



hochschule mannheim

in Kooperation mit



miracum

Biomedizinische Informatik und Data Science (M.Sc.)

Master of Science – Zertifikatskurse/-programme

## PROJEKTARBEIT

# Technische Abbildung des Lebenszyklus von ICD-10-GM Klassifikationen

*Abel* HODELIN HERNANDEZ 

betreut von  
Marcus WILL

9. September 2021



# Inhaltsverzeichnis

|   |           |
|---|-----------|
| Abbildungsverzeichnis                                 | ii        |
| Tabellenverzeichnis                                   | ii        |
| <b>1 Einführung</b>                                   | <b>4</b>  |
| <b>2 ICD-10-GM</b>                                    | <b>4</b>  |
| 2.1 Alphabetisches Verzeichnis (Alphabet)             | 4         |
| 2.2 Systematisches Verzeichnis (Systematik)           | 5         |
| 2.3 Metadaten   | 5         |
| <b>3 Datenbank Schema</b>                             | <b>6</b>  |
| 3.1 Aufbau der Datenbank                              | 6         |
| 3.2 Funktion der Tabelle <code>kodes</code> von BfArM | 6         |
| 3.3 Funktion der neuen Tabellen                       | 7         |
| 3.4 Fluss der Information                             | 8         |
| <b>4 ETL-Strecke</b>                                  | <b>8</b>  |
| 4.1 Extraktion  | 9         |
| 4.2 Transformation                                    | 9         |
| 4.3 Laden   | 11        |
| <b>5 Data Analyse</b>                                 | <b>11</b> |
| 5.1 Neue ICD-10-GM                                    | 12        |
| 5.2 Modifizierte ICD-10-GM                            | 15        |
| 5.3 Gelöschte ICD-10-GM                               | 17        |
| 5.4 Wiederverwendete ICD-10-GM                        | 19        |
| 5.5 Strukturelle Änderungen                           | 19        |
| <b>6 Schlussfolgerung</b>                             | <b>20</b> |



## Abbildungsverzeichnis

|   |  |    |
|---|--|----|
| 1 | Datenbankstruktur . . . . .                  | 7  |
| 2 | Datenfluss des Prozesses . . . . .           | 9  |
| 3 | ETL-Strecke . . . . .                        | 10 |
| 4 | Neue ICD-10-GM pro Jahr . . . . .            | 13 |
| 5 | Kapitel der ICD-10-GM pro Jahr . . . . .     | 14 |
| 6 | Meldepflichtige ICD-10-GM pro Jahr . . . . . | 14 |
| 7 | Gelöschte ICD-10-GM pro Jahr . . . . .       | 18 |
| 8 | Meist betroffene Kapitel in 2013 . . . . .   | 18 |
| 9 | Hierarchie der gelöschte ICD-10-GM . . . . . | 19 |

## Tabellenverzeichnis

|   |   |    |
|---|---|----|
| 1 | Verschiedene Bezeichnungen von COVID-19 . . . . . | 5  |
| 2 | Inhalt im Ordner Klassifikationsdateien . . . . . | 6  |
| 3 | ICD-10-GM in den CSV-Dateien . . . . .            | 11 |
| 4 | ICD-10-GM in der DB . . . . .                     | 12 |
| 5 | Meldepflichtige ICD-10-GM . . . . .               | 15 |
| 6 | Änderungen in den Metadaten . . . . .             | 16 |
| 7 | Diphtherie und Poliomyelitis . . . . .            | 17 |



## Abkürzungsverzeichnis

|                  |   |    |
|------------------|---|----|
| <b>ICD-10-GM</b> | International Statistical Classification of Diseases and Related Health Problems, 10. Revision, German Modification ..... | 4  |
| <b>CDW</b>       | Clinical Data Warehouse .....   | 4  |
| <b>DIZ</b>       | Datenintegrationszentrum .....  | 4  |
| <b>MIRACUM</b>   | Medical Informatics in Research and Care in University Medicine ....  | 4  |
| <b>MII</b>       | Medizin Informatik Initiative .....   | 4  |
| <b>DB</b>        | Datenbank .....   | 4  |
| <b>ETL</b>       | Extract, Transform, Load .....  | 4  |
| <b>BMG</b>       | Bundesministeriums für Gesundheit .....   | 4  |
| <b>BfArM</b>     | Bundesinstitut für Arzneimittel und Medizinprodukte .....   | 4  |
| <b>SGB</b>       | Sozialgesetzbuch .....  | 4  |
| <b>Alpha-ID</b>  | Alpha-Identifikator .....   | 4  |
| <b>COVID</b>     | Corona Virus Disease  |    |
| <b>ZIP</b>       | Zipper .....  | 5  |
| <b>CSV</b>       | Comma-Separated Values .....  | 5  |
| <b>SQL</b>       | Structured Query Language .....   | 5  |
| <b>RDBMS</b>     | Relational Database Management System .....   | 5  |
| <b>BASH</b>      | Bourne-again shell .....  | 8  |
| <b>KRINKO</b>    | Kommission für Krankenhaushygiene und Infektionsprävention .....  | 12 |
| <b>G-DRG</b>     | German Diagnosis Related Groups .....   | 12 |
| <b>RKI</b>       | Robert-Koch-Institut .....  | 17 |
| <b>EBM</b>       | Einheitlicher Bewertungsmaßstab .....   | 16 |
| <b>WHO</b>       | Weltgesundheitsorganisation .....   | 15 |



# 1 Einführung

Die International Statistical Classification of Diseases and Related Health Problems, 10. Revision, German Modification (ICD-10-GM) auf Deutsch Internationale statistische Klassifikation der Krankheiten und verwandter Gesundheitsprobleme [7] gilt als Standard bei der Auszeichnung von Gesundheitsdaten. Häufig bleibt die temporäre Natur dieses Schemas in Applikationen und Auswertungen unberücksichtigt.

Im Rahmen der Konzeption des Clinical Data Warehouse (CDW) für ein Datenintegrationszentrum (DIZ) des Konsortiums Medical Informatics in Research and Care in University Medicine (MIRACUM) der Medizin Informatik Initiative (MII) muss diese Situation abgebildet werden [21].

Das Ziel dieses Projekts ist die Darstellung der Besonderheiten eines Lebenszyklus der historischen ICD-10-GM Auffassungen von 2007 bis 2021 in einer Datenbank (DB). Um dieses Ziel zu erreichen wurde eine DB-Schema für die Speicherung der Information der ICD-10-GM entwickelt und eine Extract, Transform, Load (ETL)-Strecke konzipiert für den Import der Daten in die DB.

## 2 ICD-10-GM

Die amtliche Klassifikation zur Verschlüsselung von Diagnosen in der ambulanten (§295 Sozialgesetzbuch (SGB) V) und stationären (§301 SGB V) Versorgung in Deutschland ist die ICD-10-GM. Die Versionen und Formaten davon gelten vom Anfang bis zum Ende eines Jahres und werden im Auftrag des Bundesministeriums für Gesundheit (BMG) von dem Bundesinstitut für Arzneimittel und Medizinprodukte (BfArM) jährlich aktualisiert und herausgegeben [7].

Die ICD-10-GM ist eine monohierarchisch strukturierte, alphanumerische Klassifikation mit bis zu 5 Ebenen und deren Aufbau besteht aus zwei Teilen, das systematisches und alphabetisches Verzeichnis [7].

### 2.1 Alphabetisches Verzeichnis (Alphabet)

Die alphabetische Zuordnung des Codes entsteht aus der gebräuchlich Diagnostexte. Die Systematik enthält nicht alle Diagnosen des Alphabets und eine Bezeichnung dient dazu als Verzugsbezeichnung, sodass das Alphabet auch andere Diagnosenbezeichnungen enthält, die auch veraltete oder ungenau sind [1]. Ein Beispiel davon ist textsfCorona Virus Disease COVID-19 (sehen Sie die Tabelle 1). Aus diesem Grund existiert der Alpha-Identifikator (Alpha-ID). Der nutzt die Bezeichnungen als Basis für einen nicht klassifizierenden Code [1].



Tabelle 1: Verschiedene Bezeichnungen von COVID-19 mit deren Alpha-ID und ICD-10-GM Identifikatoren

| Alpha-ID | ICD-10-GM | Bezeichnung  |
|----------|-----------|--|
| I130805  | U07.1!    | Coronavirus-Infektion-2019, durch Labortest nachgewiesen |
| I130804  | U07.1!    | Coronavirus-Infektion-2019, Virus nachgewiesen           |
| I130797  | U07.1!    | Coronavirus-Krankheit-2019, Virus nachgewiesen           |
| I130809  | U07.1!    | COVID-19-Infektion, durch Labortest nachgewiesen         |
| I130796  | U07.1!    | COVID-19-Infektion, Virus nachgewiesen                   |

## 2.2 Systematisches Verzeichnis (Systematik)

Die Systematik ist eine hierarchisch geordnete Liste der vierstelligen Systematik des Kodes [18, 3]. Die Hierarchieebenen der ICD-10-GM sind Kapitel, Gruppe/Bereiche und Kode, nämlich Kategorie/Dreisteller, Subkategorien/Vier- und Fünfsteller [3]. Für die Durchführung und den Zweck dieses Projekts wird mit dem systematischen Verzeichnis der ICD-10-GM und deren vom BfArM veröffentlichten Dateien und Metadaten gearbeitet.

## 2.3 Metadaten

Das BfArM veröffentlicht zwischen September und Dezember eines Jahres die Metadaten der neuen Fassung der ICD-10-GM in einer Zipper (ZIP)-Datei mit Kodierungen und weiteren Informationen. Der Ordner Klassifikationsdateien enthält die aktuelle Version der Dateien für die Klassifikation der ICD-10-GM (sehen Sie die Tabelle 2). Diese Dateien sind in Comma-Separated Values (CSV)-Format mit „;“ als Trennzeichen.

Die ZIP-Datei enthält auch eine Liesmich-Datei mit der Beschreibung aller CSV-Dateien, vorkommenden Erneuerungen in der Struktur der CSV-Dateien, und Structured Query Language (SQL)-Statements für den Aufbau einer Datenbank oder eines Schemas, je nach welchem Relational Database Management System (RDBMS) benutzt wird [15]. Solche Tabellen sind mit Kopfzeilen in rot in der Abbildung 1 dargestellt. Diese ZIP-Dateien befinden sich auf der Seite Downloads der Klassifikationen der BfArM Webseite in der Sektion ICD-10-GM. Hier befinden sich die Version des laufenden Jahres und die Vorgängerversionen. Die Namen der Links zu den ZIP-Dateien haben die Struktur ICD-10-GM Version JJJJ. Die Struktur der Namen der ZIP-Dateien ist ICD-10-GMVersionJJJJ.zip.

Tabelle 2: Liste der Dateien im Ordner Klassifikationsdateien. „JJJJ“ stellt das Jahr der Fassung der ICD-10-GM dar.

| Dateiname                    | Information                      |
|------------------------------|----------------------------------|
| icd10gmJJJJsysst_gruppen.txt | Kapitel der ICD-10-GM-Systematik |
| icd10gmJJJJsysst_kapitel.txt | Kapitel der ICD-10-GM-Systematik |
| icd10gmJJJJsysst_kodes.txt   | Kodes der ICD-10-GM-Systematik   |
| morbl_JJJJ.txt               | Morbiditätsliste                 |
| mortl1_JJJJ.txt              | Mortalitätsliste 1               |
| mortl1grp_JJJJ.txt           | Gruppen der Mortalitätsliste 1   |
| mortl2_JJJJ.txt              | Mortalitätsliste 2               |
| mortl3_JJJJ.txt              | Mortalitätsliste 3               |
| ortl3grp_JJJJ.txt            | Gruppen der Mortalitätsliste 3   |
| mortl4_JJJJ.txt              | Mortalitätsliste 4               |

## 3 Datenbank Schema

### 3.1 Aufbau der Datenbank

Für die Durchführung dieses Projekts wurde ein relationales DB Schema in PostgreSQL entwickelt. Damit wurden die SQL-Anweisungen, nach Empfehlung in der Liesmich-Dateien der Fassungen, angepasst [15]. Außerhalb von den vorher genannten Anpassungen wurde die Tabelle **kodes** auch modifiziert. In dieser Tabelle wurde die Spalte **ver** für die Speicherung der Auffassungen oder Versionen eingefügt und die Spalte **fünfsteller** wurde zu **fuenfsteller** umbenannt, um Probleme mit Zeichen-Kodierung zu verhindern. Diese Veränderungen sind in schwarz in der Tabelle **kodes** der Abbildung 1 gekennzeichnet. Noch dazu wurden drei neuen Tabelle eingefügt, um die Veröffentlichung Daten (**icd10gm\_release\_info**), Speicherung (**icd10gm**) und Historisierung **icd10gm\_history** aller verfügbaren ICD-10-GM von 2007 bis 2021 zu steuern. Solche Tabellen sind mit Kopfseile in gelb in der Abbildung 1 repräsentiert.

### 3.2 Funktion der Tabelle **kodes** von BfArM

Die Tabellen von **kodes** vom BfArM-Schema beinhalten die Information der CSV-Dateien von nur eine Fassung und enthält damit die Anfangsdaten für die Tabellen **icd10gm** und **icd10gm\_history**.

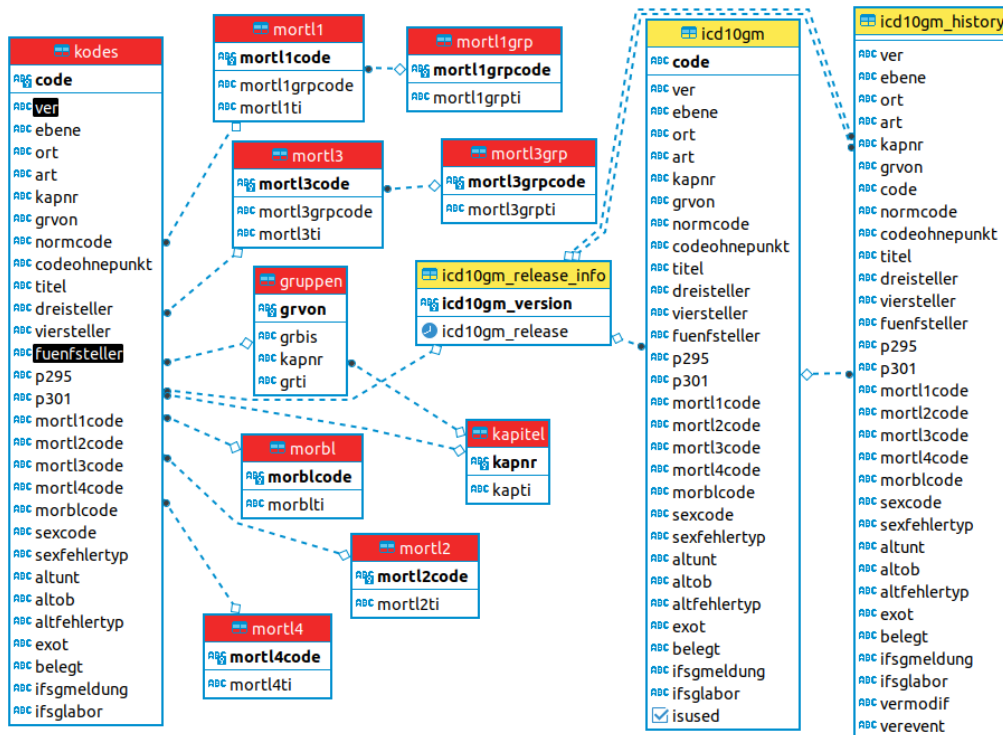


Abbildung 1: Datenbankstruktur für die Steuerung der ICD-10-GM.

### 3.3 Funktion der neuen Tabellen

Für die Erfassung der Datierung der Veröffentlichungen der Versionen wurde die Tabelle `icd10gm_release_info` erstellt. Diese speichert den Identifikator oder das Jahr der Version in der Spalte `icd10gm_version` und das Datum der Veröffentlichung in der Spalte `icd10gm_release`. Diese hat das Format JJJJ-MM-TT. Wobei J das Jahr, M der Monat und T der Tag darstellen. Dadurch wir nur mit der neuesten Veröffentlichung einer Fassung arbeiten, wurde der Identifikator der Version als Hauptschlüssel deklariert.

Die ICD-10-GM von 2007 bis 2021, werden in der Tabelle `icd10gm` gespeichert. In dieser Tabelle wird die Information der noch gültigen ICD-10-GM aktualisiert. Die Spalte `code` dieser Tabelle ist die Hauptschlüssel, genau wie in der Tabelle `kodes` vom BfArM-Schema. `ver` speichert den Identifikator der Version und ist ein Fremd Schlüssel, der zeigt die Spalte `icd10gm_version` der Tabelle `icd10gm_release_info`. Die boolesche Spalte `isused` wird benutzt, um zu markieren welche ICD-10-GM bei anderen Systemen angewendet wird. Auf diesem Grund kann die Tabelle `icd10gm` auch für





Plausibilitätsabfragen benutzt werden.

Die Tabelle `icd10gm_history` enthält die Information der ICD-10-GM, die mit der Laufe der Zeit eingefügt, gelöscht oder geändert wurden. Die Besonderheiten dieser Tabelle sind die Spalten `ver`, `vermodif` und `verevent`. Die Identifikatoren vergangener Versionen einer ICD-10-GM werden in der Spalte `ver` gespeichert. Die Spalte `vermodif` enthält den Identifikator der Fassungen bei deren eine ICD-10-GM eingefügt, gelöscht oder modifiziert wurde. Die Ereignisse einer Insertion, Änderung oder Löschung werden mit den Buchstaben I „insert“, U „update“ und D „delete“ in der Spalte `verevent` kodiert und gespeichert.

### 3.4 Fluss der Information

Mit Hilfe einer ETL-Strecke werden die Daten aus der ZIP-Dateien in der DB importiert. Die Information der Veröffentlichung der ICD-10-GM wird in der Tabelle `icd10gm_release_info` eingefügt. Bei jeder Ladung wird die Information der Tabelle `kodes` vorher gelöscht und mit neuen ICD-10-GM-Datensätzen geladen. Die Codes, die vorher nicht in der Tabelle `icd10gm` vorhanden waren, werden in dieser und in der Tabelle `icd10gm_history` kopiert. Diese Datensätze werden in `icd10gm_history` als eingefügt markiert. Nicht in `kodes` existierende ICD-10-GM werden aus `icd10gm` in `icd10gm_history` kopiert. Diese Kopie enthält auch die Indikatoren der alten und neuen Fassungen und wird als gelöscht markiert. Existierende Kodierungen mit neuer Information werden vorher in der `icd10gm_history` hinzugefügt, danach wird die Information in `icd10gm` aktualisiert. In diesem Fall werden auch die Indikatoren der alten und neuen Fassung registriert und der Datensatz als modifiziert markiert. Dieser Fluss der Information ist im Datenflussdiagramm der Abbildung 2 repräsentiert. Der automatisierte Prozessablauf in der DB ist durch Triggers in den Tabellen `kodes` und `icd10gm` gesteuert.

## 4 ETL-Strecke

Für den Import der Information der ICD-10-GM in der DB wurde eine ETL-Strecke entwickelt. Dazu wurden Bourne-again shell (BASH)-Skripts unter Ubuntu 20.04 programmiert und durchgeführt. Die Anfang Punkt für die Durchführung der ETL sind die ZIP-Dateien der ICD-10-GM Metadaten von 2007 bis 2021 aus der Download Seite vom BfArM für die Klassifikationen. Diese Dateien wurden zuerst manuell heruntergeladen. Der Abruf und die Reihenfolge der Skripts für den Durchlauf der ETL werden von dem BASH-Skript `icd_etl.sh` definiert. Die Abbildung 3 stellt das Flussdiagramm der ETL-Strecke dar.

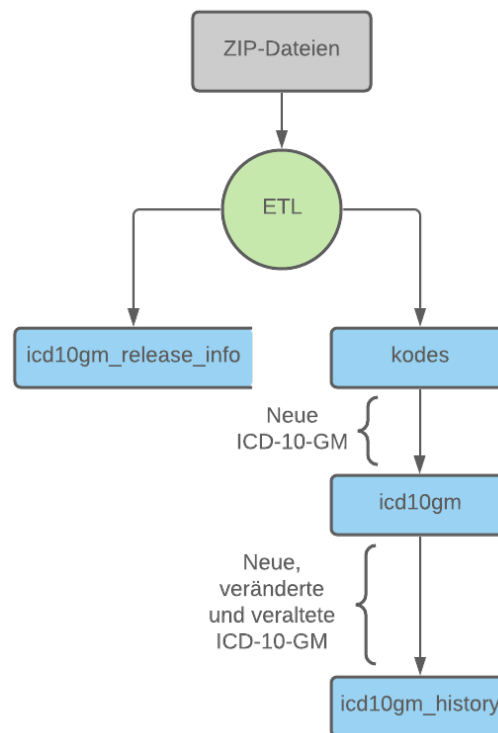


Abbildung 2: Datenflussdiagramm der Information von den ZIP-Dateien bis zum Import in der DB.

## 4.1 Extraktion

Zuerst werden die Ordner der ZIP-Dateien mit Hilfe des Skripts `unzipper.sh` entpackt. Das Skript `copy_codes.sh` generiert einen neuen Ordner für die Speicherung der Code-Dateien. Diese Dateien werden ausgewählt und in diesem Ordner kopiert. Die Information des Datums der Freigabe und Fassung wird mit dem Skript `extra_info.sh` aus der Liesmisch-Dateien extrahiert und in einer CSV-Datei importiert.

## 4.2 Transformation

Die Code-Dateien von 2007 bis 2009 haben den Windows-Standardzeichensatz (ISO-8859-15) als Zeichenkodierung. Das verursacht Probleme bei dem Datenaustausch zwischen Plattformen. Aus diesem Grund wandelt das Skript `iso_2_utf8.sh` das Format dieser Code-Dateien von ISO-8859-15 in UTF-8 um.

Mit der Laufe der Zeit sind neue Spalten in der CSV-Dateien entstanden und andere Felder werden veraltet und entnommen [13, 14]. Deswegen das Skript `se-`

lect.columns.sh, wählt die aktuell benutzte Spalten aus, fügt leere Felder an der Positionen die vorher nicht vorhanden waren ein, fügt ein neues Feld mit der Version an Anfang jede Zeile ein, und importiert der Datensatz einer neuen generierten CSV-Dateien.

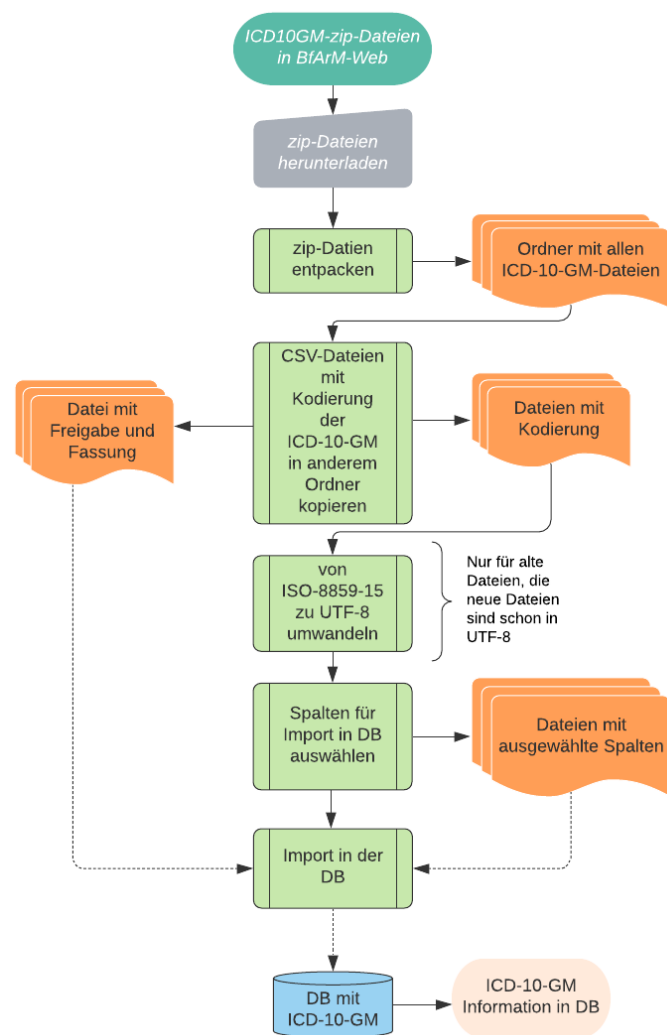


Abbildung 3: Flussdiagramm der ETL-Strecke für den Import der Information der ICD-10-GM aus den ZIP-Dateien in der DB.



## 4.3 Laden

Am Ende des Skript `insert_into_db.sh` importiert die Information der CSV-Dateien in der Tabelle `kodes` der DB.

# 5 Data Analyse

Die Analyse der Information der ICD-10-GM wurde mit Hilfe SQL-Befehle und Python unter dem Framework Jupyter-Notebook durchgeführt.

Wie in der Tabelle 3 dargestellt wird, jede Kode-Datei enthält mehr als 15450 ICD-10-GM, obwohl manche Kodierungen gelöscht werden, nimmt die Menge neuer ICD-10-GM immer zu.

Tabelle 3: Anzahl an ICD-10-GM per Jahr in den CSV-Dateien

| Anzahl        | Fassung       |
|---------------|---------------|
| 15455         | 2007          |
| 15498         | 2008          |
| 15523         | 2009          |
| 15598         | 2010          |
| 15633         | 2011          |
| 15643         | 2012          |
| 15668         | 2013          |
| 15688         | 2014          |
| 15761         | 2015          |
| 15821         | 2016          |
| 15930         | 2017          |
| 16059         | 2018          |
| 16126         | 2019          |
| 16131         | 2020          |
| 16203         | 2021          |
| <i>236737</i> | <b>Gesamt</b> |

Nach dem Durchlauf der ETL sind insgesamt 16520 ICD-10-GM von 2007 bis 2021 in der DB eingefügt. In der Tabelle 4 ist das Verhältnis dieser Kodierungen dargestellt. Wobei die Spalten für den Titel der drei- vier- und fünften Stellen der Codes nicht berücksichtigt wurden, weil diese Felder erst im Jahr 2013 entstanden sind [13]. Ein weiterer Aspekt in Bezug auf die 10180 Änderung ist, dass die Werte verschiedener Felder bei manchen ICD-10-GM in früheren Fassungen noch nicht definiert waren, wie in der Tabelle 6 bei den Fällen der Mortaliätslisten verdeutlicht wird.

Tabelle 4: Charakterisierung der ICD-10-GM in der DB

| Anzahl | Information               |
|--------|---------------------------|
| 16520  | <b>Gesamt</b>             |
| 6188   | Unverändert               |
| 10011  | Nur geändert              |
| 152    | Nur gelöscht              |
| 165    | Geändert und gelöscht     |
| 4      | Wieder verwendet          |
| 10180  | <b>Insgesamt geändert</b> |
| 317    | <b>Insgesamt gelöscht</b> |

## 5.1 Neue ICD-10-GM

Von 2008 bis 2021 sind insgesamt 966 neuen ICD-10-GM entstanden. In der Abbildung 4 ist ein interessanter Aspekt davon repräsentiert, nämlich die Anzahl neuer ICD-10-GM pro Jahr ist unregelmäßig. Es gibt Jahre wie 2017 und 2018 mit mehr als 120 neuen Einträgen. Dieses Phänomen passiert am meisten beim Bedarf neuer Subklassifikationen, um die Diagnosen spezifischer zu kodieren und an diverse klinischen Systeme anzupassen. Ein Beispiel dieses Phänomens ist der Code U81! Bakterien mit Multiresistenz gegen Antibiotika, dieser wurde in 2017 entnommen und stattdessen entstand die Kodierung U81.-! Gramnegative Erreger mit bestimmten Antibiotikaresistenzen, die besondere therapeutische oder hygienische Maßnahmen erfordern zusammen mit 43 weiteren Subklassifikationen von U81.0.-! bis U81.8.-!. Die Ursache davon war eine Umstrukturierung der Bereiche U81!, um diese Codes an die Nomenklatur der Kommission für Krankenhaushygiene und Infektionsprävention (KRINKO) anzupassen [11].

Im Jahr 2018 wurden mehr als 80 neue ICD-10-GM bei M14.-\* Arthropathien bei sonstigen anderenorts klassifizierten Krankheiten eingefügt. Diese Zunahme ist auch in der Abbildung 5 widerspiegelt, denn diese Diagnosen sind im Kapitel Krankheiten des Muskel-Skelett-Systems und des Bindegewebes repräsentiert. Die Ursache dieser Zunahme war die Insertion einer fünften Stelle, um die Abbildung im German Diagnosis Related Groups (G-DRG)-System zu ermöglichen [12].

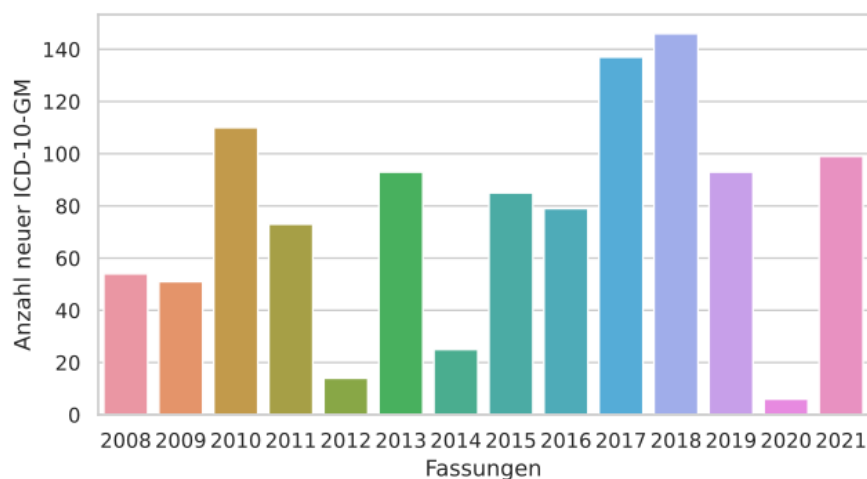


Abbildung 4: Anzahl neuer ICD-10-GM zwischen den Jahren 2008 und 2021

Ein wichtiger und aktueller Punkt sind die meldepflichtigen Krankheiten in der Laufe der Jahre. Dieses Phonemen ist in der Abbildung 6 dargestellt. Es ist zu erkennen, dass sei 2007 nur in den Jahren 2010, 2016 und 2021 wurden neue meldepflichtigen Codes definiert (Tabelle 5). Ursachen davon sind Pandemien und Epidemien wie die Influenza Varianten zwischen 2009 und 2010 [2, 16], die Verbreitung des Dengue Fiebers in Europa als Effekt der Globalisierung mit der steigenden Mobilität [17] und Verbreitung der asiatischen Tigermücke *Aedes (Stegomyia) albopictus* zwischen 2015 und 2016 in der Region als Konsequenz der milden Winter [19], und noch aktuell seit Februar 2020 die Verbreitung des Corona Virus in Europa [5] und deren gesundheitlichen Folgen [4].

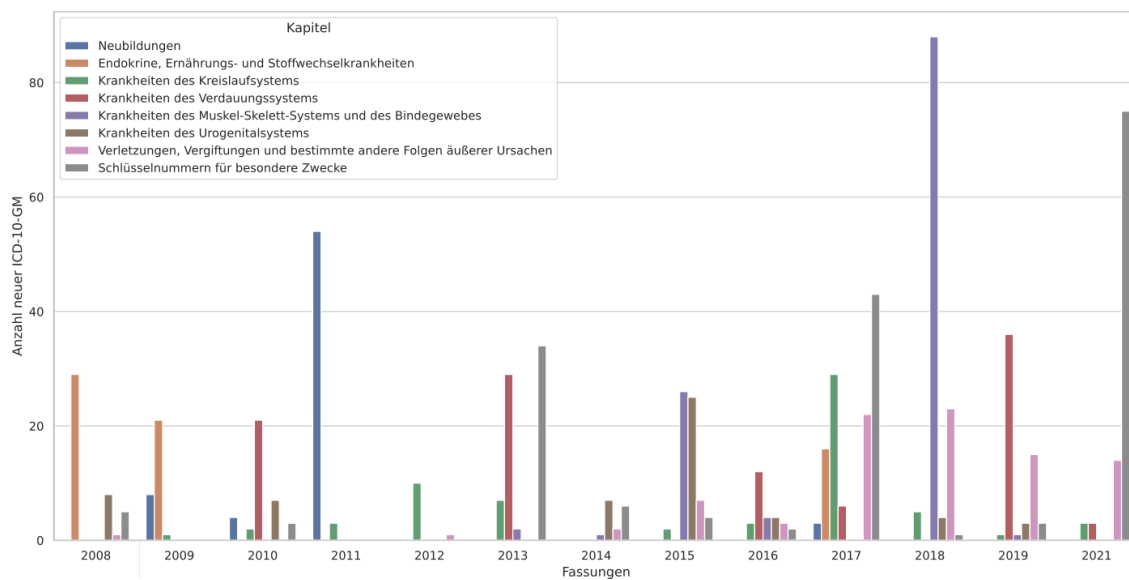


Abbildung 5: Kapitel der neuen ICD-10-GM zwischen den Jahren 2008 und 2021

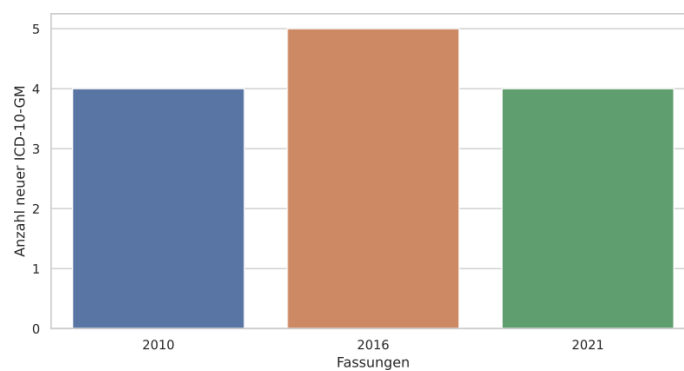


Abbildung 6: Meldepflichtige Krankheiten zwischen den Jahren 2008 und 2021

Tabelle 5: Meldepflichtige Krankheiten

| Fassung | ICD-10-GM | Titel   |
|---------|-----------|---|
| 2010    | B17.9     | Akute Virushepatitis, nicht näher bezeichnet  |
| 2010    | U69.2-!   | Sekundäre Schlüsselnummern für besondere epidemiologische Zwecke                        |
| 2010    | U69.20!   | Influenza A/H1N1 Pandemie 2009 [Schweinegrippe]   |
| 2010    | U69.21!   | Influenza A/H5N1 Epidemie [Vogelgrippe]   |
| 2016    | A97.-     | Dengue  |
| 2016    | A97.0     | Dengue ohne Warnzeichen   |
| 2016    | A97.1     | Dengue mit Warnzeichen  |
| 2016    | A97.2     | Schweres Dengue   |
| 2016    | A97.9     | Dengue, nicht näher bezeichnet  |
| 2021    | U07.1!    | COVID-19, Virus nachgewiesen  |
| 2021    | U07.2!    | COVID-19, Virus nicht nachgewiesen  |
| 2021    | U10.-     | Multisystemisches Entzündungssyndrom in Verbindung mit COVID-19                         |
| 2021    | U10.9     | Multisystemisches Entzündungssyndrom in Verbindung mit COVID-19, nicht näher bezeichnet |

## 5.2 Modifizierte ICD-10-GM

Jedes Jahr werden Änderungen in der Metadaten der ICD-10-GM vorgenommen. Solche Modifikationen sind in der Tabelle 6 aufgelistet. In dieser Tabelle ist es zu erkennen, dass die meisten Änderungen in Bezug zur Mortalitätsliste entstanden sind, nämlich 6949 ICD-10-GM bei dem Bezug zur Mortalitätsliste 3 und 6348 Kodierungen bei dem Bezug zur Mortalitätsliste 1. Die Ursache davon ist, dass die Werte solche Listen bei der meisten Codes im Jahr 2007 nicht definiert waren, und mit dem Wert UNDEF gesetzt. Erst in 2008 wurden diese Felder durch einen spezifischen Wert ersetzt. Die zahlreiche Modifikationen an der Klassentitel entstehen durch Vorschläge zur Weiterentwicklung der Klassifikationen. Ein Beispiel davon ist die Schlüsselnummer E10.31 mit dem Titel Primär insulinabhängiger Diabetes mellitus [Typ-1-Diabetes] mit Augenkomplikationen: Als entgleist bezeichnet. Dieser Titel wurde zu Primär insulinabhängiger Diabetes mellitus [Typ-1-Diabetes] : Mit Augenkomplikationen: Als entgleist bezeichnet in 2009 geändert. Die Änderung wurde in diesem Jahr durchgeführt, da eine fünfte Stelle in der Klassifikationen von E10 bis E14 eingefügt wurde [8]. Der Titel dieser ICD-10-GM wurde nochmal in 2014 zu Diabetes mellitus, Typ 1: Mit Augenkomplikationen: Als entgleist bezeichnet ersetzt. In diesem Jahr wurden die Titel von der Weltgesundheitsorganisation (WHO) an die gebräuchliche Terminologie angepasst [10].



Tabelle 6: Liste deDr Metadaten und Anzahl an Änderungen

| Änderungen | Metadaten   |
|------------|---|
| 6949       | <b>Bezug zur Mortalitätsliste 3</b>   |
| 5          | Bezug zur Mortalitätsliste 3 (davon definiert)                              |
| 6348       | <b>Bezug zur Mortalitätsliste 1</b>   |
| 2          | Bezug zur Mortalitätsliste 1 (davon definiert)                              |
| 1111       | Klassentitel  |
| 308        | <b>Untere Altersgrenze</b>  |
| 42         | Untere Altersgrenze (davon relevant)  |
| 299        | <b>Bezug zur Morbiditätsliste</b>   |
| 11         | Bezug zur Morbiditätsliste (davon definiert)                                |
| 286        | <b>Obere Altersgrenze</b>   |
| 19         | Obere Altersgrenze (davon relevant)   |
| 162        | Auswahl der Laborauschlussziffer des einheitlichen Bewertungsmaßstabs (EBM) |
| 118        | Art der Vier- und Fünfsteller   |
| 69         | <b>Bezug zur Mortalitätsliste 4</b>   |
| 0          | Bezug zur Mortalitätsliste 4 (davon definiert)                              |
| 68         | Geschlechtsbezug  |
| 64         | <b>Bezug zur Mortalitätsliste 2</b>   |
| 2          | Bezug zur Mortalitätsliste 2 (davon definiert)                              |
| 44         | Erster Dreisteller der Gruppe   |
| 40         | Meldepflicht  |
| 31         | Sehr seltene Krankheit in Mitteleuropa                                      |
| 6          | Belegte Nummer  |
| 3          | Paragraph 295   |
| 2          | Klassifikationsebene  |

Ein interessantes Aspekt ist, dass 52 Diagnosen mit Geschlechtsbezug der 68, Krankheiten der Brustdrüse sind. Diese ICD-10-GM waren frühe auf das weibliche Geschlecht bezogen, jetzt ist der Geschlechtsbezug diese Kodierungen irrelevant. Viele dieser Änderungen wurden in den Jahren 2008 und 2016 vorgenommen. An dieser Stelle können wir nicht erklären, wieso diese Art von Krankheiten nicht immer ohne Geschlechtsbezug waren, denn seit 1930 existieren medizinische Publikationen über Brustdrüse Anomalien bei Männer [6].

In Bezug auf die sehr seltene Krankheiten in Mitteleuropa sind 20 von den 31 belegte Schlüsselnummer. Davon gehören 14 zu den Bereichen A36.- Diphtherie und A80.- Akute Poliomyelitis [Spinale Kinderlähmung]. Diese sind in der Tabelle 7 aufgelistet. Die



Tabelle 7: ICD-10-GM der Bereiche A36.- Diphtherie und A80.- Akute Poliomyelitis [Spinale Kinderlähmung]. J Ja, N Nein.

| Änderung | ICD-10-GM | Titel   | Alt | Neue |
|----------|-----------|---|-----|------|
| 2011     | A36.-     | Diphtherie  | N   | J    |
| 2011     | A36.0     | Rachendiphtherie  | N   | J    |
| 2011     | A36.1     | Nasenrachendiphtherie   | N   | J    |
| 2011     | A36.2     | Kehlkopfdiphtherie  | N   | J    |
| 2011     | A36.3     | Hautdiphtherie  | N   | J    |
| 2011     | A36.8     | Sonstige Diphtherie   | N   | J    |
| 2011     | A36.9     | Diphtherie, nicht näher bezeichnet                                    | N   | J    |
| 2011     | A80.-     | Akute Poliomyelitis [Spinale Kinderlähmung]                           | N   | J    |
| 2011     | A80.0     | Akute paralytische Poliomyelitis durch Impfvirus                      | N   | J    |
| 2011     | A80.1     | Akute paralytische Poliomyelitis durch importiertes Wildvirus         | N   | J    |
| 2011     | A80.2     | Akute paralytische Poliomyelitis durch einheimisches Wildvirus        | N   | J    |
| 2011     | A80.3     | Sonstige und nicht näher bezeichnete akute paralytische Poliomyelitis | N   | J    |
| 2011     | A80.4     | Akute nichtparalytische Poliomyelitis                                 | N   | J    |
| 2011     | A80.9     | Akute Poliomyelitis, nicht näher bezeichnet                           | N   | J    |

Ursachen der Änderungen sind die erfolgreiche Impfkampagne in Deutschland, sodass die vom Robert-Koch-Institut (RKI) jährlich selten gemeldeten Fällen importiert sind [20].

### 5.3 Gelöschte ICD-10-GM

In jeder Fassung von 2008 bis 2021 ICD-10-GM werden gelöscht. Dieses Verhalten ist in der Abbildung 7 repräsentiert. Die meiste davon im Jahr 2013.

Die Kapitel, die in diesem Jahr, am meisten getroffen wurden, sind in der Abbildung 8 dargestellt. In diesem Jahr wurde eine umfangreiche Umstrukturierung in der Semantik vorgenommen, bei denen bestimmte Codebereiche im Kapitel 09 Krankheiten des Kreislaufsystems umfänglich überarbeitet wurden, viele fünfstellige Klassifizierungen wurden von Kapitel 13 Krankheiten des Muskel-Skelett-Systems und des Bindegewebes entnommen, und im Kapitel 22 Schlüsselnummern für besondere Zwecke wurden zahlreiche ICD-10-GM umgebaut [9].

Von den 317 gelöschten Schlüsselnummer 126 wurden erweitert, wie in Sektion 5.1 beschrieben wurde,

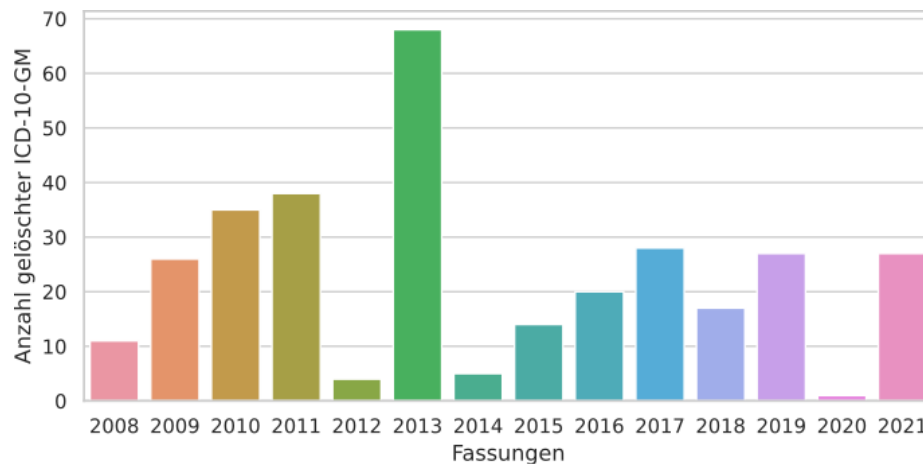


Abbildung 7: Anzahl gelöschter ICD-10-GM zwischen den Jahren 2008 und 2021

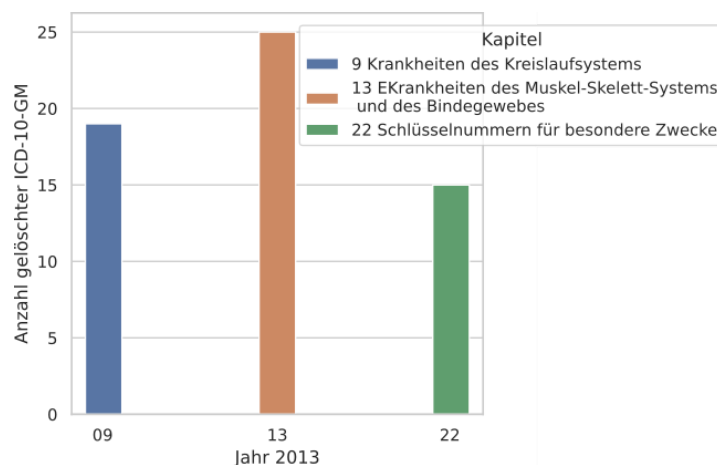


Abbildung 8: Meist betroffene Kapitel in 2013

Die Abbildung 9 zeigt, dass die meiste Codes, nämlich 295 der 327 gelöschten Kodierungen, terminale Schlüsselnummer waren, also ICD-10-GM ohne Subkategorien. Davon wurden 115 erweitert wie in der Sektion 5.1 erläutert wurde. Von der nicht terminale Kodierungen wurden dann 11 erweitert. Ein Beispiel davon ist **Z75.2- Wartezeit auf eine Untersuchung oder Behandlung**, diese nicht terminale Kodierung wurde gelöscht und zu der terminalen Code **Z75.2** mit demselben Titel umgewandelt. Der Rest der ICD-10-GM wurden einfach nicht mehr weiter benutzt.

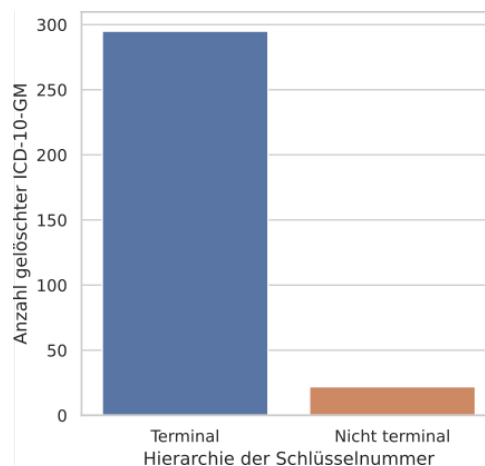


Abbildung 9: Hierarchie der gelöschte ICD-10-GM

## 5.4 Wiederverwendete ICD-10-GM

4 Schlüsselnummer wurden wiederverwendet  
...tabelle...

## 5.5 Strukturelle Änderungen

Wie in der Subsektion 4.2 genannt wurde, es gab auch Änderungen in der Struktur des semantischen Verzeichnisses. Im Jahr 2013 wurde die Spalte **titel** für den Klassentitel der Tabelle **kodes** aus dem Dreisteller-, Viersteller- und gegebenenfalls Fünfstellertitel zusammengesetzt. Auch in diesem Jahr wurden drei neue Felder für die einzelnen Bestandteile des zusammengesetzten Klassentitels eingefügt [13]. Eine weitere Änderung in der Tabelle **kodes**, die schon sei 2015 geplant wurde, und im Jahr 2018 vorgenommen wurde, war die Entnahme der Felder der Altersklassenformaten für die untere und obere Grenze des Altenbezuges mit Angabe von Tagen, Wochen, Monaten und Jahren; und damit blieben nur die Felder mit dem Format in Tagen und Jahren, nämlich **altunt** für die untere Grenze und **altob** für die obere Grenze des Altenbezuges [14].



## 6 Schlussfolgerung

Mit der Durchführung dieses Projekts konnten wir die Besonderheiten eines Lebenszyklus der historischen ICD-10-GM Auffassungen von 2007 bis 2021 darstellen. Damit wurde eine DB-Schema zur Speicherung der Daten zusammen mit eine ETL-Strecke zum Import der Information in der DB programmiert. Am Ende der Implementierung konnten die Besonderheiten der ICD-10-GM in dem genannten Zeitraum, nämlich Insertion, Modifikation und Löschung in Python analysiert werden. Damit wurden verschiedene Ursachen für die Änderungen in der Auffassungen erkannt und detailliert analysiert.



## Literatur

- [1] *Alphabetisches Verzeichnis zur ICD-10-GM*. Zugriff am 22-09-2021. Bundesministeriums für Gesundheit (BMG). <https://www.dimdi.de/dynamic/de/klassifikationen/icd/icd-10-gm/alphabet/>.
- [2] A Amato-Gauci, P Zucs, R Snacken, B Ciancio, V Lopez, E Broberg, P Penttinen, A Nicoll und collective on behalf of the European Influenza Surveillance Network (EISN). „Surveillance trends of the 2009 influenza A(H1N1) pandemic in Europe“. In: *Eurosurveillance* 16.26, 19903 (2011). DOI: <https://doi.org/10.2807/ese.16.26.19903-en>. <https://www.eurosurveillance.org/content/10.2807/ese.16.26.19903-en>.
- [3] *Aufbau der Vierstelligen ausführlichen Systematik der ICD-10-GM*. Zugriff am 30-09-2021. Bundesministeriums für Gesundheit (BMG). <https://www.dimdi.de/dynamic/de/klassifikationen/icd/icd-10-gm/systematik/systematik/>.
- [4] Abdul Mannan Baig. „Chronic COVID syndrome: Need for an appropriate medical terminology for long-COVID and COVID long-haulers“. In: *Journal of Medical Virology* 93.5 (2021), S. 2555–2556. DOI: <https://doi.org/10.1002/jmv.26624>. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/jmv.26624>. <https://onlinelibrary.wiley.com/doi/abs/10.1002/jmv.26624>.
- [5] Feng He, Yu Deng und Weina Li. „Coronavirus disease 2019: What we know?“. In: *Journal of Medical Virology* 92.7 (2020), S. 719–725. DOI: <https://doi.org/10.1002/jmv.25766>. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/jmv.25766>. <https://onlinelibrary.wiley.com/doi/abs/10.1002/jmv.25766>.
- [6] Mohd. Ajmal Husain. „Scirrhous Cancer of the Breast in the Male“. In: *The Indian Medical Gazette* 65.6 (Juni 1930), S. 330. ISSN: 0019-5863. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5157223/> (besucht am 09.09.2021).
- [7] *ICD-10-GM*. Zugriff am 22-09-2021. Bundesministeriums für Gesundheit (BMG), 2021. <https://www.dimdi.de/dynamic/de/klassifikationen/icd/icd-10-gm/>.
- [8] *Kommentar zur ICD-10-GM Version 2009*. Bundesministeriums für Gesundheit (BMG), Sep. 2008.
- [9] *Kommentar zur ICD-10-GM Version 2013*. Bundesministeriums für Gesundheit (BMG), Sep. 2012.
- [10] *Kommentar zur ICD-10-GM Version 2014*. Bundesministeriums für Gesundheit (BMG), Sep. 2013.

- [11] *Kommentar zur ICD-10-GM Version 2017*. Bundesministeriums für Gesundheit (BMG), Sep. 2016.
- [12] *Kommentar zur ICD-10-GM Version 2018*. Bundesministeriums für Gesundheit (BMG), Sep. 2017.
- [13] *LIESMICHDATEI*. Bundesministeriums für Gesundheit (BMG), Sep. 2012.
- [14] *LIESMICHDATEI*. Bundesministeriums für Gesundheit (BMG), Sep. 2016.
- [15] *LIESMICHDATEI*. Bundesministeriums für Gesundheit (BMG), Nov. 2020.
- [16] Robyn Martin, Alexandra Conseil, Abie Longstaff, Jimmy Kodo, Joachim Siegert, Anne-Marie Duguet, Paula Lobato de Faria, George Haringhuizen, Jaime Espin und Richard Coker. „Pandemic influenza control in Europe and the constraints resulting from incoherent public health laws“. In: *BMC Public Health* 10.1 (Sep. 2010), S. 532. ISSN: 1471-2458. DOI: 10.1186/1471-2458-10-532. <https://doi.org/10.1186/1471-2458-10-532>.
- [17] Donald Salami, César Capinha, Maria do Rosário Oliveira Martins und Carla Alexandra Sousa. „Dengue importation into Europe: A network connectivity-based approach“. In: *PLOS ONE* 15.3 (März 2020), S. 1–19. DOI: 10.1371/journal.pone.0230274. <https://doi.org/10.1371/journal.pone.0230274>.
- [18] *Systematisches Verzeichnis der ICD-10-GM*. Zugriff am 22-09-2021. Bundesministeriums für Gesundheit (BMG). <https://www.dimdi.de/dynamic/de/klassifikationen/icd/icd-10-gm/systematik/>.
- [19] D. Walther, D.E. Scheuch und H. Kampen. „The invasive Asian tiger mosquito *Aedes albopictus* (Diptera: Culicidae) in Germany: Local reproduction and overwintering“. In: *Acta Tropica* 166 (2017), S. 186–192. ISSN: 0001-706X. DOI: <https://doi.org/10.1016/j.actatropica.2016.11.024>. <https://www.sciencedirect.com/science/article/pii/S0001706X16305472>.
- [20] Miriam Wiese-Posselt, Christine Tertilt und Fred Zepp. „Vaccination Recommendations for Germany“. In: *Dtsch Arztebl International* 108.45 (2011), S. 771–780. DOI: 10.3238/arztebl.2011.0771. eprint: <https://www.aerzteblatt.de/pdf.asp?id=112587>. <https://www.aerzteblatt.de/int/article.asp?id=112587>.
- [21] Marcus Will. persönliches Gespräch. 2021.