

Auswertung der Forschungsdatenzentrum-Gesundheit-Daten bei Menschen mit Osteogenesis imperfecta (ICD-10 Q78.0)

Vorschlag #2

H.M.Č.

2025-12-17

Inhaltsverzeichnis

1	Beschreibung	3
1.1	Datengrundlage	3
1.2	Struktur der Ergebnistabellen	3
2	Tabelle I - Aggregierte Darstellung der Alter- und Geschlechtsverteilung	3
2.1	Dimensionsberechnung	4
3	Tabelle II - Aggregierte Darstellung der spezifischen ICD-10 Diagnosen bei der Population der Q78.0	4
3.1	Ansatz 1: Gruppierte Codes	4
3.2	Ansatz 2: Einzelcodes-Abfrage	5
3.3	Dimensionsberechnung	6
4	Tabelle III - Aggregierte Darstellung der spezifischen ICD-10 Diagnosen bei Restpopulation	7
4.1	Dimensionsberechnung	7
5	Tabelle IV - Aggregierte PLZ und Geschlechter für PSIDs mit und ohne den Code Q78.0	8
5.1	Dimensionsberechnung	8
6	Finale Dimensionberechnung der gesamten Ergebnismenge	9
7	Berechnungen	9
7.1	Relative Aussagen	10

8	FDZ Q&A	11
8.1	ICD-Codes mit Zusatzkennzeichen (G, A, V, Z)	11
8.2	Bedeutung der Symbole in der Spalte DIAGSICH der Tabelle KHDIAG:	11
8.3	PSID-Geschlecht Beziehungen	12

1 Beschreibung

Dieses Dokument beschreibt die methodischen Ansätze zur Analyse der vom Forschungsdatenzentrum Gesundheit (FDZ-G) bereitgestellten Daten sowie die daraus resultierenden Ergebnistabellen mit dem Fokus auf Osteogenesis imperfecta (ICD-10 Q78.0). Das erstellte Skript eignet sich vollständig für alle vorhandenen Berichtsjahren. Das Alter der PSIDs wird auf Basis des jeweiligen Berichtsjahres als Differenz zwischen Berichtsjahr und Geburtsjahr berechnet.

1.1 Datengrundlage

Die Analysen basieren auf dem **Datenmodell 3** des FDZ-G. Folgende Tabellen werden für die Abfragen verwendet:

- **AMBDIAG**: Ambulante Diagnosen (PSID, ICDAMB_CODE, DIAGSICH, BJAHHR)
- **KHDIAG**: Stationäre Diagnosen (PSID, ICDKH_CODE, BJAHHR)
- **VERS**: Versichertenstammdaten (PSID, GEBJAHR, PLZ)
- **VERSQ**: Versichertenmerkmale (PSID, GESCHLECHT)

1.2 Struktur der Ergebnistabellen

Für jede der nachfolgend beschriebenen Ergebnistabellen werden folgende Informationen dargestellt:

- **Inhalt**: Beschreibung der erfassten Variablen
- **Dimensionen**: Anzahl der Zeilen und Spalten der Ergebnistabelle

Hinweis zur Dimensionierung: Gemäß den Programmcoderegeln des FDZ-G darf die gesamte Ergebnismenge maximal 2.400 Datenzellen (Zeilen \times Spalten) umfassen. Die Dimensionsangaben dienen der Überprüfung dieser Vorgabe. Referenz auf Seite 8 im Nebenbestimmungen – Anlage 2: Programmcoderegeln und auf Seite 16 im Nutzerhandbuch für Forschende.

2 Tabelle I - Aggregierte Darstellung der Alter- und Geschlechtsverteilung

Die Zuordnung der Geschlechtscodes erfolgt deterministisch auf Grundlage der VERSQ-Tabelle. Im 1 %-PUF treten PSIDs auf, denen mehr als ein Geschlecht zugeordnet ist. Da im Produktivdatensatz plausiblerweise überwiegend eindeutige 1:1-Beziehungen zwischen PSID und Geschlecht zu erwarten sind, werden im Skript ausschließlich PSIDs mit einer eindeutigen Geschlechtszuordnung in die Analyse eingeschlossen. PSIDs mit mehrfacher Geschlechtszuordnung werden ausgeschlossen, da diese entweder auf Datenbankinkonsistenzen oder auf tatsächliche Geschlechtsänderungen zurückzuführen sein können und somit außerhalb des Analysefokus liegen.

RD cohort = patients with at least one confirmed (DIAGSICH = 'G') diagnosis of Q78.0, and age defined at first recorded Q78.0 diagnosis year.

NON RD cohort = Age at first recorded diagnosis of any kind.

Die Tabelle 1 (**RT_RD_NON_RD_AGE_SEX_DIST**) beinhaltet alle PSIDs mit einer Q78.0 ICD-10 Diagnose, gruppiert basierend auf Alter und Geschlecht, sowie die Gruppierungen der nicht-OI-Population (gekennzeichnet durch !Q78.0) und ist folgendermaßen zu interpretieren:

Der Code Q78.0 würde in der weiblichen Altersgruppe x-y bei 55 PSIDs gefunden.

[1] "Die Ergebnistabelle 1 hat 12 Zeilen und 4 Spalten."

Table 1: RT_RD_NON_RD_AGE_SEX_DIST

ICD_CODE	ALTERSGRUPPE	GESCHLECHT_LABEL	CNT_D_PSID
Q780	0-18	Male	3
Q780	19-49	Male	6
Q780	50+	Male	10
Q780	19-49	Female	5
Q780	50+	Female	17
Q780	0-18	Female	5
!Q780	0-18	Male	24888
!Q780	19-49	Female	75511
!Q780	19-49	Male	62768
!Q780	0-18	Female	30573
!Q780	50+	Male	75752
!Q780	50+	Female	91119

2.1 Dimensionsberechnung

OI-Population: 3 Altersgruppen x 2 Geschlechter x 4 Spalten = 24 Datenzellen.

Nicht-OI Population: 3 Altersgruppen x 2 Geschlechter x 4 Spalten = 24 Datenzellen.

Insgesamt: OI-Population + Nicht-OI Population = 48 Datenzellen

PSID-Zellen = 12

[TBD] If value for CNT_D_PSID of OI is <5, then write <5. if value 0, write 0. Only for this table?

3 Tabelle II - Aggregierte Darstellung der spezifischen ICD-10 Diagnosen bei der Population der Q78.0

Die Tabelle 2 (**RT_SPECIFIC_ICD_AGE_SEX_DIST**) zeigt die Häufigkeit spezifischer ICD-10 Diagnosen bei Patienten mit Q78.0, stratifiziert nach Alter und Geschlecht für alle Berichtsjahre.

Zur Erfassung relevanter Komorbiditäten werden zwei Abfrageansätze verwendet.

3.1 Ansatz 1: Gruppierte Codes

ICD-Codes werden auf Ebene der dreistelligen Kategorien gruppiert. Alle Unterkategorien eines Codes werden zusammengefasst (z.B. werden S02.0, S02.1, S02.2 etc. zur Gruppe S02 aggregiert).

Kapitel IX, Krankheiten des Kreislaufsystems:

- Essentielle (primäre) Hypertonie (I10.-);
- Sekundäre Hypertonie (I15.-);

Kapitel XIII, Krankheiten des Muskel-Skelett-Systems und des Bindegewebes:

- Rückenschmerzen (M54.-)
- Sonstige Gelenkrankheiten, anderenorts nicht klassifiziert (M25.-)
- Sonstige Krankheiten des Weichteilgewebes, anderenorts nicht klassifiziert (M79.-) Aus dem Kapitel XVIII, Symptome und abnorme klinische und Laborbefunde, die anderenorts nicht klassifiziert sind:
- Schmerz, anderenorts nicht klassifiziert (R52.-)

Kapitel XV, Schwangerschaft, Geburt und Wochenbett

- Mehrlingsschwangerschaft (O30.-);

Kapitel XIX, Verletzungen, Vergiftungen und bestimmte andere Folgen:

- Fraktur des Schädels und der Gesichtsschädelknochen (S02.-)
- Fraktur im Bereich des Halses (S12.-)
- Fraktur der Rippe(n), des Sternums und der Brustwirbelsäule (S22.-)
- Fraktur der Lendenwirbelsäule und des Beckens (S32.-)
- Fraktur im Bereich der Schulter und des Oberarmes (S42.-)
- Fraktur des Unterarmes (S52.-)
- Fraktur im Bereich des Handgelenkes und der Hand (S62.-)
- Fraktur des Femurs (S72.-)
- Fraktur des Unterschenkels, einschließlich des oberen Sprunggelenkes (S82.-)
- Fraktur des Fußes [ausgenommen oberes Sprunggelenk] (S92.-)
- Frakturen mit Beteiligung mehrerer Körperregionen (T02.-)
- Fraktur der Wirbelsäule, Höhe nicht näher bezeichnet (T08.-)
- Fraktur der oberen Extremität, Höhe nicht näher bezeichnet Gebrochener Arm o.n.A. Fraktur des Armes o.n.A. (T10.-)
- Fraktur der unteren Extremität, Höhe nicht näher bezeichnet Gebrochenes Bein o.n.A. Fraktur des Beines o.n.A. (T12.-)

3.2 Ansatz 2: Einzelcodes-Abfrage

Folgende Codes werden als vollständige, spezifische Diagnosen ohne Gruppierung abgefragt:

- Unwohlsein und Ermüdung (R53);
- Neurasthenie (F480);
- Erschöpfung durch übermäßige Anstrengung (T733).
- Spontangeburt eines Einlings (080);
- Geburt eines Einlings durch Zangen- oder Vakuumextraktion (O81);
- Geburt eines Einlings durch Schnittentbindung [Sectio caesarea] (O82).

Table 2: RT_SPECIFIC_ICD_AGE_SEX_DIST

ICD_GROUP	GESCHLECHT_LABEL	AGE_GROUP	CNT_D_PSID
F480	Female	19-49	1
F480	Female	50+	3
F480	Male	50+	1
I10	Female	0-18	5
I10	Female	19-49	5
I10	Female	50+	15
I10	Male	0-18	1
I10	Male	19-49	13
I10	Male	50+	11
M25	Female	19-49	2
M25	Female	50+	2
M25	Male	0-18	1
M25	Male	19-49	2
M25	Male	50+	1
M54	Female	0-18	4
M54	Female	19-49	3
M54	Female	50+	9
M54	Male	0-18	2
M54	Male	19-49	4
M54	Male	50+	4

3.3 Dimensionsberechnung

Die Abfrage umfasst 29 ICD-10 Codes, stratifiziert nach 3 Altersgruppen und 2 Geschlechtern. Dies ergibt maximal 174 Zeilen ($29 \times 3 \times 2$) * 4 Spalten = 696 Datenzellen. Diese Berechnung setzt voraus, dass alle abgefragten Codes in den Produktivdaten vorhanden sind. Die tatsächliche Zeilenanzahl kann geringer ausfallen, falls einzelne Kombinationen nicht vorkommen.

```
## [1] "Die Ergebnistabelle 2 hat 44 Zeilen und 4 Spalten."
```

```
## Spezifische Codes gefunden in dem 1% PUF:
```

```
##
## F480 I10 M25 M54 M79 O80 R52 R53 S02 S12 S42 S52 S62 S72 S82 S92
## 3 6 5 6 2 1 5 2 2 1 2 3 1 1 3 1
```

PSID-Zellen = 174

4 Tabelle III - Aggregierte Darstellung der spezifischen ICD-10 Diagnosen bei Restpopulation

Tabelle 3 **RT_SPECIFIC_ICD_AGE_SEX_DIST_REMAIN** dient als Vergleichstabelle zur Tabelle 2 und enthält die identischen ICD-10 Codes, jedoch für die Vergleichspopulation ohne Q78.0-Diagnose (Restpopulation).

```
## [1] "Die Ergebnistabelle 3 hat 162 Zeilen und 4 Spalten."
```

```
## Spezifische Codes gefunden in dem 1% PUF:
```

```
##
## F480 I10 I15 M25 M54 M79 O30 O80 O81 O82 R52 R53 S02 S12 S22 S32
##      6   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6
## S42 S52 S62 S72 S82 S92 T02 T08 T10 T12 T733
##      6   6   6   6   6   6   6   6   6   6   6
```

Table 3: RT_SPECIFIC_ICD_AGE_SEX_DIST_REMAIN

ICD_GROUP	GESCHLECHT_LABEL	AGE_GROUP	CNT_D_PSID
F480	Female	0-18	795
F480	Female	19-49	1952
F480	Female	50+	2328
F480	Male	0-18	604
F480	Male	19-49	1615
F480	Male	50+	1884
I10	Female	0-18	13585
I10	Female	19-49	32567
I10	Female	50+	38676
I10	Male	0-18	11004
I10	Male	19-49	26987
I10	Male	50+	32096
I15	Female	0-18	181
I15	Female	19-49	445
I15	Female	50+	573
I15	Male	0-18	142
I15	Male	19-49	352
I15	Male	50+	447
M25	Female	0-18	2135
M25	Female	19-49	5440

4.1 Dimensionsberechnung

Siehe Dimensionsberechnung Tabelle 2 (**RT_SPECIFIC_ICD_AGE_SEX_DIST**).

Die Abfrage umfasst die gleichen 29 ICD-10 Codes, stratifiziert nach 3 Altersgruppen und 2 Geschlechtern. Dies ergibt maximal 174 Zeilen ($29 \times 3 \times 2$) * 4 Spalten = 696 Datenzellen. Diese Berechnung setzt voraus, dass alle abgefragten Codes in den Produktivdaten vorhanden sind. Die tatsächliche Zeilenanzahl kann geringer ausfallen, falls einzelne Kombinationen nicht vorkommen.

5 Tabelle IV - Aggregierte PLZ und Geschlechter für PSIDs mit und ohne den Code Q78.0

Die Tabelle 4 (RT_PLZ_OI_NON_OI_SEX) dient als Vergleichstabelle der OI- und Nicht-OI-Population hinsichtlich der Verteilung nach Postleitzahlen. Die Auswahl der PLZ-Stelligkeit wird im Rahmen der Antragsstellung festgelegt und in entsprechender Form im FDZ-Analyse Raum bereitgestellt. Aus diesem Grund ist das Skript so gestaltet, dass die Postleitzahlen ohne weitere Aggregation direkt in die Ergebnistabelle übernommen werden. Entsprechend sind im vorliegenden 1%-PUF fünfstellige Postleitzahlen ausgewiesen, da diese in diesem Format im Datensatz enthalten sind.

- Die Q78.0-Population, aufgeteilt nach Geschlecht und PLZ-Region
- Die Restpopulation (Patienten ohne Q78.0), mit denselben Merkmalen.

[1] "Die Ergebnistabelle 4 hat 8179 Zeilen und 6 Spalten."

Table 4: RT PLZ OI NON OI SEX

PLZ	CNT_D_PSID_TOTAL	CNT_D_PSID_OI_F	CNT_D_PSID_OI_M	CNT_D_PSID_REMAIN_F	CNT_D_PSID_REMAIN_M
00000	1835	0	0	977	858
50354	128	0	0	66	62
76437	127	0	0	78	49
65428	125	0	0	70	55
40764	123	0	0	69	54
52249	122	0	0	65	57
52525	120	0	0	67	53
59192	119	0	0	61	58
53332	118	0	0	57	61
33378	118	0	0	61	57
78224	117	0	0	64	53
94315	116	0	0	60	56
47533	115	0	0	59	56
70794	114	0	0	60	54
52477	112	0	0	64	48
91126	112	0	0	53	59
50259	110	0	0	60	50
41334	110	0	1	61	48
41812	109	0	0	51	58
14612	109	0	0	56	53

Die Spaltensummen entsprechen den jeweiligen Gesamtzahlen der Gruppen:

- CNT_D_PSID_TOTAL - Summe der zwei Gruppen/Subpopulationen
- CNT_D_PSID_OI_F - Anzahl der weiblichen PSIDs mit OI
- CNT_D_PSID_OI_M - Anzahl der männlichen PSIDs mit OI
- CNT_D_PSID_REMAIN_F - Anzahl der weiblichen PSIDs ohne OI
- CNT_D_PSID_REMAIN_M - Anzahl der männlichen PSIDs ohne OI

5.1 Dimensionsberechnung

Bei 1-stelligen PLZ sind folgende Dimensionen zu erwarten: 10 PLZ * 5 Spalten = 50 Datenzellen

[TBD] Die Spalte CNT_D_PSID_TOTAL entfernen?

[TBD] Zellenberechnung für PSIDs: 10*4=40

6 Finale Dimensionberechnung der gesamten Ergebnismenge

Folgend der Aussage auf Seite 16 Nutzerhandbuch für Forschende: *So würde zum Beispiel bei einer Tabelle mit den Spalten "Geschlecht", "Alter", "ICD-Code" und "CNT_D_PSID" (=Anzahl Versicherte) nur die Spalte "CNT_D_PSID" zu der Berechnung Maximalanzahl an erlaubten Zellen beitragen.* liegt die Zellenzahl der finalen Ergebnismenge:

(Tabelle 1) **12** + (Tabelle 2) **174** + (Tabelle 3) **174** + (Tabelle 4) **50** = 410

Und bei der Berechnung von allen Spalten:

(Tabelle 1) **48** + (Tabelle 2) **696** + (Tabelle 3) **696** + (Tabelle 4) **50** = 1490

7 Berechnungen

```
cat("The number of total PSIDs, with both populations included: ")
```

```
## The number of total PSIDs, with both populations included:
```

```
sum(RT_PLZ_OI_NON_OI_SEX$CNT_D_PSID_TOTAL)
```

```
## [1] 180353
```

```
cat("The total number of Q78.0 population: ")
```

```
## The total number of Q78.0 population:
```

```
sum(RT_RD_NON_RD_AGE_SEX_DIST$CNT_D_PSID)
```

```
## [1] 360657
```

```
cat("The number of females in the Q78.0 population: ")
```

```
## The number of females in the Q78.0 population:
```

```
sum(RT_PLZ_OI_NON_OI_SEX$CNT_D_PSID_OI_F)
```

```
## [1] 27
```

```
cat("The number of males in the Q78.0 population: ")
```

```
## The number of males in the Q78.0 population:
```

```

sum(RT_PLZ_OI_NON_OI_SEX$CNT_D_PSID_OI_M)

## [1] 19

cat("The total number of the remaining population: ")

## The total number of the remaining population:

sum(RT_PLZ_OI_NON_OI_SEX$CNT_D_PSID_REMAIN_F)+sum(RT_PLZ_OI_NON_OI_SEX$CNT_D_PSID_REMAIN_M)

## [1] 180307

cat("The number of females in the remaining population: ")

## The number of females in the remaining population:

sum(RT_PLZ_OI_NON_OI_SEX$CNT_D_PSID_REMAIN_F)

## [1] 98603

cat("The number of males in the remaining population: ")

## The number of males in the remaining population:

sum(RT_PLZ_OI_NON_OI_SEX$CNT_D_PSID_REMAIN_M)

## [1] 81704

```

7.1 Relative Aussagen

Die nachfolgenden Beispiele zeigen, wie sich die Tabellen für weiterführende Analysen kombinieren lassen.

7.1.1 OI-Population

Aus Tabelle 1 die Gesamtzahl der OI-ler erfahren (`sum(CNT_D_PSID)`):

```
## [1] 360657
```

Aus Tabelle 2 beispielsweise folgenden Eintrag über die Anzahl der weiblichen OI-PSIDs der Altersgruppe 50+ mit einem F48.0 Code entnehmen:

```
##   ICD_GROUP GESCHLECHT_LABEL AGE_GROUP CNT_D_PSID
## 2      F480           Female      50+          3
```

Daraus stellt man fest:

```
## Im Jahr 2019, gab es 3 weibliche OI-PSIDs mit dem Code F48.0 (Neurasthenie) in der
## Altersgruppe 50+, bei einer OI-Gesamtzahl von: 360657 .
```

```
## Dies führt zur Aussage, dass es von insgesamt 360657 OI-PSIDs 3 weibliche OI-PSIDs
## mit dem Code F48.0 in der Altersgruppe 50+.
```

7.1.2 Nicht-OI-Population

Aus der Tabelle 4, die Gesamtzahl der nicht-OI-PSIDs erfahren, in dem man die Anzahl der weiblichen und männlichen nicht-OI-PSIDs addiert:

Aus der Tabelle 3, entsprechend der Auswahl oben, folgenden Eintrag über die Anzahl der weiblichen nicht-OI-PSIDs der Altersgruppe 50+ mit einem F48.0 Code entnehmen:

```
##   ICD_GROUP GESCHLECHT_LABEL AGE_GROUP CNT_D_PSID
## 3      F480           Female      50+      2328
```

Daraus stellt man fest:

```
## Im Jahr 2019, gab es 2328 weibliche nicht-OI-PSIDs mit dem Code F48.0 (Neurasthenie) in der
## Altersgruppe 50+
## bei einer nicht-OI-Gesamtzahl von: 180307
```

```
## Dies führt zur Aussage, dass es von insgesamt: 180307 nicht-OI-PSIDs 2328 weibliche
## nicht-OI-PSIDs mit dem Code F48.0 in der Altersgruppe 50+ gibt.
```

7.1.3 Vergleich der Häufigkeit bei OI- und Nicht-OI-PSIDs

Anhand der Gesamtzahlen der Populationen und der ausgewählten Beispielsgruppe weiblicher PSIDs in der Altersgruppe 50+ mit dem Code F48.0 lassen sich die Häufigkeiten berechnen:

- OI-Population: 3 von 217 PSIDs → Jede 72. OI-PSID in dieser Gruppe weist F48.0 auf.
- Nicht-OI-Population: 9.702 von 760.194 PSIDs → Jede 78. nicht-OI-PSID in dieser Gruppe weist F48.0 auf.

8 FDZ Q&A

8.1 ICD-Codes mit Zusatzkennzeichen (G, A, V, Z)

In den Produktivdaten werden diese Zusatzkennzeichen bereinigt und in das Feld `_ZUSATZ` verschoben, sodass im Feld `_CODE` nur der reine ICD-Code verbleibt. Für Auswertungen kann die Tabelle `DIAGSICH` herangezogen werden. Bitte beachten Sie, dass Datenfehler seitens der Leistungserbringer nicht vollständig ausgeschlossen werden können.

8.2 Bedeutung der Symbole in der Spalte `DIAGSICH` der Tabelle `KHDIAG`:

„+“ = Kreuz-Code

„#“ = Stern-Code Diese beiden Kennzeichen entsprechen der Kreuz-Stern-Systematik der ICD-10-GM

„-“ = ohne besondere Bedeutung. Dieses Zeichen bedeutet, dass keine Kreuz- oder Stern-Kodierung vorliegt „!“ = Zusatzcode Das Ausrufezeichen weist auf einen Zusatzcode zur Primärdiagnose hin, also eine ergänzende Kodierung

Wichtiger Hinweis zum Public Use File (PUF): Aufgrund der Art der Erzeugung des PUFs und der Auflösung von Korrelationen kann es vorkommen, dass Zusatzcodes (z. B. mit „!“) auch ohne zugehörige Primärdiagnose im Datensatz erscheinen.

Zudem möchten wir darauf hinweisen, dass diese Kennzeichnungen auch in den Spalten mit der Endung `„_ZUSATZ“` bei den ICD-Kodierungen vorkommen können.

8.3 PSID-Geschlecht Beziehungen

Das Verhalten der Daten können Sie jedoch im PUF nachvollziehen; in den Echtdaten treten voraussichtlich deutlich weniger Nicht-1:1-Fälle auf als im PUF.