzt Radiologie
15	Facharzt für Unfallchirurgie
16	Facharzt für Urologie
17	Facharzt für Zahn-, Mund- und Kieferheilkunde (ohn
18	Facharzt für Neurochirurgie
19	Facharzt für Neurologie
20	Facharzt für Psychiatrie
21	Facharzt für Plastische Chirurgie
22	Facharzt für Kinderchirurgie
23	Facharzt für Mund-, Kiefer- und Gesichtschirurgie
24	Facharzt für Nuklearmedizin
25	Facharzt für medizinische Radiologie-Diagnostik
26	Facharzt für Strahlentherapie - Radioonkologie
27	Dr. med. dent
28	Facharzt für Blutgruppenserologie und Transfusions
29	Facharzt für Immunologie
30	kfo
32	Facharzt für Kinder- und Jugendpsychiatrie
32	Facharzt für Kinder- und Jugendpsychiatrie
33	Facharzt für Histologie und Embryologie
34	Facharzt für medizinische Biologie
35	Facharzt für Virologie
36	Gruppenpraxen (Fachgebiet wurden aufgelassen)
37	Facharzt für Arbeits- und Betriebsmedizin

Code	Erklärung
38	Facharzt für Pharmakologie und Toxikologie
39	medizinische Genetik
40	Hörgeräteakustiker
41	Leihfirmen für Hör- und Sprechgeräte
42	Leihfirmen für Heilbehelfe und Hilfsmittel
43	Augenprothetiker
49	Sonst. Nicht kurativ tätige, freiberufliche Ärzte
50	Facharzt für medizinische und chemische Labordiagn
51	Labor, EEG
52	Labor, zytodiagnostisch
53	Facharzt für Pathologie
55	Facharzt für Hygiene u. Mikrobiologie (FA f. mikro
56	Kinderkranken. Und Säuglingsschwester/-pfleger
59	Hausapothekenführender Arzt
60	öffentliche Apotheke
61	Bandagist
62	Dentist
63	Diplomierter Physiotherapeut
64	gewerblicher Masseur
65	Rettungsdienste, Beförderungsgewerbe
66	Diplomierter Logopädie
67	Augenoptiker / Kontaktlinsenoptiker
68	Orthopädieschuhmacher
69	Orthopädietechnik (Prothesenhersteller usw.)
70	Hebamme
71	Diplomierte(r) Krankenschwester/-pfleger
72	Psychotherapeut
73	klinischer Psychologe
74	Psychotherapie und klinische Psychologie (Fachgebi
75	Diplomierter Ergotherapeut
80	Krankenanstalt - stationäre und ambulante Behandlu
84	CT/MR und and. Leistungen (nuklear)
85	selbständiges Ambulatorium, ausgenommen für ZMK un
86	selbständiges Ambulatorium für Zahn-, Mund- und Ki
87	Pflegeheim für chronisch Kranke
90	Genesungsheim, Kurheim, sonstige Heime
91	selbständiges Ambulatorium für physikalische Mediz
92	Hauskrankenpflegeorganisation
99	Andere Vertragspartner, soweit sie nicht den Fachgebieten 40-43, 50-91 zuzuordnen sind, z.B.: Universitätsinstitut (sofern nicht als Teil einer Krankenanstalt), Bundesstaatliche Untersuchungsanstalt, Sonstige Unternehmen (Handelsunternehmen, Frisöre und Perückenmacher, etc.)

## 6 Tabellenverzeichnis

Tabelle 1: Übersicht der Variablen der Pflegegeld-Datenbank ( <i>Pfif_Eintritte</i> ).....	7
Tabelle 2: Datenprofil der Patienten ID.....	7
Tabelle 3: Datenprofil der Pflegestufe. ....	8
Tabelle 4: Datenprofil des Übergangsgrundes. ....	9
Tabelle 5: Datenprofil der Postleitzahl. ....	9
Tabelle 6: Datenprofil des Bundeslandes.....	10
Tabelle 7: Definition der Codes der Bundesländer. ....	10
Tabelle 8: Datenprofil der Variable zu Alter.....	11
Tabelle 9: Datenprofil der Variable zum Sterbedatum. ....	12
Tabelle 10: Datenprofil der Variable zu Beginndatum. ....	13
Tabelle 11: Übersicht der Variablen zu den Versichertendaten ( <i>Vers_daten</i> ). ....	14
Tabelle 12: Datenprofil der Variable zu den Altersdekaden der Personen in den Versichertendaten. ....	14
Tabelle 13: Datenprofil der Variable <i>Geschlecht</i> .....	15
Tabelle 14: Datenprofil der Variable zum Sterbedatum in den Versichertendaten. ....	16
Tabelle 15: Datenprofil der Variable zum Geburtsdatum in den Versichertendaten. ....	17
Tabelle 16: Datenprofil der Variable zur Postleitzahl in den Versichertendaten.....	18
Tabelle 17: Datenprofil der Variable <i>Von</i> zum Beginn der Versicherung in den Versichertendaten....	18
Tabelle 18: Datenprofil der Variable <i>Bis</i> zum Ende der Versicherung in den Versichertendaten. ....	19
Tabelle 19: Datenprofil des Dimensionsschlüssels.....	20
Tabelle 20: Datenprofil der Variable der Patienten ID.....	21
Tabelle 21: Übersicht der Variablen zu den Leistungsdaten ( <i>Med_versorgung</i> ).....	21
Tabelle 22: Datenprofil der Variable <i>Leistungsart</i> .....	22
Tabelle 23: Definition der Leistungsarten .....	22
Tabelle 24: Datenprofil der Variable Positionsnummer.....	23
Tabelle 25: Datenprofil der Variable <i>Positionsbezeichnung</i> .....	24
Tabelle 26: Datenprofil der Variable <i>Fachgruppe</i> . ....	25
Tabelle 27: Datenprofil der Variable Leistungszeitraum.....	26
Tabelle 28: Datenprofil des Schlüssels des Versichertendateneintrages.....	26
Tabelle 29: Datenprofil des Dimensionsschlüssels der Leistungsdimension .....	27
Tabelle 30: Datenprofil der Variable <i>Anzahl</i> . ....	28
Tabelle 31: Übersicht der Variablen zu den Pflegeheimdaten (Pflegeheime) .....	28
Tabelle 32: Datenprofil der Variable <i>Name</i> .....	29
Tabelle 33: mehrfach vorkommende Namen von Pflegeheimen .....	29
Tabelle 34: Datenprofil der Variable <i>Ort</i> . ....	30
Tabelle 35: Datenprofil der Postleitzahl aus <i>Pflegeheime</i> .....	30
Tabelle 36: Datenprofil der Variable <i>Wohnplätze</i> .....	31
Tabelle 37: Datenprofil der Variable Pflegeplätze. ....	31
Tabelle 38: Datenprofil der Variable <i>Wohnplätze mit Pflegemöglichkeit</i> .....	32
Tabelle 39: Datenprofil der Variable <i>Betreutes/betreubares Wohnen</i> .....	33
Tabelle 40: Datenprofil der Variable <i>Ärztliche Betreuung text</i> . ....	34
Tabelle 41: Datenprofil der Variable <i>Ärztliche Betreuung code</i> . ....	34
Tabelle 42: Erklärung der Codes für ärztliche Betreuung. ....	34
Tabelle 43: Zusammenfassung der Grunddaten .....	35
Tabelle 44: Fachgruppencodes Erklärung .....	37

## 7 Abbildungsverzeichnis

Abbildung 1: Rohdatenschema der vier Datensätze zu Pflegeheimen, Personen in Pflegeheimen sowie Leistungen und Versicherungsdaten dieser Personen.....	6
Abbildung 2: Verteilung der Variable <i>Pseudonr_patient</i> (Patienten ID).....	7
Abbildung 3: Verteilung der Variable <i>Pstu_ext</i> (Pfleigestufe).....	8
Abbildung 4: Verteilung der Variablen <i>Uebg</i> (Übergangsgrund).....	9
Abbildung 5: Verteilung der Variable <i>Plz</i> (Postleitzahl).....	9
Abbildung 6: Verteilung der Variable <i>Bld</i> (Bundesland), Legende siehe Tabelle 7.....	10
Abbildung 7: Verteilung der Variable <i>AAlder</i> (Alter).....	11
Abbildung 8: Verteilung der Variable <i>Tdat</i> (Todesdatum).....	12
Abbildung 9: Verteilung der Variable <i>Begi</i> (Beginndatum).....	13
Abbildung 10: Verteilung der Variable <i>ALTDEKA</i> (Altersdekaden in 10er Schritten).....	15
Abbildung 11: Verteilung der Variable <i>Geschlecht</i> .....	15
Abbildung 12: Verteilung der Variable <i>Toddatum</i> (Todesdatum).....	16
Abbildung 13: Verteilung der Variable <i>Gebdatum</i> (Geburtsdatum).....	17
Abbildung 14: Verteilung der Variable <i>PLZ</i> (Postleitzahl).....	18
Abbildung 15: Verteilung der Variable <i>Von</i> (Beginn der Versicherung).....	19
Abbildung 16: Verteilung der Variable <i>Bis</i> (Ende der Versicherung einer Person).....	20
Abbildung 17: Verteilung der Variablen <i>Vers_key</i> (Dimensionsschlüssel der Versichertendimension).....	20
Abbildung 18: Verteilung der Variable <i>Pseudonr_patienten</i> (Patienten ID).....	21
Abbildung 19: Verteilung der Variable <i>Leistungsart</i> .....	22
Abbildung 20: Verteilung der Variable <i>Positionsnummer</i> .....	23
Abbildung 21: Verteilung der Variable <i>Positionsbezeichnung</i> .....	24
Abbildung 22: Verteilung der Variable <i>Fachgruppe</i> .....	25
Abbildung 23: Verteilung der Variable <i>Leistungszeitraum</i> .....	26
Abbildung 24: Verteilung der Variable <i>Vers_key</i> (Dimensionsschlüssel der Versichertendimension).....	27
Abbildung 25: Verteilung der Variable <i>Lstg_key</i> (Dimensionsschlüssel der Leistungsdimension).....	27
Abbildung 26: Verteilung der Variable <i>Anzahl</i> .....	28
Abbildung 27: Verteilung der Variable <i>Name</i> .....	29
Abbildung 28: Verteilung der Variable <i>Ort</i> .....	30
Abbildung 29: Verteilung der Variable <i>Plz</i> (Postleitzahl).....	30
Abbildung 30: Verteilung der Variable <i>Wohnplätze</i> .....	31
Abbildung 31: Verteilung der Variable <i>Pflegeplätze</i> .....	31
Abbildung 32: Verteilung der Variable <i>Wohnplätze mit Pflegemöglichkeit</i> .....	32
Abbildung 33: Verteilung der Variable <i>Betreutes Wohnen</i> .....	33
Abbildung 34: Verteilung der Variable <i>ärztliche Betreuung text</i> .....	34
Abbildung 35: Verteilung der Variable <i>ärztliche Betreuung. code</i> .....	34

## 8 Quellenverzeichnis

[1] I. Wilbacher, S. Scheffel, B. Glock, M. Zechmeister. *Medizinische Versorgung in Pflegeheimen In Österreich*. Hauptverband der österreichischen Sozialversicherungsträger, 2017.

<http://www.hauptverband.at/cdscontent/load?contentid=10008.647827&version=1513857256>