

FDPG Cohort Explorer – Eine Shiny Web App zur Visualisierung von FDPG-Kohorten

Datenintegrationszentrum Cottbus
Dr. Franziska Bathelt, f.bathelt@mul-ct.de
Ahmad Albenny, a.albenny@mul-ct.de
Dennis Hübner, d.huebner@mul-ct.de

Datenintegration & -standardisierung



Christian Papritz



GB IT



Labor

Forschungsinfrastruktur, Anwendungen & Qualitätssicherung

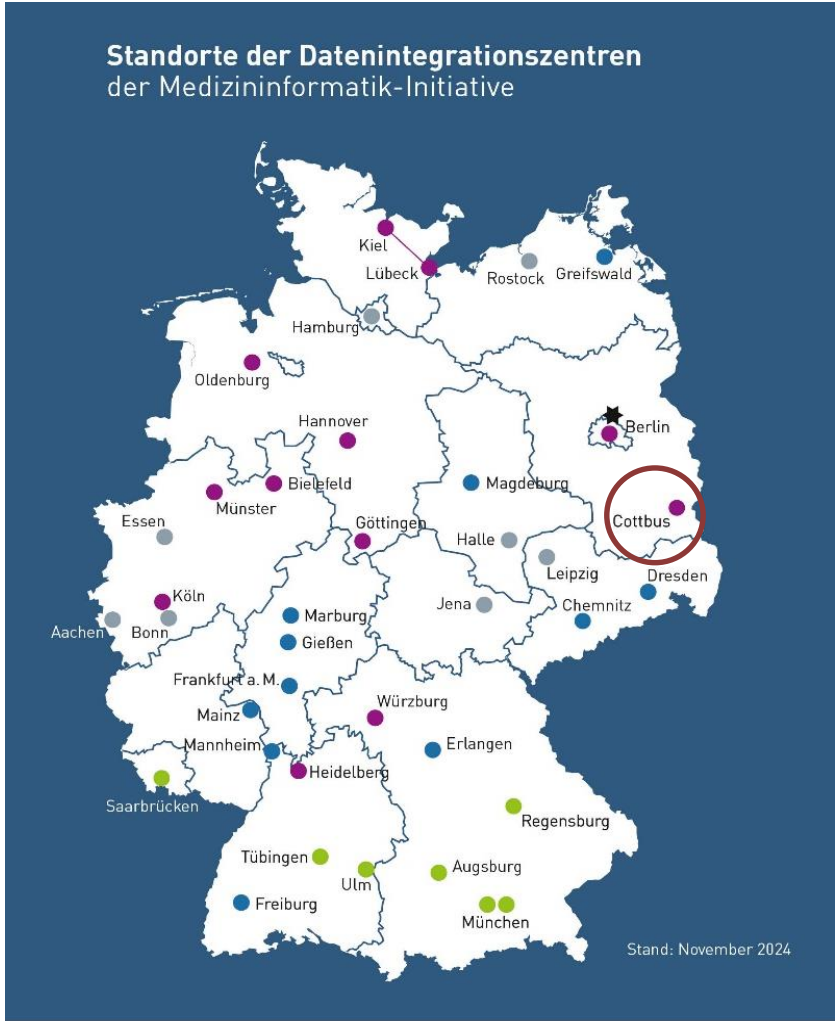


Regulatorik & Datenschutz



Interdisziplinäre Studienzentrale

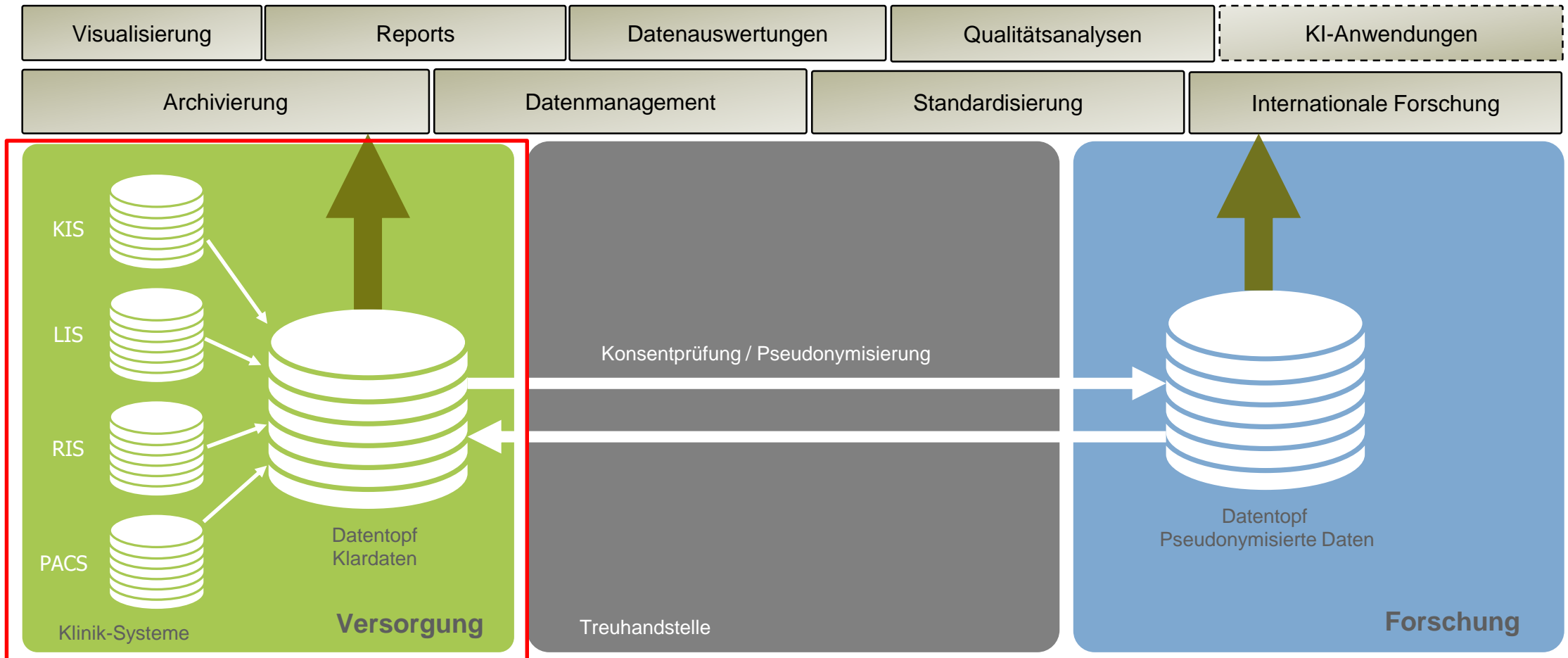
Use Cases



DIFUTURE
HiGHmed
MIRACUM
SMITH
Koordinationsstelle: TMF/MFT/VUD

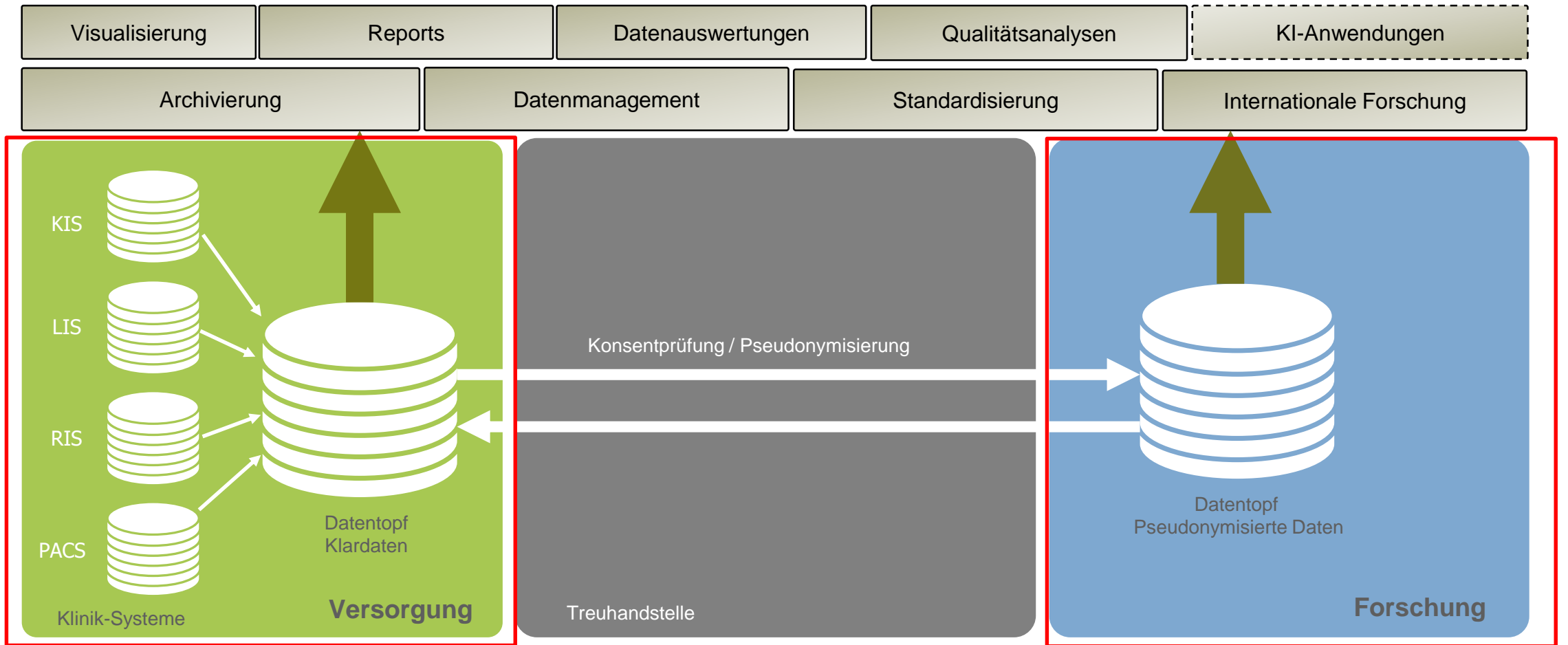
Zielstellung

Mehrwertgenerierung durch Erschließung, Harmonisierung, Standardisierung, Qualitätssicherung und Bereitstellung von **Routinedaten** zu Forschungs- und Versorgungszwecken



Zielstellung

Mehrwertgenerierung durch Erschließung, Harmonisierung, Standardisierung, Qualitätssicherung und Bereitstellung von **Routinedaten** zu Forschungs- und Versorgungszwecken



Kohortenexplorer



- Daten Definition
- Kohorten Selektion**
- Gespeichert Abfragen
- Einstellungen

Machbarkeitsabfrage wird b

Die Anzahl an Patientinnen und Pa
exakt.

0 0 0 0 0 0

Insights

44

Einschlusskriterien

Diabetes mellitus, Typ 1: Mit Koma
Diagnose, E10.0, ICD10

FDPG-Cohort-Explorer

General Exploratory data analysis

Medizinische Universität
Lausitz – Carl Thiem

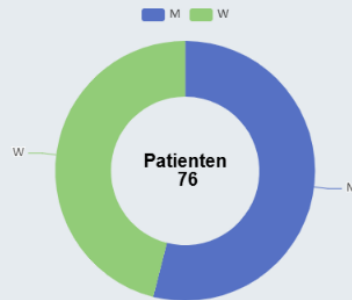
Choose .json File

Browse...

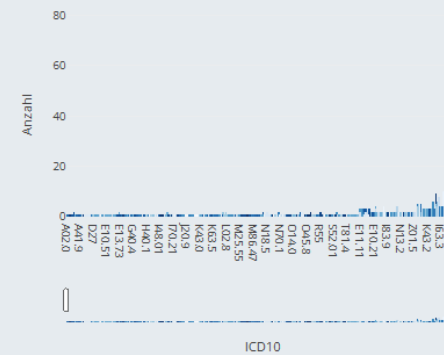
No file

Process Data

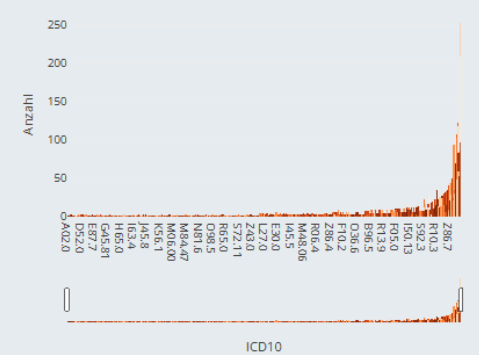
Verteilung nach Geschlecht



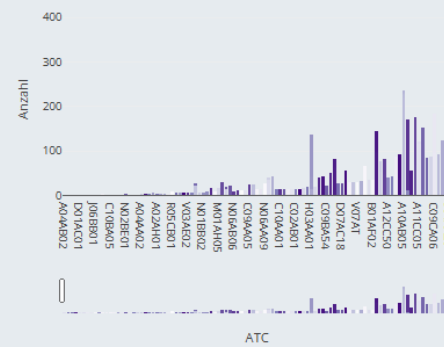
Hauptdiagnosen



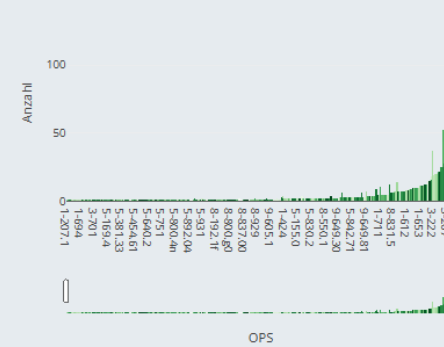
Nebendiagnosen



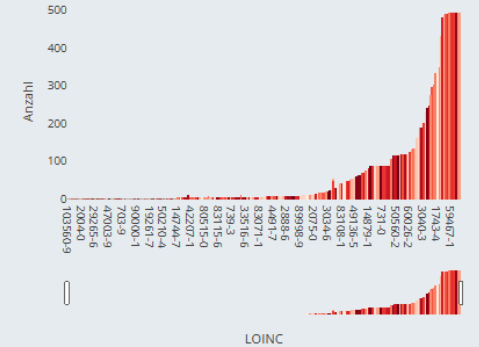
Medikation



Prozeduren



Laboruntersuchungen



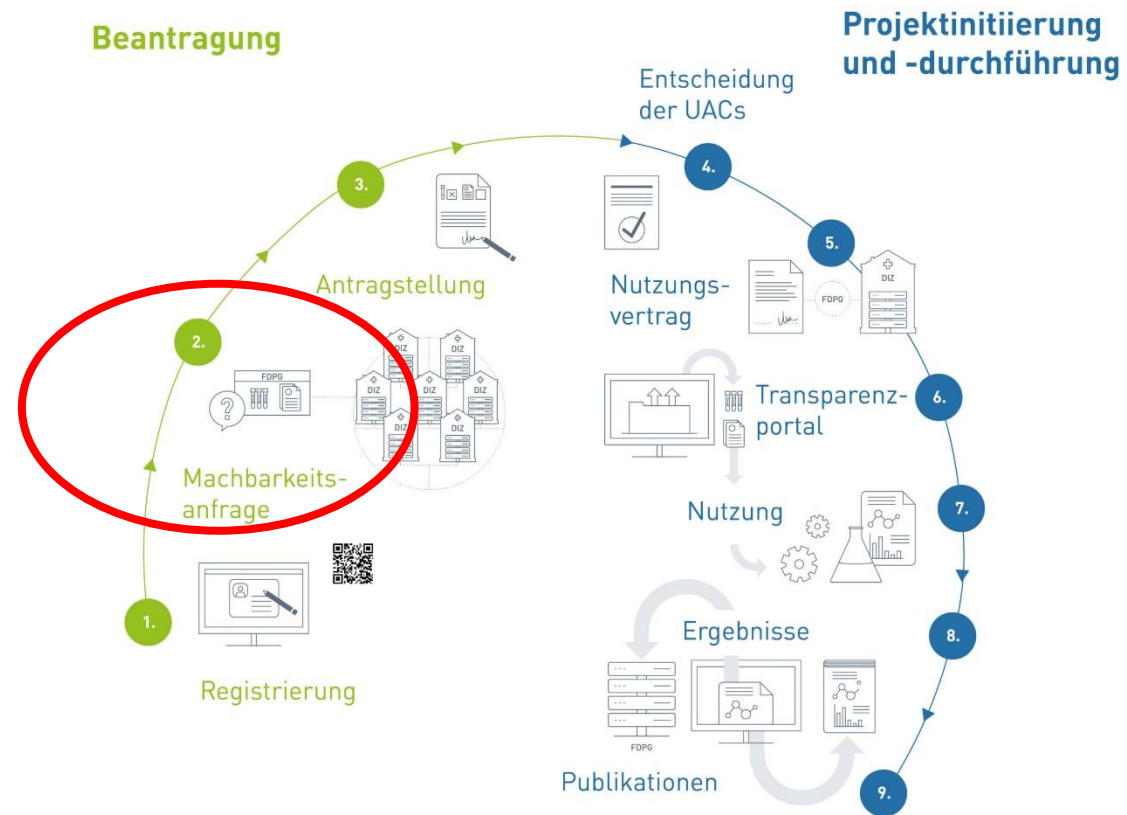


- Lernziele:
 - Nutzung von verfügbarer Software zur Visualisierung von Kohorten
 - Vergleich verschiedener Ansätze zum Extrahieren der Kohorten Daten
 - Geeignete Visualisierungen von Kohorten
 - User Experience des FDPG Cohort Explorers

Datenverarbeitung

Ahmad Albenny

Visualisierung der Kohorte Datenverarbeitung



[1]

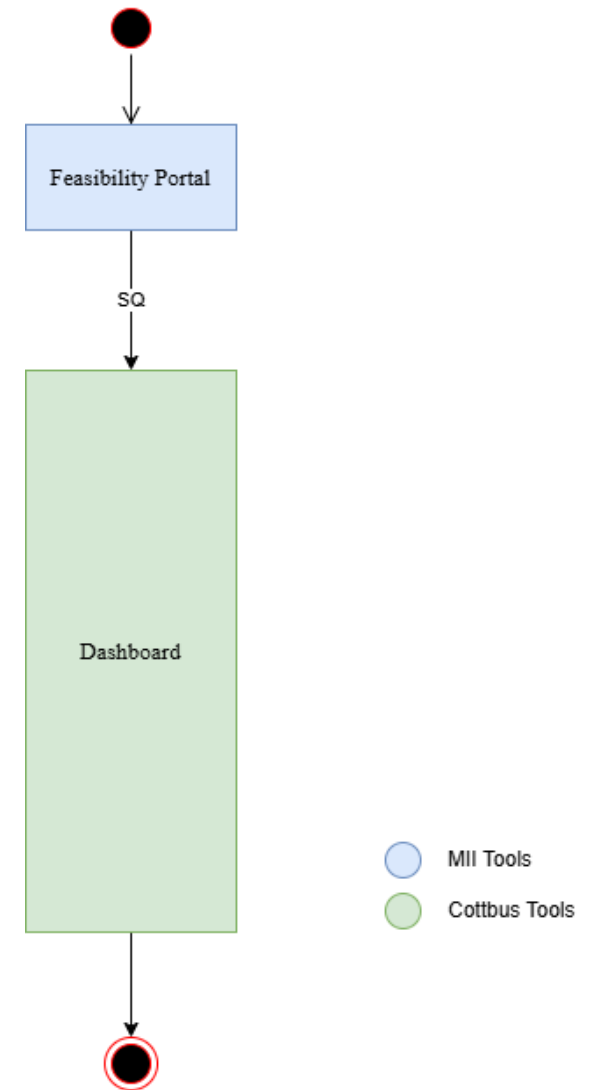
Einführung

- Die Machbarkeitsanfrage liefert die Anzahl der Patienten an, die die Ein- und Ausschlusskriterien der Studie erfüllen.
- Die Anzahl allein kann jedoch mehrere wichtige Inhalte haben.
 - die viel Zeit und Mühe sparen könnten.
 - vielleicht auch die Richtung der Forschung ändern könnten.
- Damit das Gefühl, mit dieser Anzahl zu verbessern und die Inhalte verständlicher, übersichtlicher und interpretierbarer zu machen, ist Datenvisualisierung der richtige Weg, um dies zu erreichen.
- Die Frage: Ist eine Visualisierung der FDPG-Kohorte möglich?
- Wenn Ja, Wie?...



Methode

- Ja, das ist möglich, aber nur, wenn die Daten der Patienten aus der FDPG-Kohorte extrahiert werden können.
- Wir brauchen
 - Datenbank, die die Patientendaten hat.
 - die Suchkriterien, die der Forscher in FDPG ausgewählt hat.
- Um das zu erreichen, ist der Ausgangspunkt „structured query“, die Ein- bzw. Ausschlusskriterien (Suchkriterien) beschreibt.
- Ansatz: Nach der Erstellung der Machbarkeitsanfrage wird die SQ von FDPG exportiert und dann in einem Tool importiert, um die Kohorte zu visualisieren.



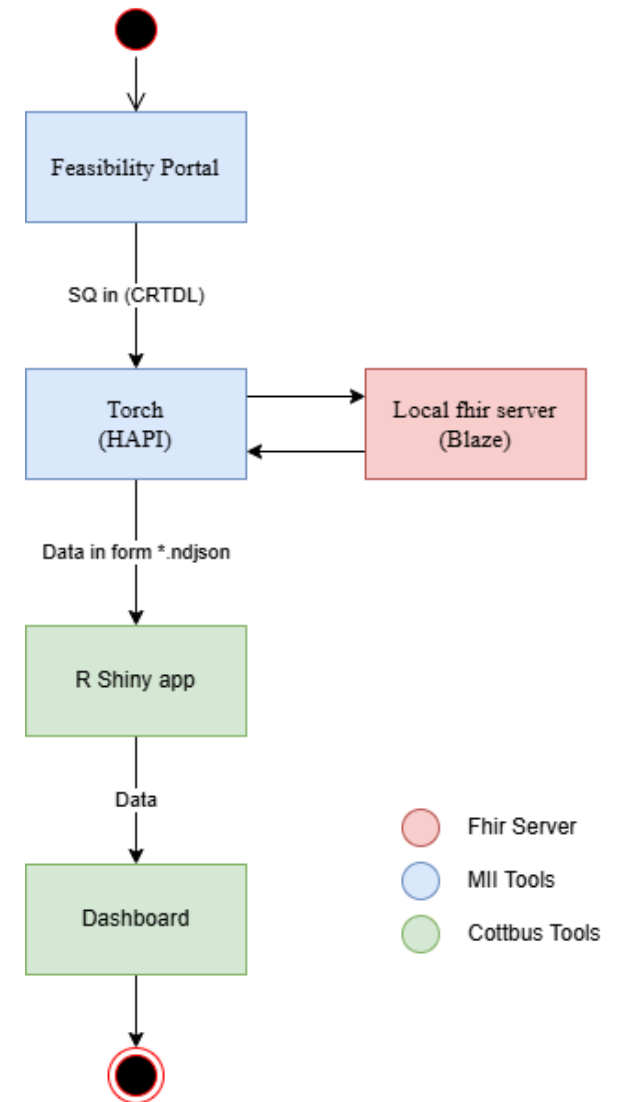
Methode – Erster Ansatz

Das erste Szenario fordert:

- Torch Tool [4].
- Lokales Fhir Server.
- Visualisierungstool.

Prozessablauf:

- SQ exportieren.
- SQ an Torch senden.
- Extraktion der Daten.
- Datenlesen von *.ndjson file (R shiny app).
- Visualisierung.



Methode – Erster Ansatz

Vorteile:

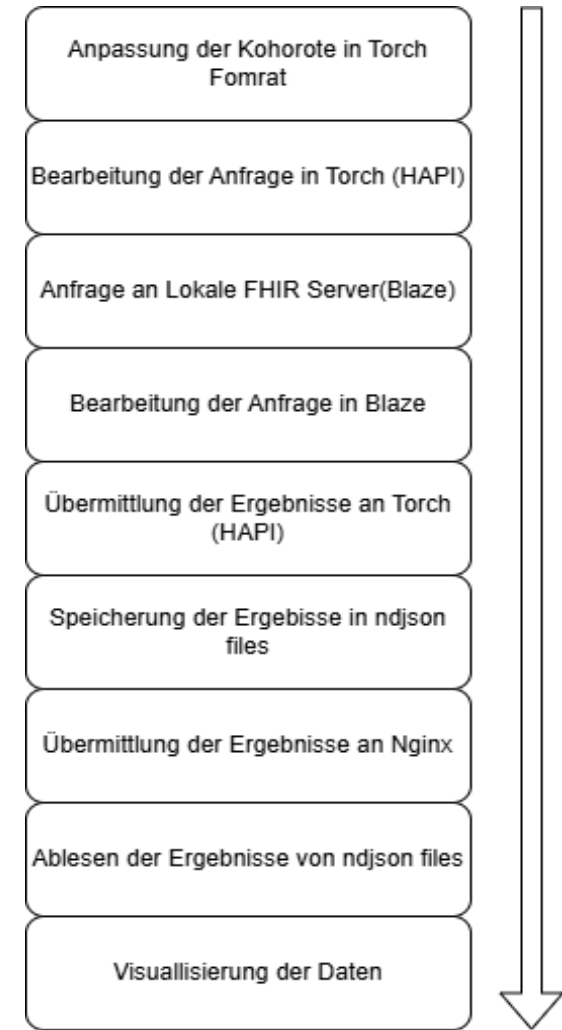
- Simple/Einfach. Benötigt nur mii Tools für Datenextraktion

Nachteile:

- Viele Duplikate.
- Die Datenextraktion aus ndjson-Dateien ist sehr aufwendig.
- Viele I/O-Operationen (read/write) können eingespart werden.
- Die Kommunikation zwischen Torch (HAPI) und dem lokalen FHIR-Server (Blaze) für die Datenextraktion ist aufwendig.

=> sehr zeit- und ressourcenintensiv.

Aus diesem Grund war es notwendig, eine effiziente Lösung zu finden, um Zeit und Ressourcen zu sparen.



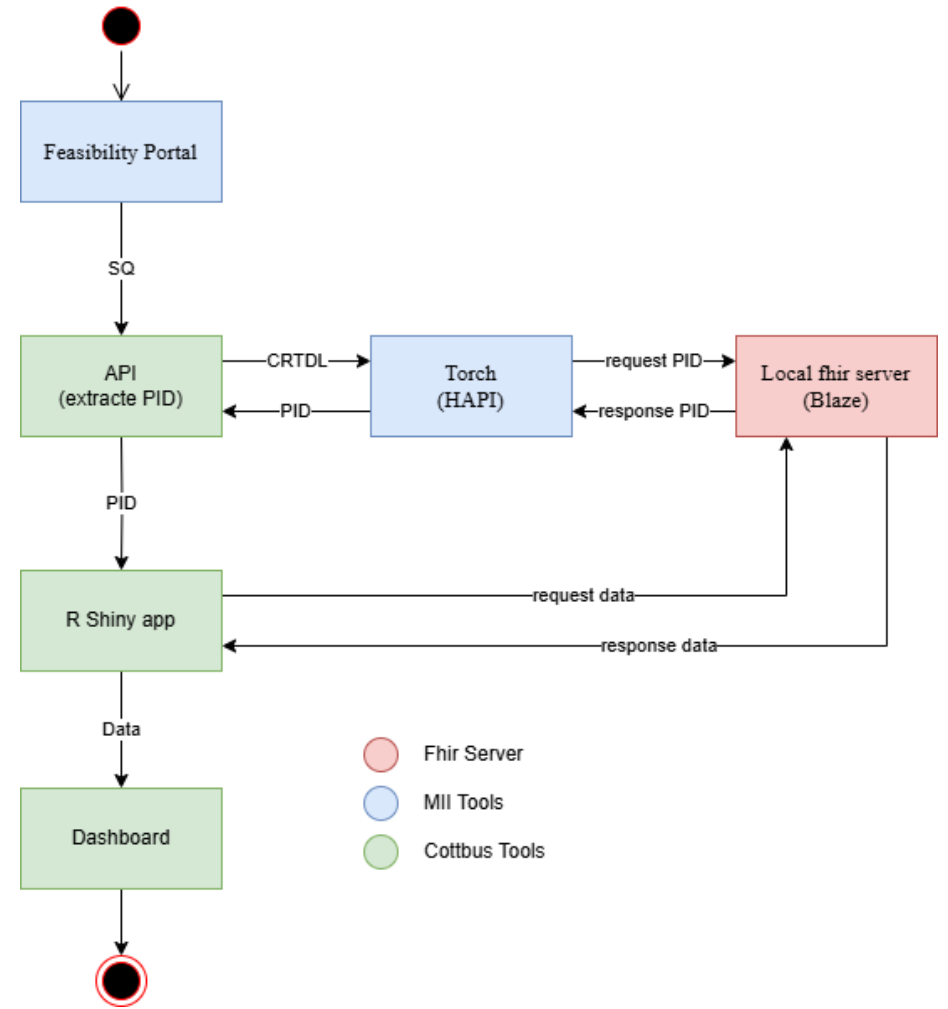
Methode - Zweiter Ansatz

Das zweite Szenario fordert:

- Torch Tool [4].
- Lokales Fhir Server.
- API (zur Extraktion der PIDs).
- Visualisierungstool.

Prozessablauf:

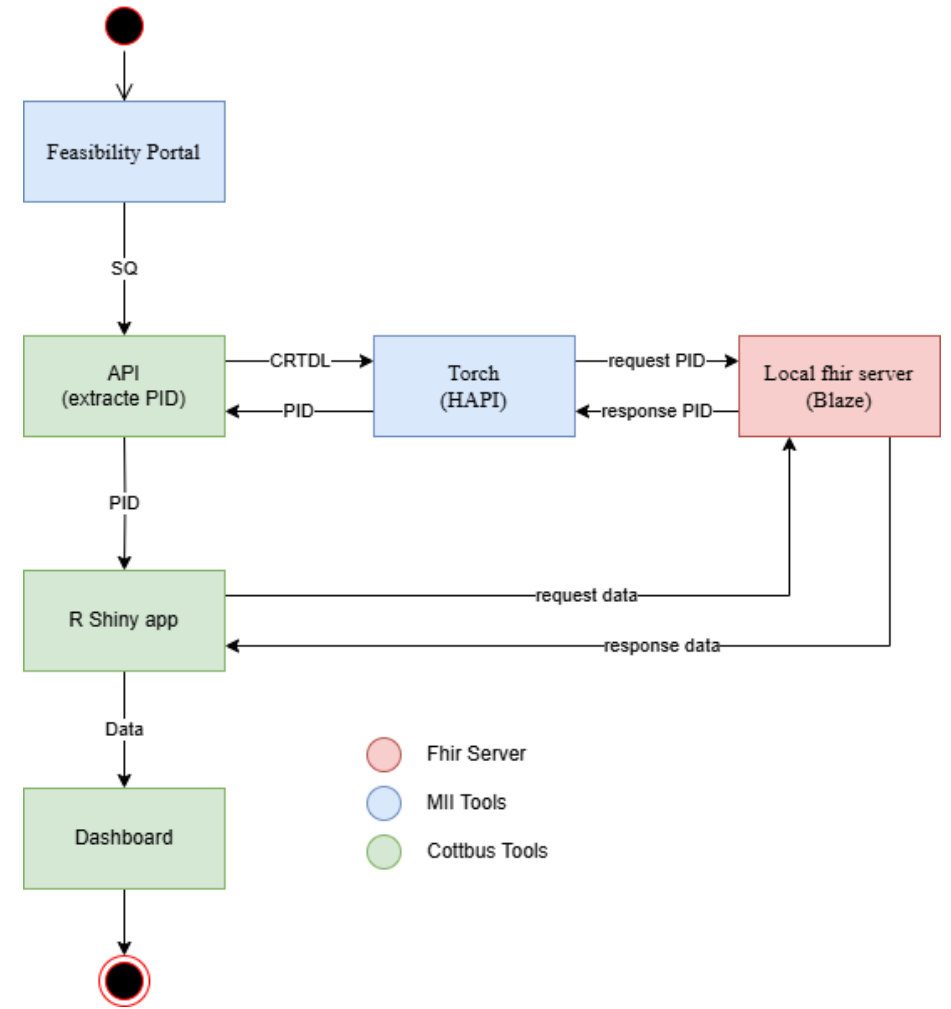
- SQ exportieren.
- SQ an API senden, um nur PIDs zu extrahieren.
- API kommuniziert mit torch, welches PID von Blaze abholt.
- API liefert PIDs an R shiny app.
- Visualisierung.



Methode - Zweiter Ansatz

Vorteile:

- Ansatz anwendbar überall.
- Mit PID-Extraktion können viele Attribute später abfragt werden.
- Die Daten Extraktion in Torch funktioniert auf minimale Anforderung, wird nur PID extrahiert -> Sparen wir Bearbeitungszeit (Schreiben/Lesen) + Speicherung der Daten in mehreren ndjson-Dateien.
- Die Suche für gewünschte Daten ist schneller als der vorheriger Ansatz, weil sie direkt auf Blaze ausgewertet wird[5].

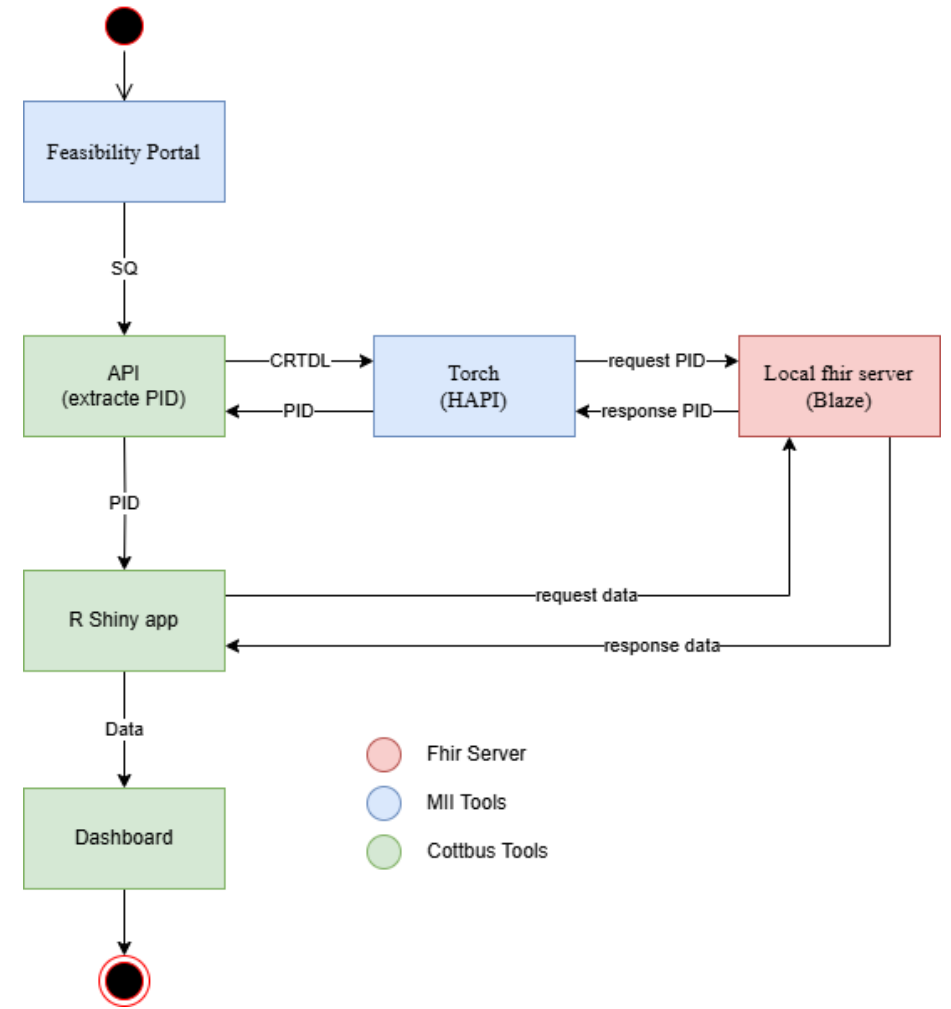


Methode - Zweiter Ansatz

Nachteile:

- PID-Extraktion folgt den ersten Ansatz.
- Zahlreiche Anfragen von R Shiny app an Blaze wegen Ablauf des Token (OAuth 2.0).
- Konvertierung vom json-Format in das Record-Format => zeit- und ressourcenintensiv.

Aus diesem Grund war es auch notwendig, eine effiziente Lösung zu finden, um Zeit und Ressourcen zu sparen. Aber die Frage wie?



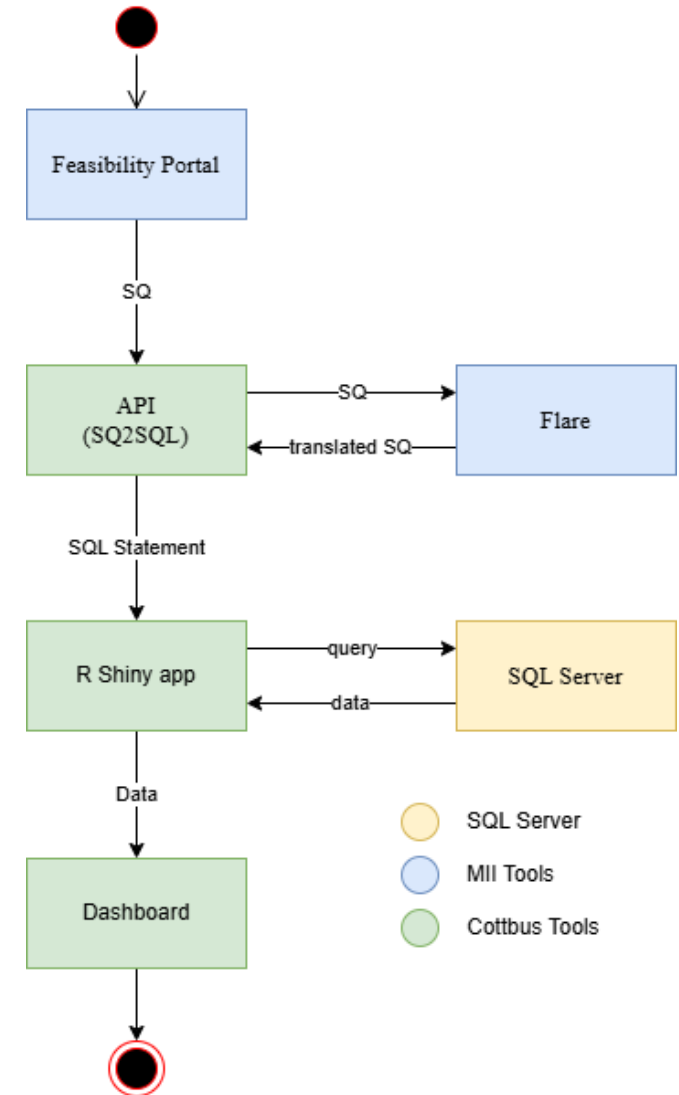
Methode - Dritter Ansatz

Das dritte Szenario fordert:

- Flare Tool [6].
- API SQ2SQL.
- SQL Server.
- Visualisierungstool.

Prozessablauf:

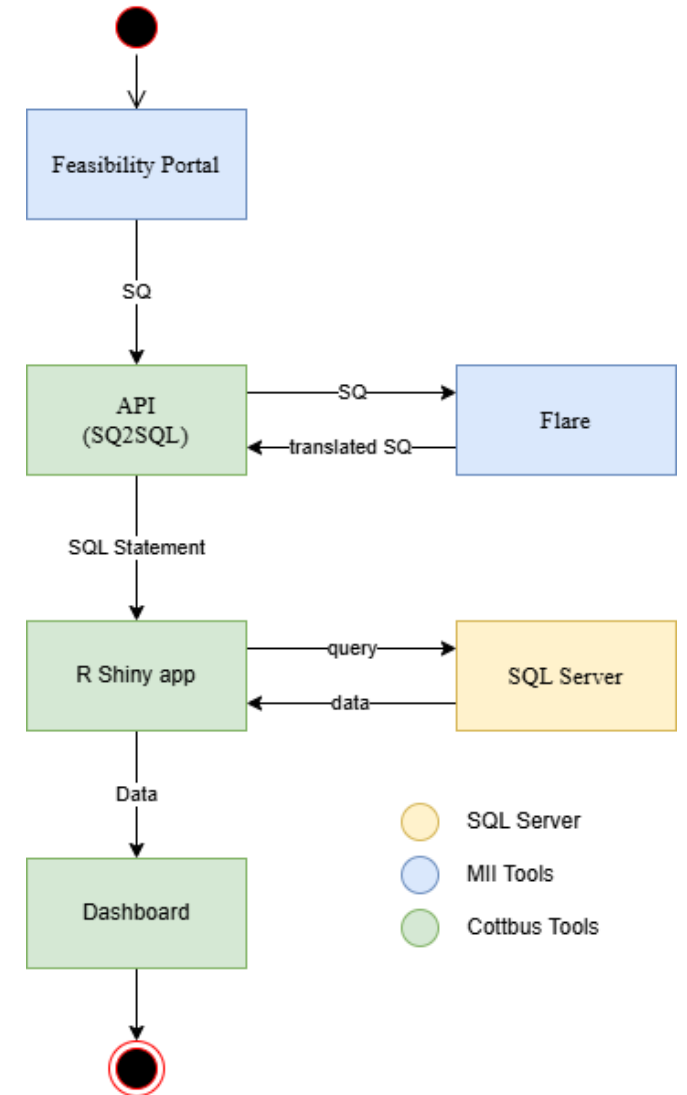
- SQ exportieren.
- SQ an API SQ2SQL senden.
- Flare übersetzt SQ in FHIR Search.
- API SQ2SQL übersetzt FHIR Search in SQL Statement.
- API SQ2SQL send generierte SQL Statement an R shiny app.
- (R shiny app) erstellt gewünschte SQL Anfrage basiert auf SQL Statement.
- Visualisierung.



Methode - Dritter Ansatz

Vorteile:

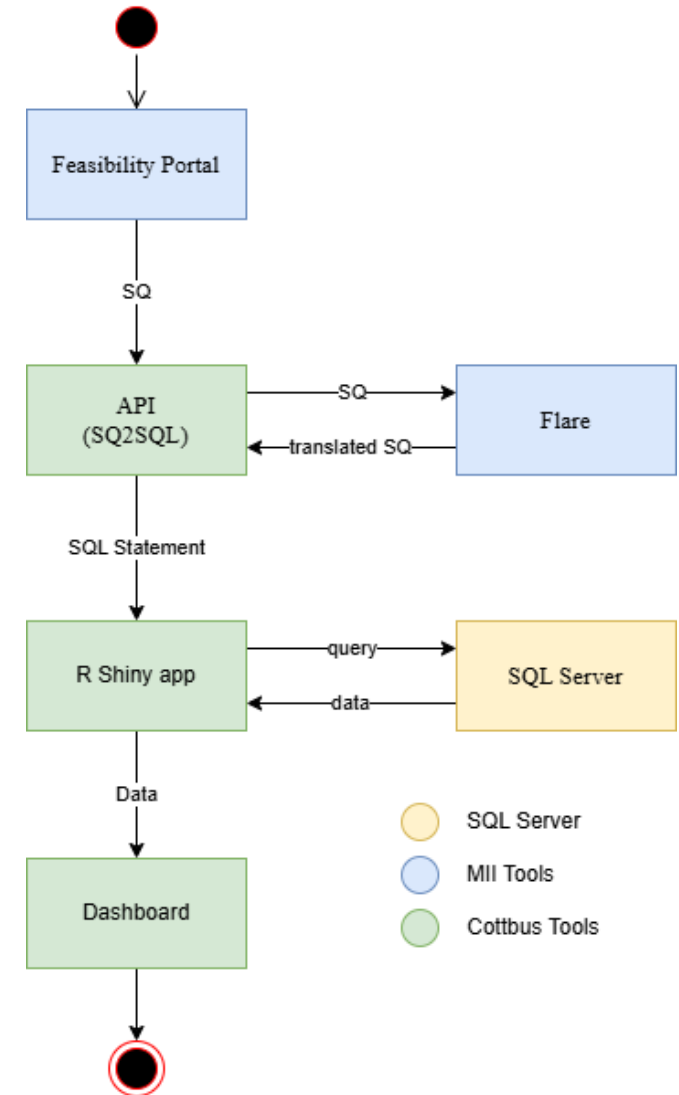
- Die Übersetzung von FDPG SQ zu SQL Statement ist nicht aufwendig, da man hier Texte bearbeitet ohne Semantik.
- Die Auswertung der Anfrage erfordert nicht Torch -> mehrere *.ndjson files.
- Kann man viele Attribute abfragen, da notwendige Tabelle verbunden werden.
- Das Ergebnis ist in Tabellenform und nicht im Json-Format.
- Nach Schließung des Dashboards werden die Daten nicht gespeichert.



Methode - Dritter Ansatz

Nachteile :

- Ansatz anwendbar bei Standorte, die Relationale Datenbank haben.



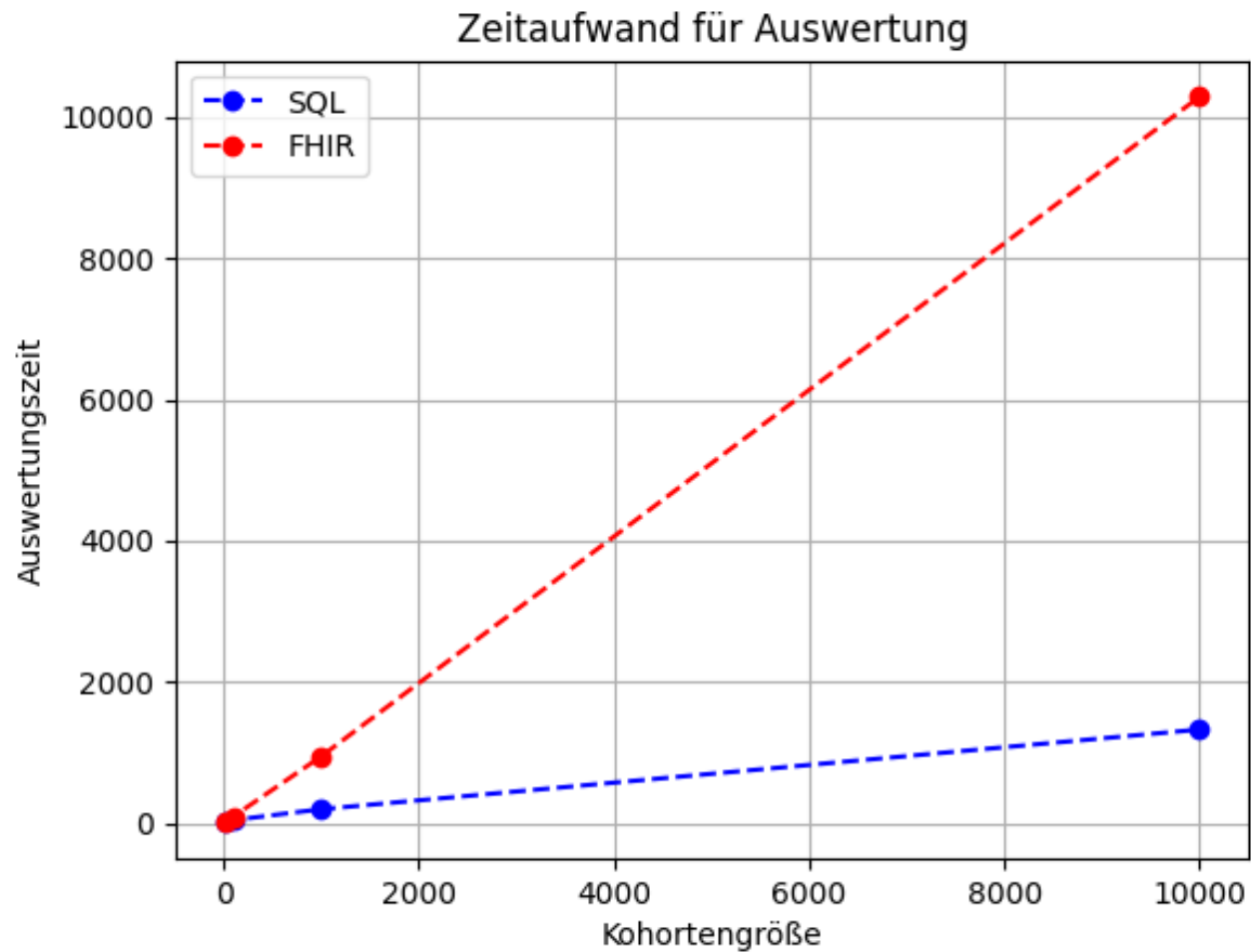
Vergleich

- Um dies zu beweisen, haben wir ein Experiment durchgeführt. In unserem Experiment haben wir mehrere Kohorten im Feasibility Portal erstellt.
- Jede Kohorte hat nur eine Diagnose, keine Filter oder etwas anderes.
- In unserem Experiment betrachten wir vier Kohorten, die mit ungefähr 10, 100, 1000, 10000 Patienten haben.
- Wir haben zwei Tools entwickelt, das erste mit einem Fhir-Ansatz und das zweite mit einem SQL-Ansatz.
- Die SQs werden dann in der Reihenfolge der Kohorten in die Tools importiert und ihre benötigte Zeit für die Auswertung wird dokumentiert.
- Für diesem Experiment haben wir Maschine mit 5 Kerne und 200 GB Arbeitsspeicher benutzt.

Vergleich

| 5 Cores, 200 GB RAM | | | | |
|---------------------|------|------------|-------------|--------------------|
| Größe der Kohorte | 10 | 100 | 1000 | 10000 |
| Zeitsaufwand SQL | 10 s | 45 s | 3 m : 19 s | 22 m : 07 s |
| Zeitsaufwand FHIR | 15 s | 1 m : 35 s | 15 m : 45 s | 2 st : 51 m : 26 s |

Vergleich

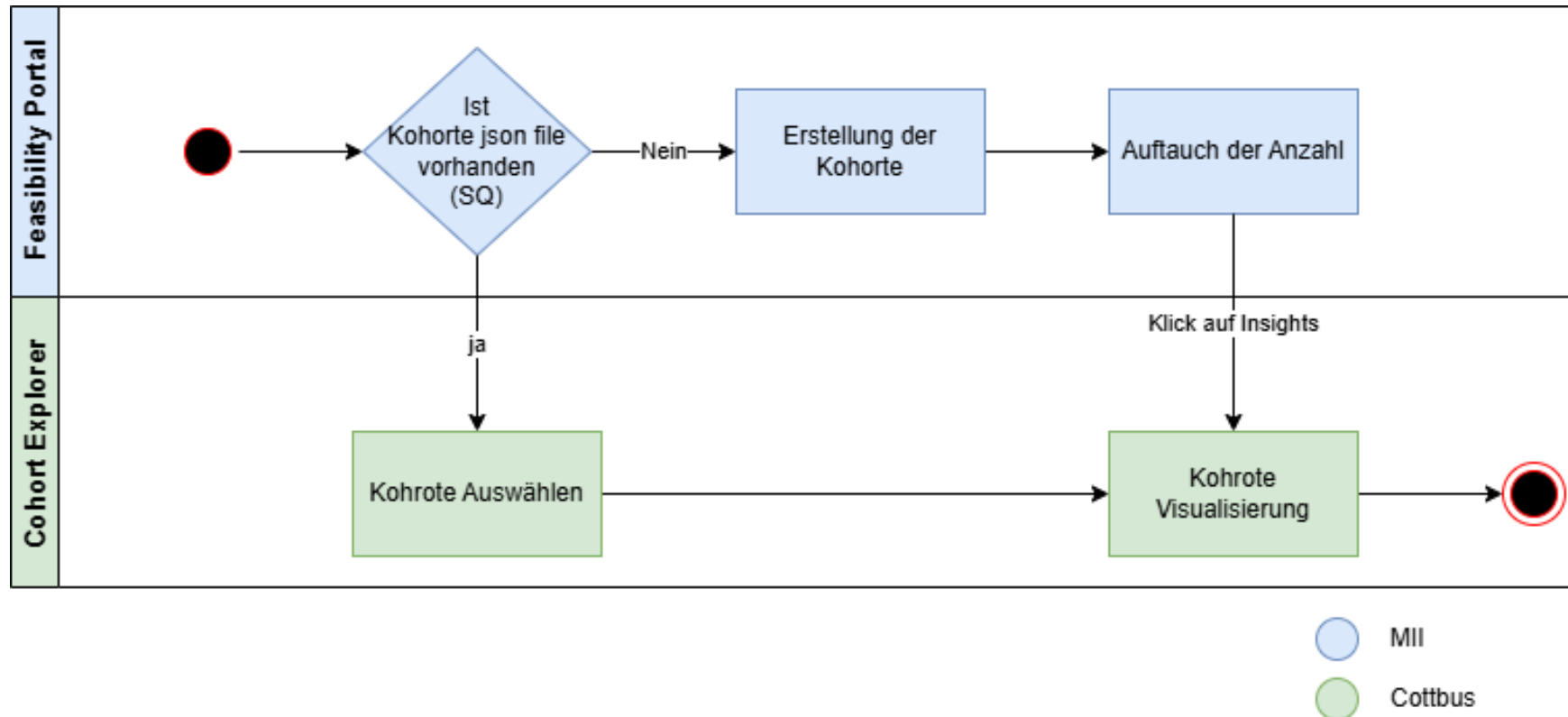


Ergebnis des Vergleichs

- SQ wurde als grundlegendes Element identifiziert.
- Die Extraktion der PID ermöglicht die Abfrage zahlreicher Informationen über den Patienten.
- Für den FHIR-Ansatz war das Tool „Torch“ eine wesentliche Komponente für die Extraktion der Patienten-ID.
- Für den SQL-Ansatz war das Konvertierungstool von SQ zu SQL Statement eine wesentliche Komponente für die Extraktion der Patienten-ID.
- Der SQL-Ansatz ist schneller und effizienter als der FHIR-Ansatz.

=> Aus diesem Grund wurde in Cottbus Standort SQL-Ansatz angewendet.

Arbeitsablauf



Arbeitsablauf - Kohorte vorhanden



Deutsch

Ahmad Albenny

Daten Definition

Kohorten Selektion

Gespeichert Abfragen

Einstellungen

1

Kohorten Selektion

Willkommen beim Machbarkeitsportal der MUL-CT. Es sind Daten ab 2003 bis heute verfügbar! Definieren Sie Ihre Kohorte anhand von Kriterien die Sie über das Suchfeld auswählen können. Für gezieltes Suchen können Sie auf bestimmte Terminologien (z.B. ICD10, LOINC, OPS) oder KDS-Module einschränken. Sie können die Kohorten Selektion direkt verwenden um die Patientenzahl deutschlandweit abzuschätzen.

2

Merkmal Selektion



Für eine kurze Anleitung bitte auf diesen Text klicken!

Neue Kohorte

Kohorten Selektion bearbeiten

Kohorten Selektion hochladen

Kohorten Selektion herunterladen

Machbarkeitsabfrage

Kohorte zu Insights hochladen

Arbeitsablauf - Kohorte nicht vorhanden



FORSCHUNGS-
DATENPORTAL
FÜR GESUNDHEIT

Deutsch

Ahmad Albenny

Daten Definition

Kohorten Selektion

Gespeichert Abfragen

Einstellungen

1

Kohorten Selektion

Willkommen beim Machbarkeitsportal der MUL-CT. Es sind Daten ab 2003 bis heute verfügbar! Definieren Sie Ihre Kohorte anhand von Kriterien die Sie über das Suchfeld auswählen können. Für gezieltes Suchen können Sie auf bestimmte Terminologien (z.b. ICD10, LOINC, OPS) oder KDS-Module einschränken. Sie können die Kohorten Selektion direkt verwenden um die Patientenzahl deutschlandweit abzuschätzen.

2

Merkmal Selektion



Für eine kurze Anleitung bitte auf diesen Text klicken!

Neue Kohorte

Kohorten Selektion bearbeiten

Kohorten Selektion hochladen

Kohorten Selektion herunterladen

Machbarkeitsabfrage

Kohorte zu Insights hochladen

Arbeitsablauf - Kohorte nicht vorhanden



Deutsch

Ahmad Albenny

Daten Definition

Kohorten Selektion

Gespeichert Abfragen

Einstellungen

Machbarkeitsabfrage wird bearbeitet

Die Anzahl an Patientinnen und Patienten, die die Kriterien Ihrer Kohorte erfüllen wird standortübergreifend ermittelt. Dieses Ergebnis ist nicht exakt.

0 0 0 0 0 0 2 9 Anzahl der Patienten

Insights

Einschlusskriterien

Ausschlusskriterien

Sonstige näher bezeichnete alimentäre Mangelzustände
Diagnose, E63.8, ICD10

Kohorten Selektion bearbeiten

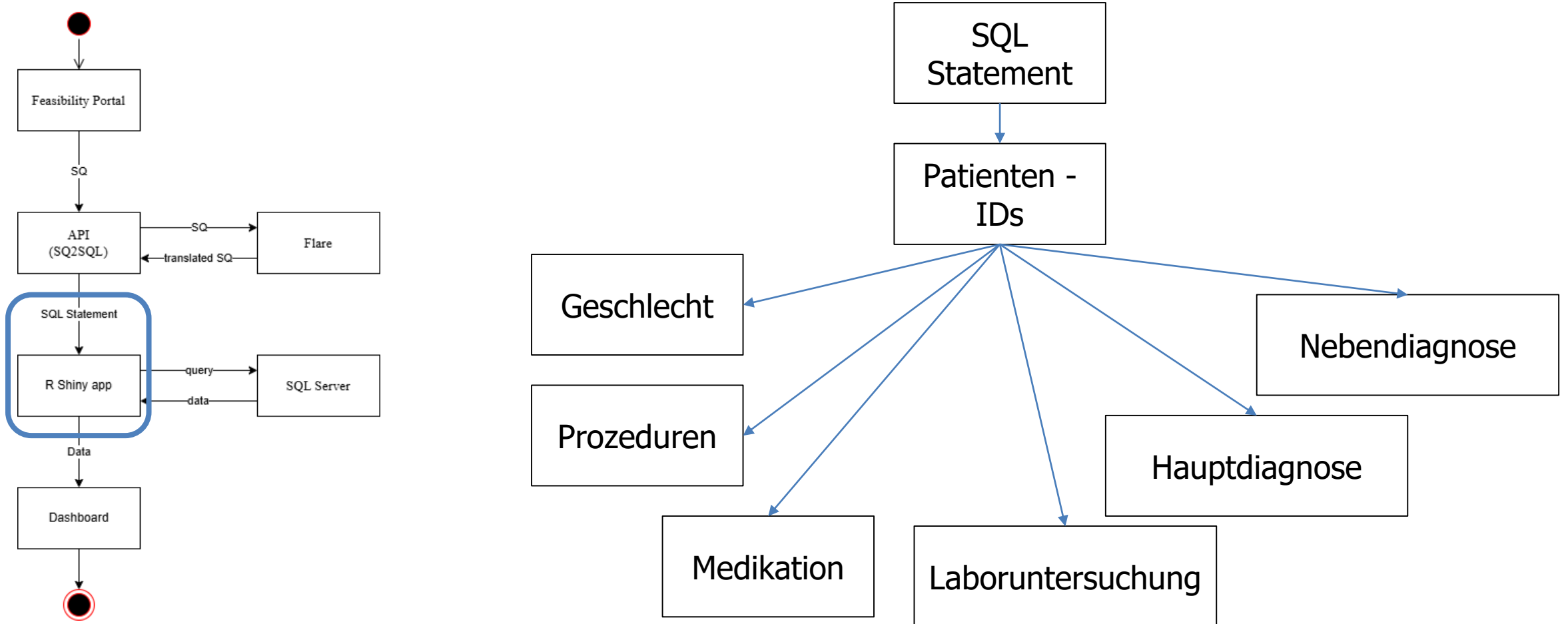
Kohorte speichern

Kohorte herunterladen

Datenvisualisierung

Dennis Hübner

Datenvisualisierung – Daten abrufen



Interaktive Visualisierungen in R



<https://plotly.com/r/>

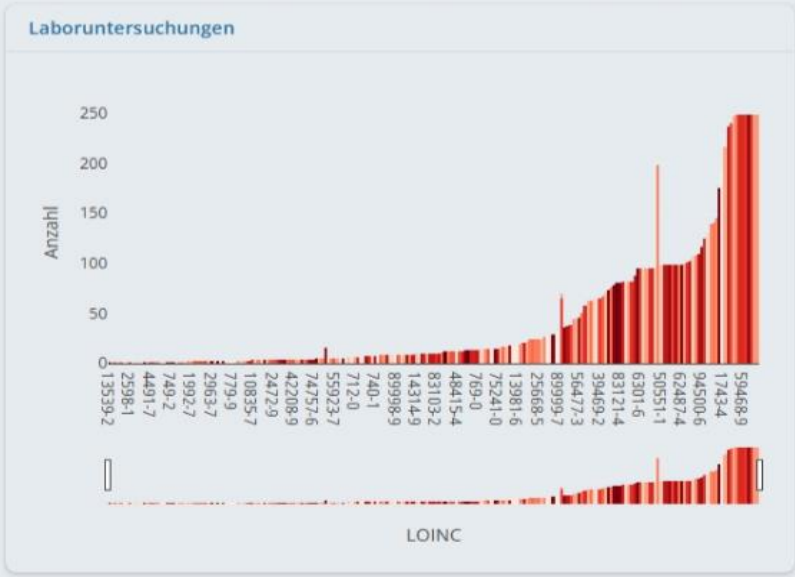
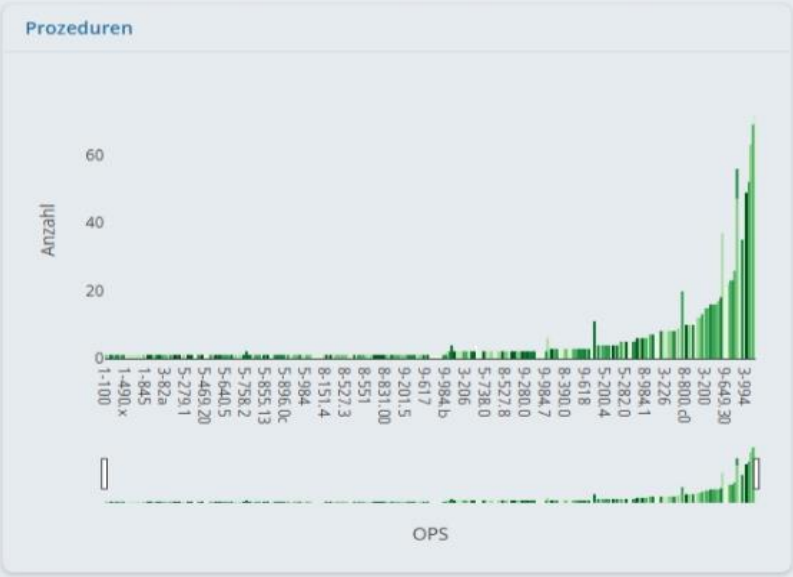
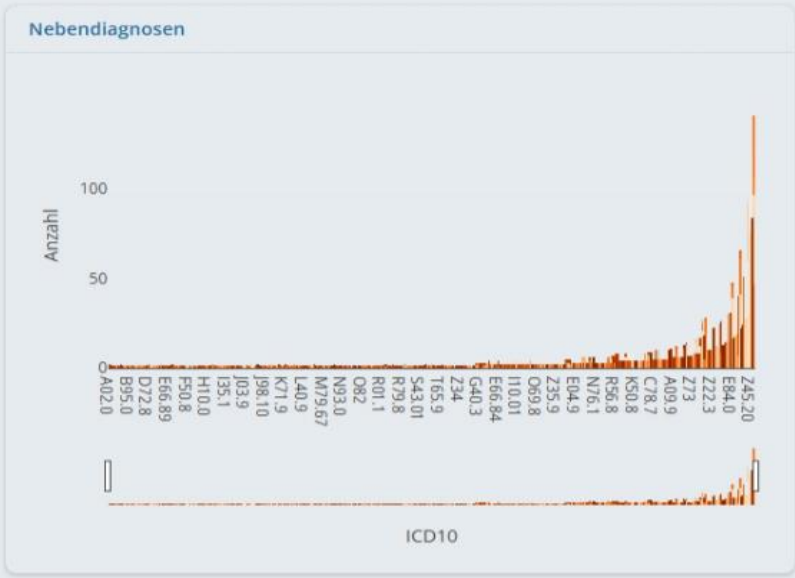
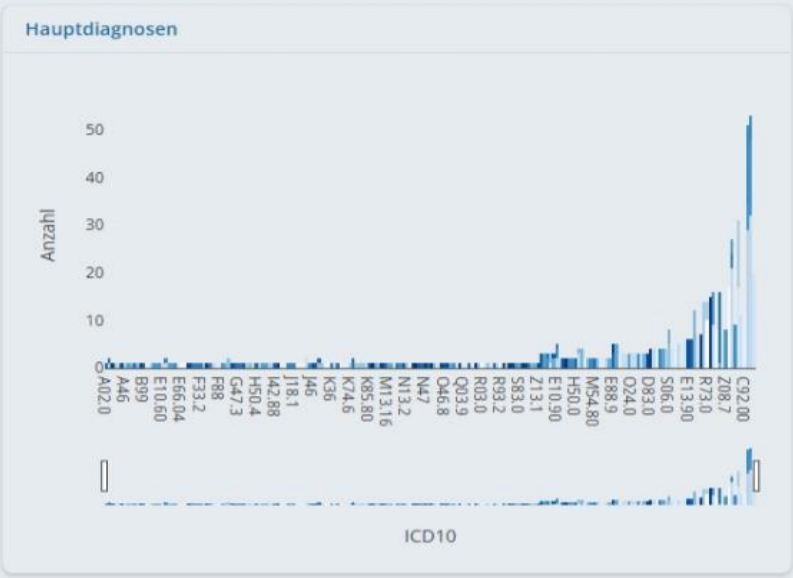


<https://echarts4r.john-coene.com/>

Shiny – Web App Framework in R



- Interaktive Web Apps in der Programmiersprache R
- User benötigt kein R
 - Link zur App
 - Browser



<

Choose .json File

Browse... 10.json

Upload complete

Process Data

Verteilung nach Geschlecht

Hauptdiagnosen

Nebendiagnosen

Medikation

Prozeduren

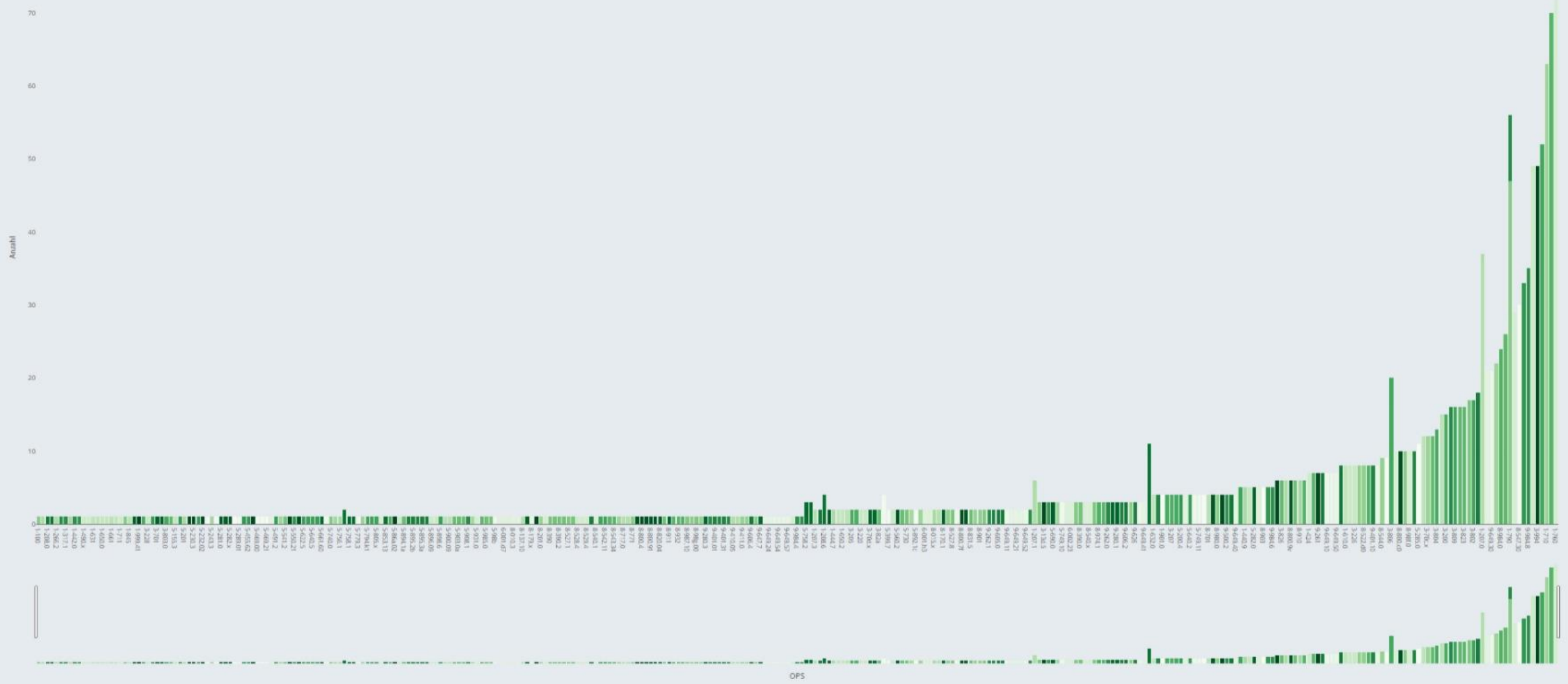
Laboruntersuchungen

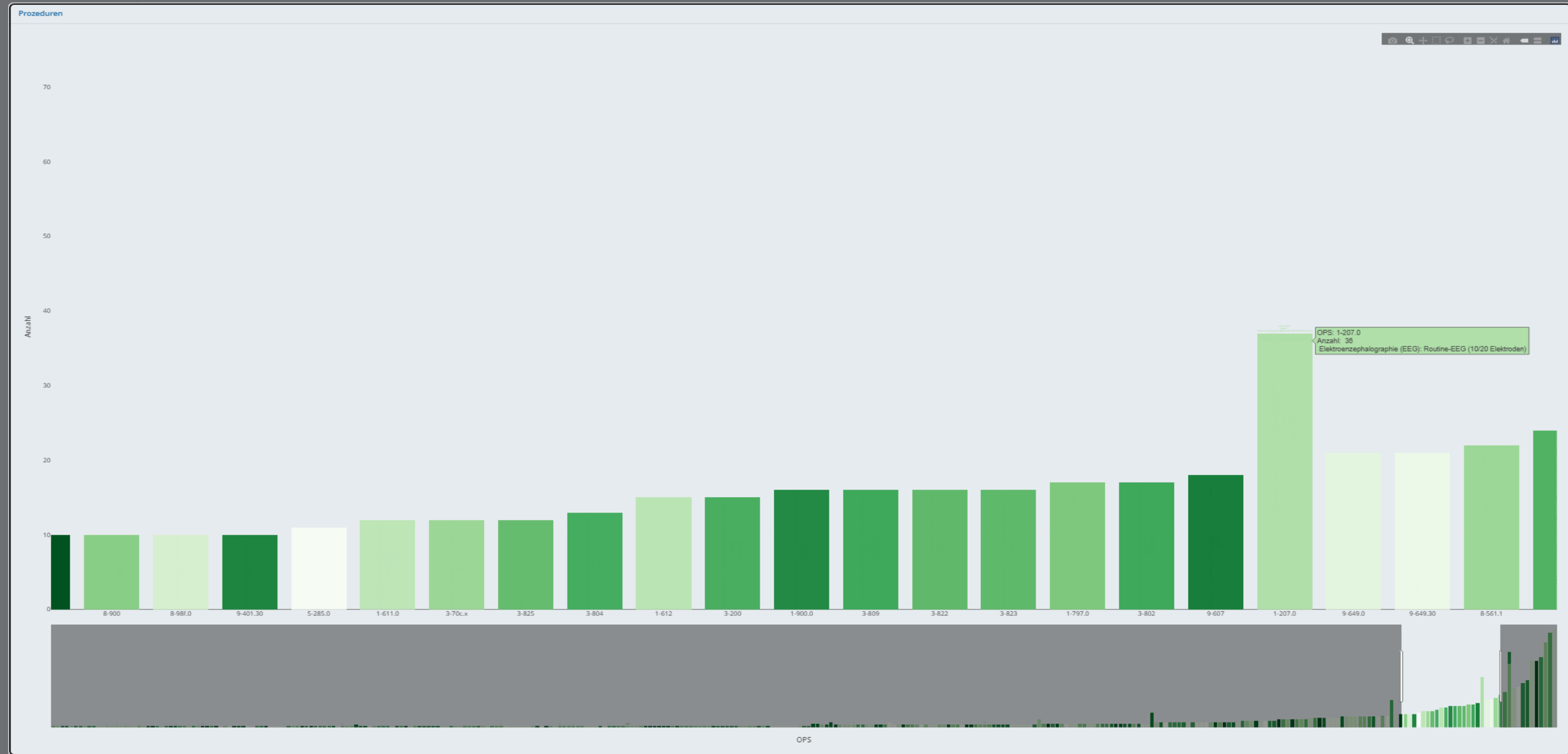
Error

Es sind weniger als 10 Patienten in Ihrer Anfrage erhalten.

Close

Prozeduren

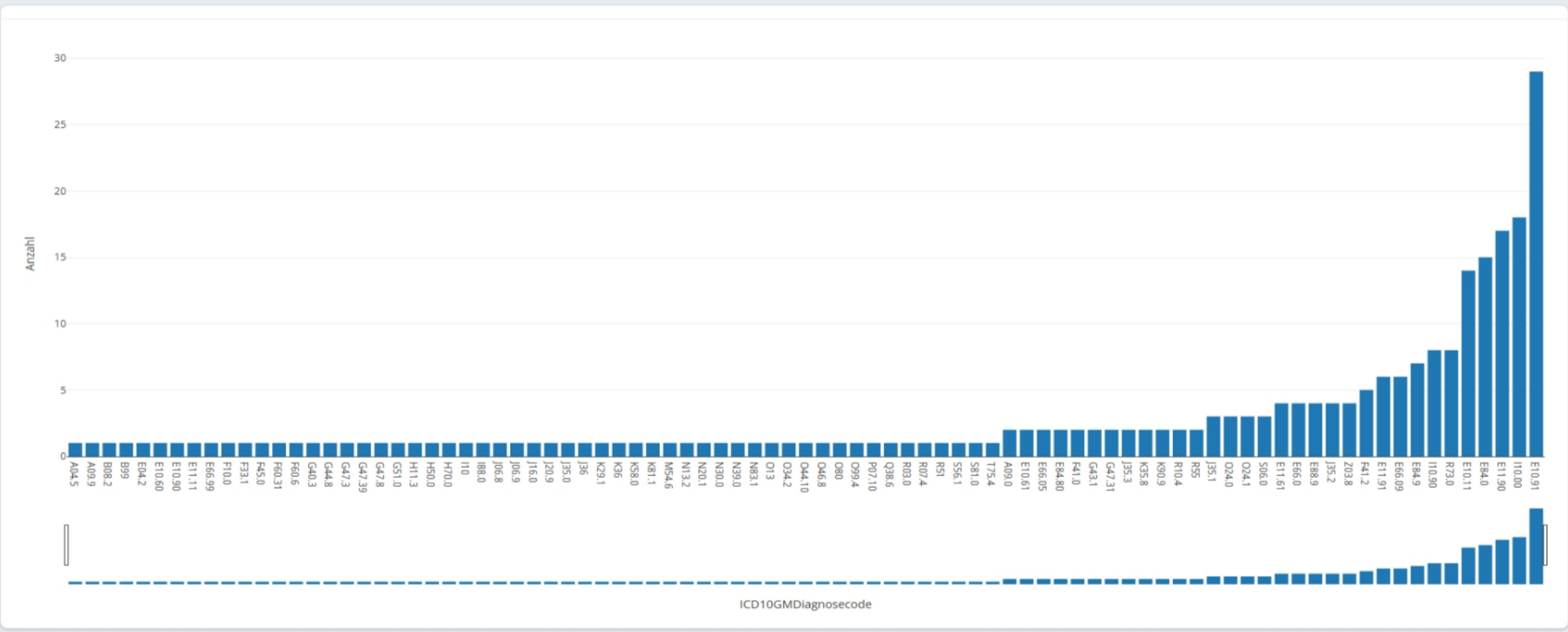




Choose x variable: Gender

Choose y variable: Hauptdiagnose

Select Gender: M W



Choose x variable:

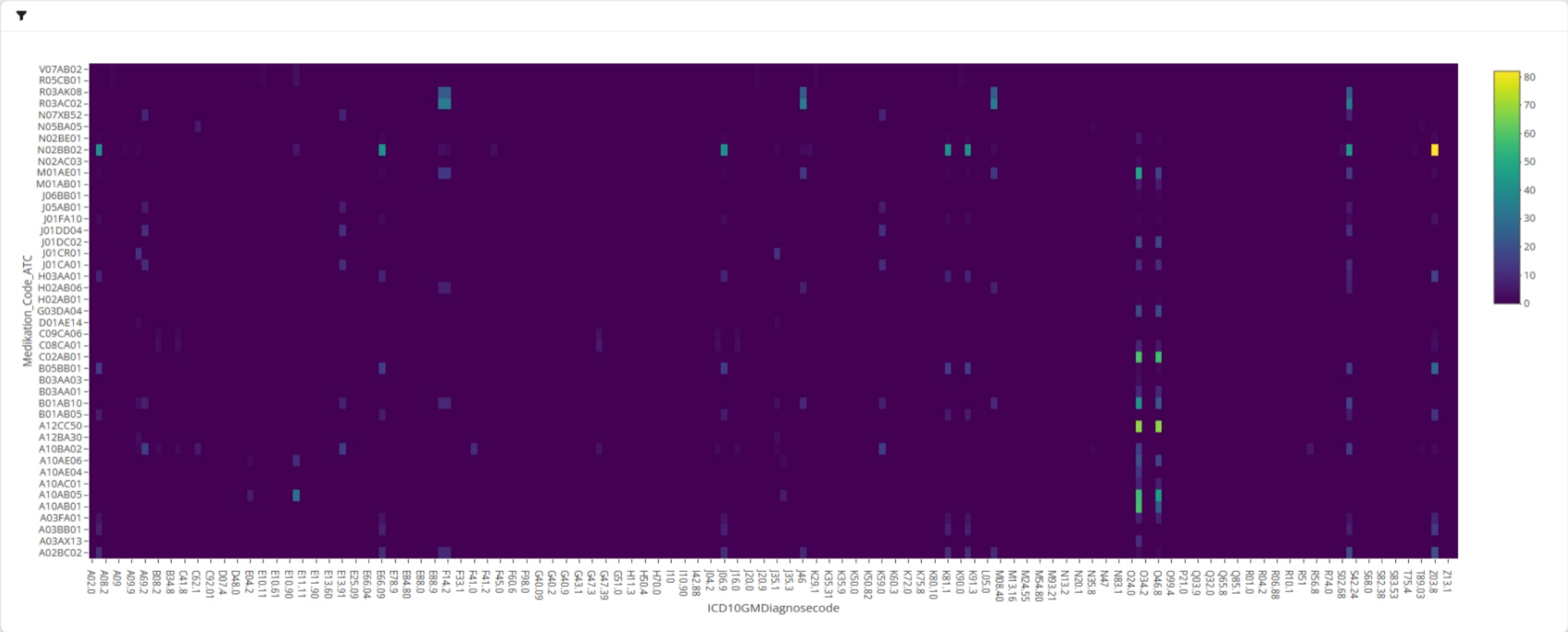
Choose y variable:

Code:

Hauptdiagnose

Medikation

Update Plot



Choose x variable:

Choose y variable:

Code:

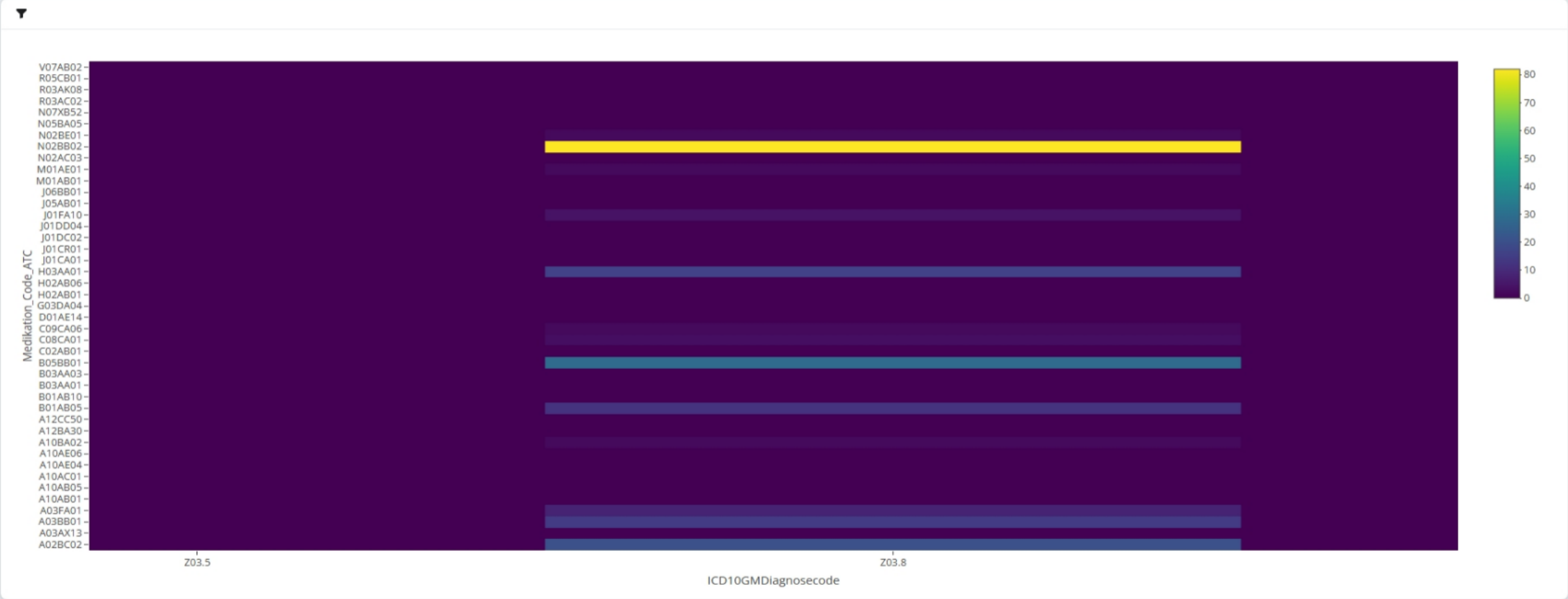
Hauptdiagnose

Medikation

Update Plot

Z03.8

• Ärztliche Beobachtung u. Beurteilung v. Verdachtsfällen, Beobachtung b. sonst. Verdachtsfällen



Choose x variable:

Hauptdiagnose

Choose y variable:

Medikation

Update Plot

Code:

Z03.8

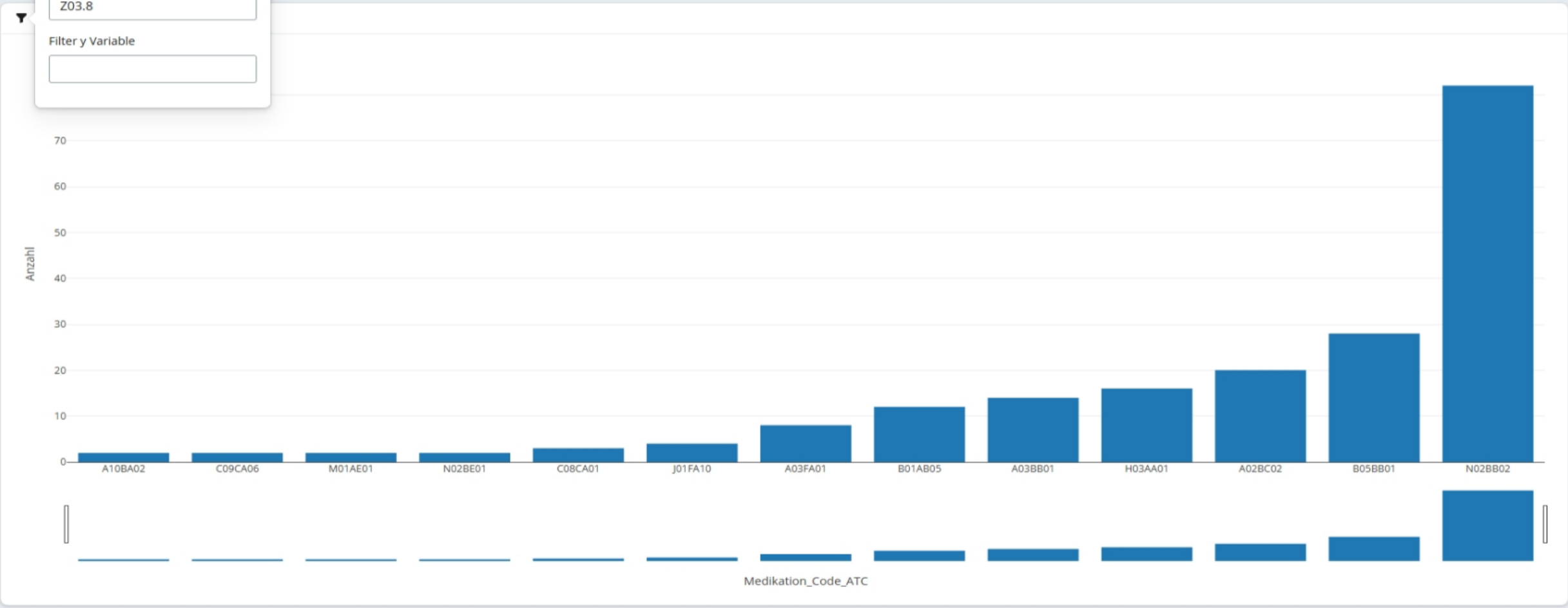
Filter Plot

Filter x Variable

Z03.8

Filter y Variable

• Ärztliche Untersuchung v. Verdachtsfällen, Beobachtung b. sonst. Verdachtsfällen



Choose x variable:

Choose y variable:

Code:

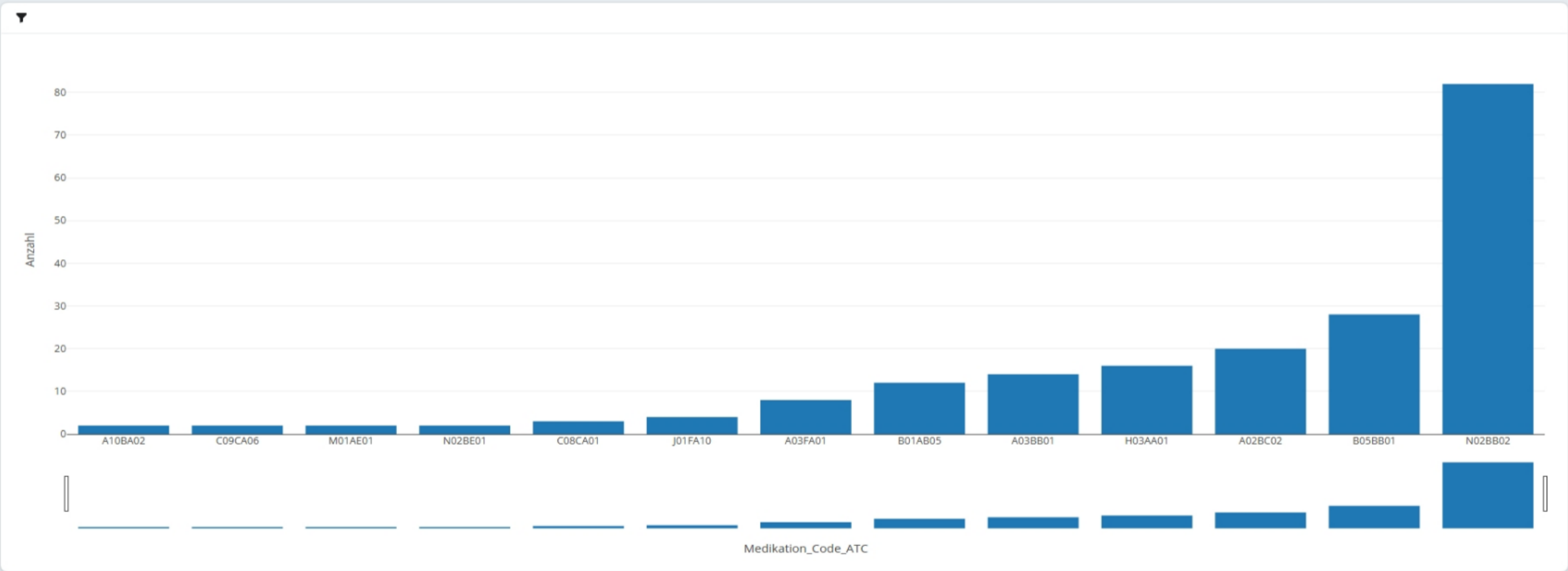
Hauptdiagnose

Medikation

Update Plot

N02BB02

- Metamizol-Natrium 500 mg TAB
- Metamizol-Natrium 500 mg/ml 50ml TEI
- Metamizol-Natrium 500 mg/ml 1000mg ILO



Referenzen

- [1] <https://forschen-fuer-gesundheit.de/daten-und-bioproben/prozesse-der-antragstellung-und-datennutzung-in-der-mii/> access 17.03.2025
- [2] <https://de.vecteezy.com/vektorkunst/4669926-mix-icon-for-count> access 17.03.2025
- [3] https://pngtree.com/freepng/dashboard-line-icon-vector_9020881.html access 17.03.2025
- [4] <https://github.com/medizininformatik-initiative/torch> access 17.03.2025
- [5] Ulrich, H., Behrend, P., Wiedekopf, J., Drenkhahn, C., Kock-Schoppenhauer, A.K. and Ingenerf, J., 2021. Hands on the Medical Informatics Initiative Core data set—lessons learned from converting the MIMIC-IV. In *German Medical Data Sciences 2021: Digital Medicine: Recognize–Understand–Heal* (pp. 119-126). IOS Press.
- [6] <https://github.com/medizininformatik-initiative/flare> access 17.03.2025

Diskussion

Danke für Ihre Aufmerksamkeit

- Welche Szenarien könnten für Sie in Frage kommen?
- Was könnte Ihrer Meinung nach verbessert werden?
- Können Sie sich vorstellen den Cohort Explorer zu nutzen?

