

Gesundheitsdaten im FDZ verstehen

Potenzial, Struktur und analytische Grenzen:

Ein Einblick für Datennutzende

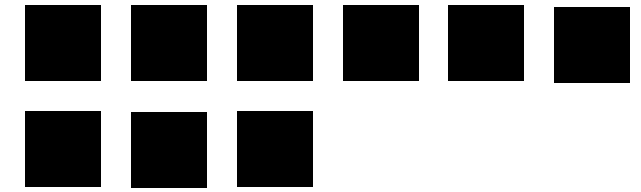
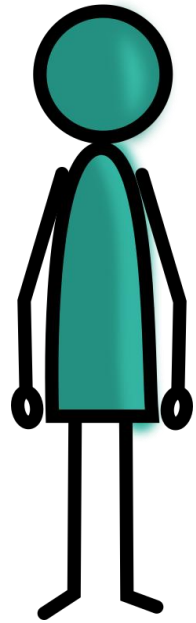
Dr. Melanie Ludwig

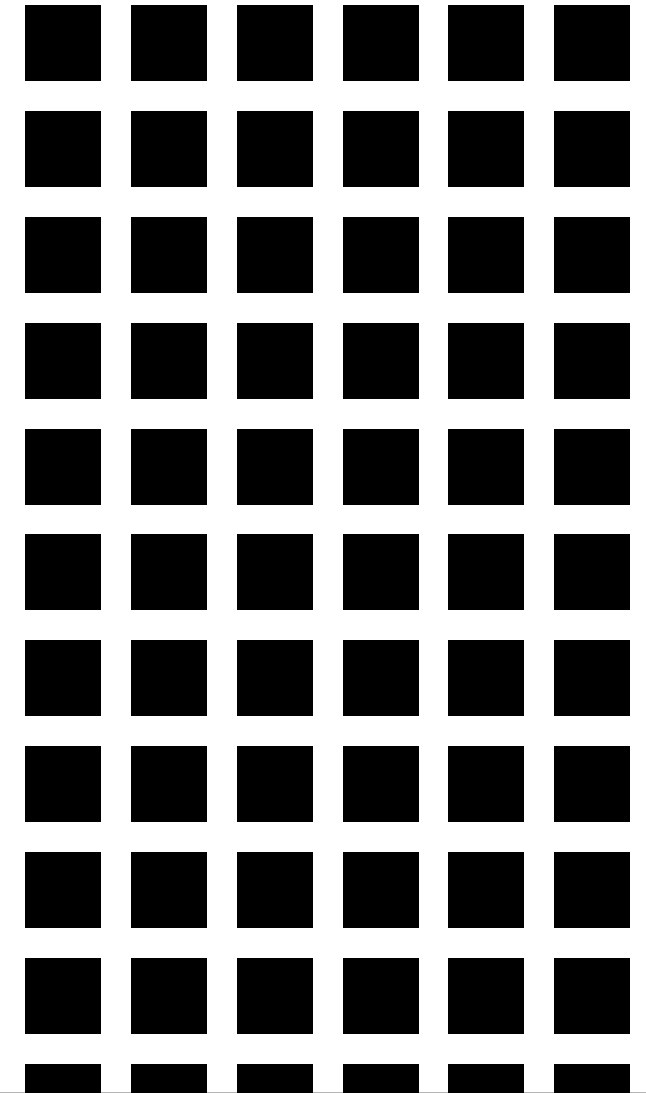
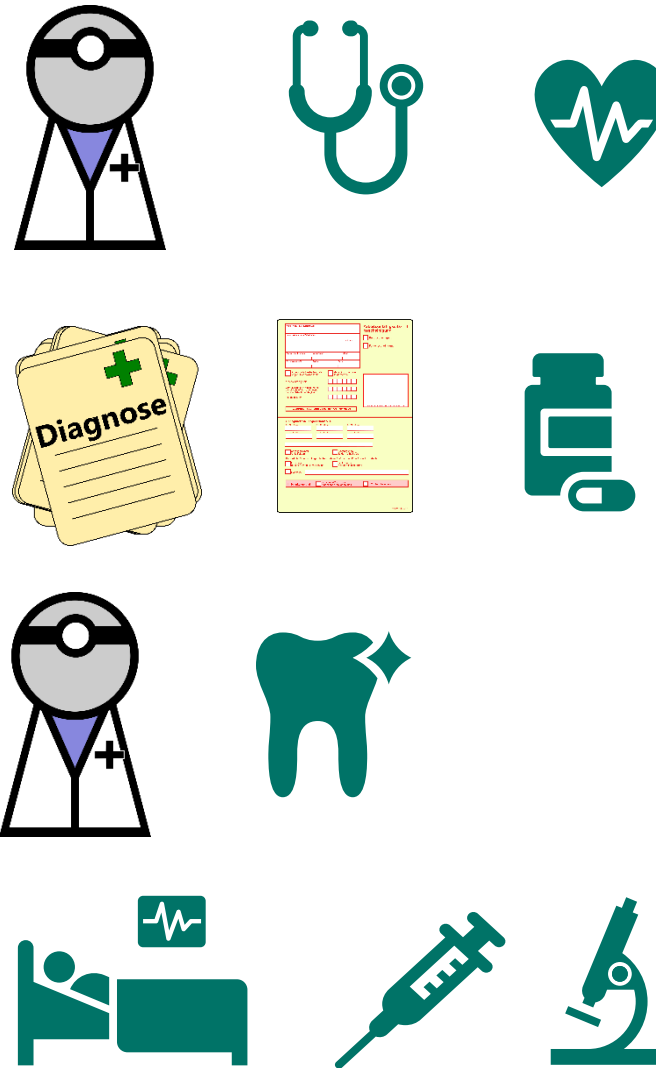
Datenkoordinatorin



Daten aus der Gesundheitsversorgung



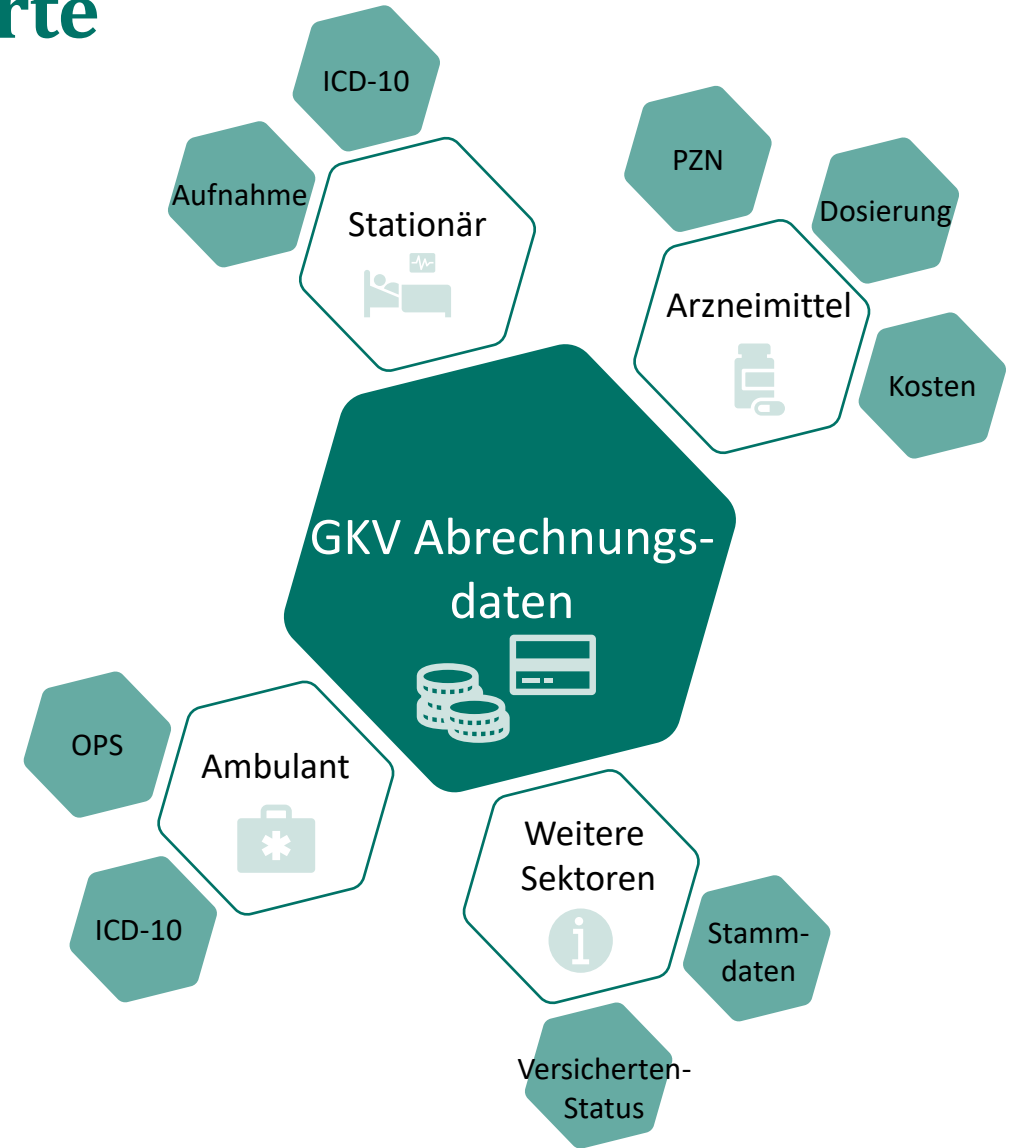


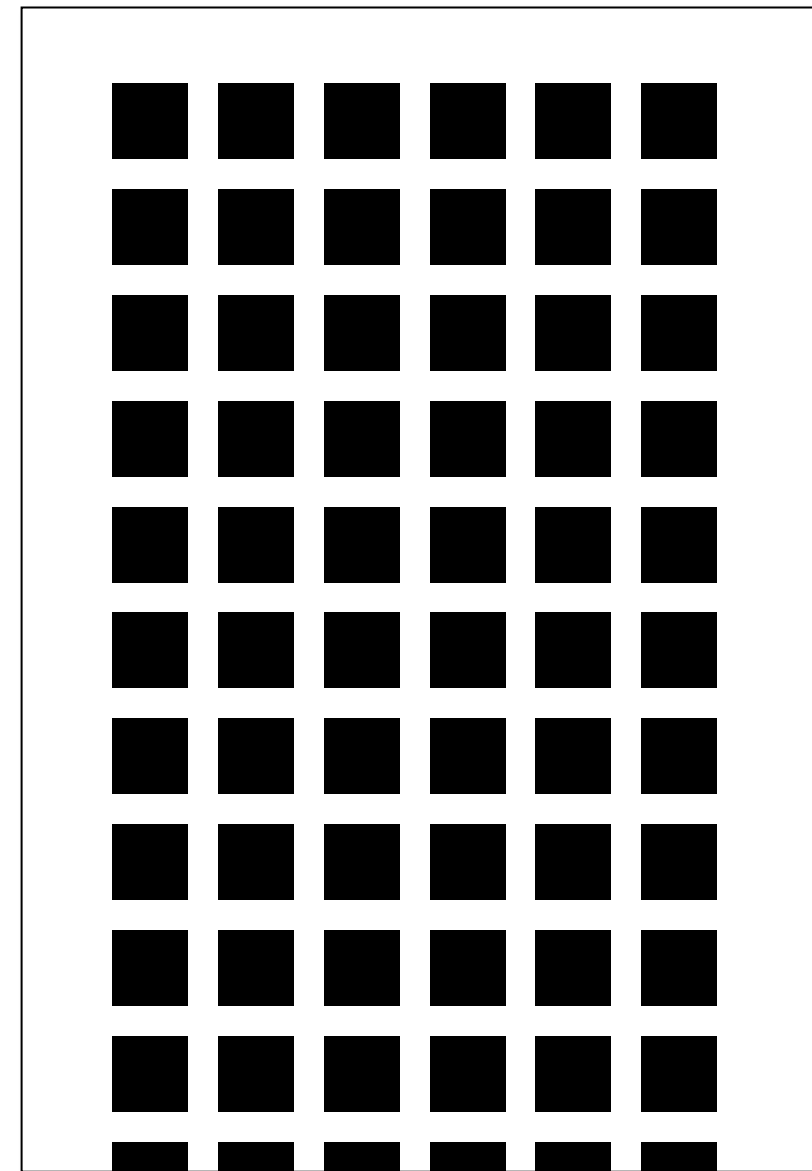
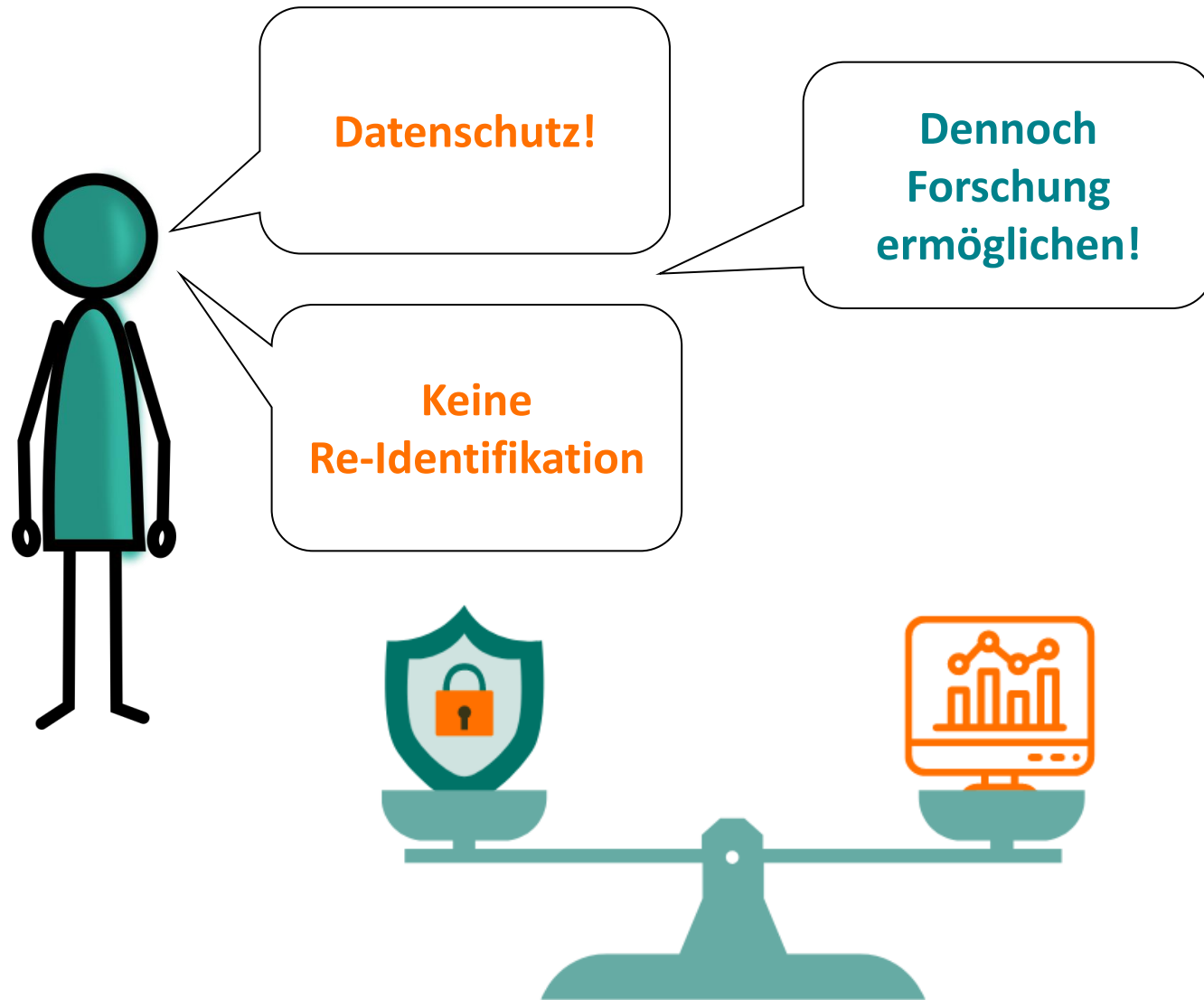


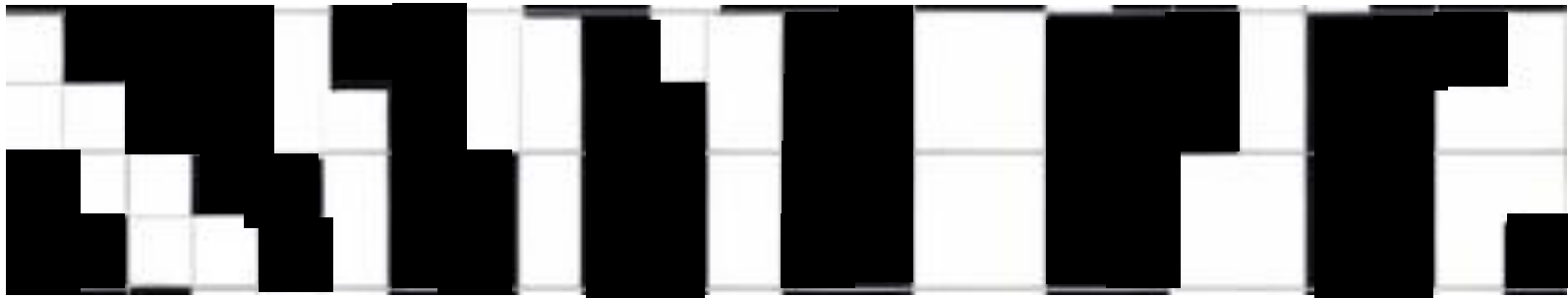
Status Quo – 74 Millionen Versicherte

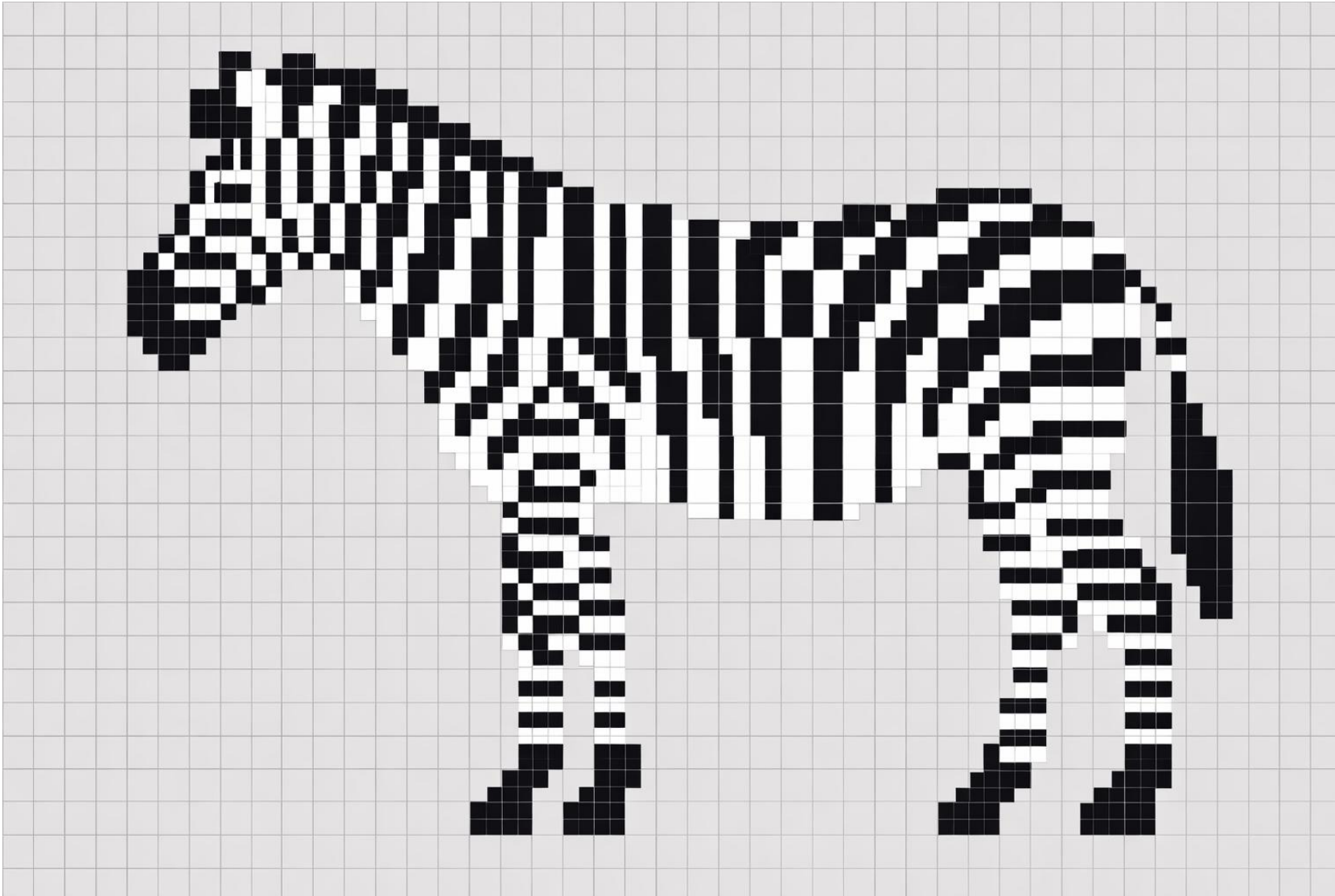
Daten seit 2009, erste große Erweiterung 2019

- Aktuellste Daten: 4 zentrale Sektoren mit Fall-Inhalten in insgesamt 17 Detailtabellen
- Zwischen 10 Mio. und 3 Mrd. Zeilen pro Jahr pro Tabelle
- Durchschnittlich rund 650 Mio. Zeilen pro Jahr pro Tabelle und 11 Mrd. Zeilen pro Jahr







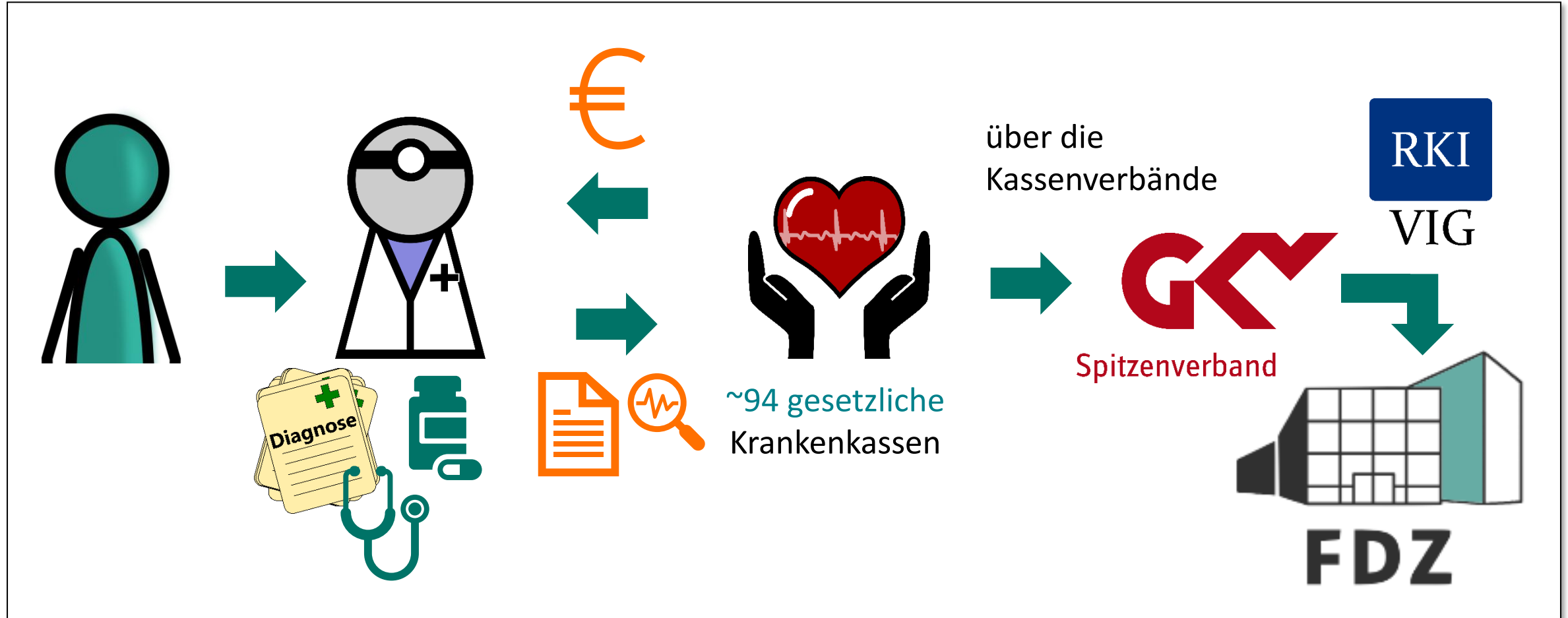


Graphik KI generiert

Woher kommen die Daten im FDZ?



Der Weg der Daten: Von den Versicherten bis ins FDZ...



**Was beinhalten die Daten im FDZ?
...und was nicht?**



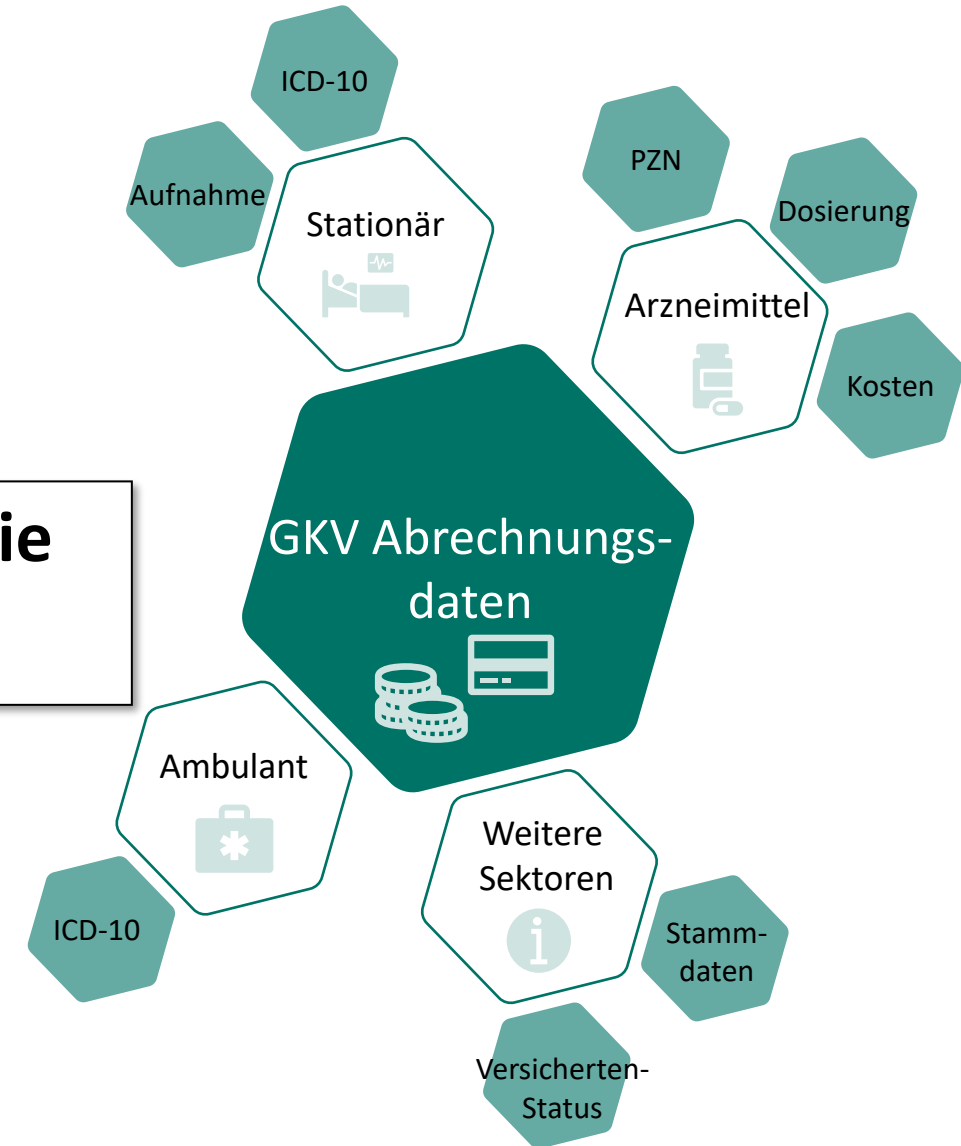
Dateninhalte und Pseudonymisierung

Diagnosen
Medikamente
Behandlungen (z.B. Hausarzt, Facharzt, Krankenhaus)
Durchgeführte Prozeduren (z.B. Operationen)
Zahnärztliche Behandlungen
Zeitliche Verläufe über viele Jahre
Versorgung in unterschiedli
...



Details aller Inhalte über die Datensatzbeschreibung

Namen
Adressdaten
Konkrete Geburts-/Sterbedaten (nur Jahr)
Krankenversicherungsnummer / Lebenslange Arztnummer
Freitexte
Arztberichte
Gesprächsinhalte
...



Ein Blick in die Tiefe



Personen u

Alte Struktur im öffentlichen PUF, in den Analyseräumen VARCHAR(32) wie PSID!

Daten stammen aus dem Public Use File.

KHFALL (Ausschnitt)

PSID	VSID	FALLIDKH	KHPSEUDO	KHKCLASS	KHREGKZ	KHPRUEF	AUFNDAT	AUFNGRUND	ENTLASSGRUND	AUFNFA	EINWEISPSEUDO	EINWEISFG	EINWEISPRUEF	VERANL
CzSUmwj37jMYLp3aUhb	615404492	52320439	015860752329	26	5	1	20190121	107	19	2800	653013731837	1.0	1.0	81363
ByqwuO65udLZTw4EUrY	274403617	231401	090751214688	26	15	1	20190416	107	19	2300	793621206343	1.0		79535
KSPvaWPJICsACenaM6Q	533881692	3197416	097260728342	26	3	1	20190210	107	31	1500	377765534505	1.0	1.0	73599
80PV6vEd1NI3J2DmHmr	045610214	36548249	079547623677	26	12	1	20190901	201	19	100	005180874549			43200
S2sca5fRX9KYYXRU48bx	412535457	3261702	071639947278	26	5	1	20190717	101	42	2425	753182326663	1.0	1.0	78217
O19ZYIRjhJBeeUXIYG	007759368	31466405	031472586282	26	16	1	20191114	107	29	2600	760950870552	1.0	1.0	07687
3yMkaXPRm4u8QTYUCI8	408035775	3555242	041233110421	26	9	1	20190415	101	19	1500	807262280327	0.0		92365
ORhjRITbAakcwrWMZE9	206094413	3256450	028973789282	26	5	1	20190213	107	19	2800	344041851841			69394
mqHdEEK5yO1jBDtRwOZ	899861165	340498	004144055804	26	3	1	20191220	107	19	1500	444045544472			59714
D8sxmodkmWwnSczOKs	872670869	39561062	001056276107	26	3	1	20190125	507	12	2600	848923075730	1.0	0.0	84885

Personen, Fälle, Leistungen

PSID: Personen-ID = *wer*

VSID: Versicherten-ID = *wer im Jahr bei einer Kasse*

FALL-ID = *was ist konkret passiert*

Analyse-Hinweise

- ✓ Überjährige Analysen immer über PSID
- ✓ FALL-IDs in Bezug zu Personen setzen (mind. PSID, „sicherer“ PSID+VSID)
- ✗ VSID über Jahre hinweg joinen

ID-Typ	Bezieht sich auf	Zeitbezug	Typische Anzahl
PSID	natürliche Person	überjährig eindeutig	1 pro Person
VSID	Person im Berichtsjahr	jahres- & kassenbezogen	i. d. R. 1 pro Jahr (Ø1,01)
FALL-ID	konkreter Versorgungsfall	fallbezogen	viele pro Jahr pro Person

Unterschiede der Datenmodelle

SA651 (DM1/2)

SA651_SATZART	SA651_AUSGLEICHSJAHR	SA651_BERICHTSJAHR	SA651_VSID	SA651_PSID	SA651_LEISTUNGSQUARTAL	SA651_ICD_CODE	SA651_ICD_ZUSATZ	SA651_QUALIFIZIERUNG	SA651_LOKALISATION	SA651_ABRECHNUNGSWEG
651		2017	81057027	d1fVfChqaNBIZDMzlxJ	2	M5499		G	0	1
651		2017	72724821	s5e8SQSvmIaZoNFIJ7z	2	N921		G	0	1
651		2017	95605202	w0JFFV4OK44oyfdj9Tv	1	E04				
651		2017	27842640	cYvV6aUSgNKKLX6Qh4T	3	H6				

AMBDIAG (DM3)

PSID	VSID	FALLIDAMB	DIAGSICH	DIAGLOKAL
CzSUMwj37jMYLp3aUHb	615404492	11025774	Other	
p9ldheRW1ohJ5pEOmvo	819765740	877822171	G	
JWXJiubAuo85CS2Tl2f	717562234	83232461	G	20190214 3 2019 H401
CvXSKhYpaMvp3NerfCI	757077748	30060009	G	3 2019 UUU

Strukturelle Änderung von DM1/2 zu DM3

Zeiträume:

- 2009–2018 → DM1/2 (Satzart-Tabellen)
- ab 2019 → DM3 (modulare Struktur)

Konsequenz:

- Strukturbruch ab 2019
- Trendanalysen über Modellwechsel hinweg → Benennung und Logik prüfen!

Kohorten sauber definieren: Versicherungszeiten

Daten stammen aus dem Public Use File.

VERSQ

PSID	VSID	VERSQ	GESCHLECHT	VERSTAGE	VERSTAGEAUSL	VERSSTATUS	VERSTAGEKG	VERSTAGEKOSTERSTWAHLT	DATENMODELL	BJAHR
CzSUmwj37jMYLp3aUhb	615404492	20194	1	90	0	10499	0	0	3	2019
DKX3KbRYBhRhDgYd0Ln	079902716	20191	2	90	0	10010	0	0	3	2019
mXT6vMsR39YQn7kpWvx	534751948	20192	1	92	0	10010	0	0	3	2019
6axCZCorQBUvR5DXyqS	714434665	20191	1	90	0	10499	0	0	3	2019
sHxeQYgzlbCkNWuO7sY	651263745	20193	2	90	0	10010	0	0	3	2019
AozlZ0Dk2PVFEYyXxsF	653954607	20193	1	92	0	11010	0	0	3	2019
dkJ75IJZqudYBiYGqfJ	429496691	20192	1	91	0	10010	0	0	3	2019
ky2LSl23kOkAlw1G48T	357097647	20193	1	92	0	11010	0	0	3	2019
nBhpi0pa3SpjWCffAbQ	019235054	20193	1	91	0	11010	0	0	3	2019
fw1DsQHjzO7H90hf0n5	621133343	20191	1	91	0	10010	0	0	3	2019

Kohorten sauber definieren: Versicherungszeiten

Analyse-Hinweise

- Versicherungszeiten prüfen (VERSTAGE)
- Auslandstage berücksichtigen (VERSTAGEAUSL < VERSTAGE)
- Vollständige Versicherung über Zeitraum prüfen

Warnung

„Keine Leistung“ ≠ „gesund“

Mögliche Gründe:

- Auslandsaufenthalt
- Unterjährige Versicherung
- Selbstzahlerleistungen

Beispiel:

VERSTAGE = 90

VERSTAGEAUSL = 30

→ Inlandstage = 60

innerhalb eines Quartals

PSID	VSID	VERSQ	GESCHLECHT	VERSTAGE	VERSTAGEAUSL	VERSSTATUS	VERSTAGEKG	VERSTAGEKOSTERSTWAHLT	DATENMODELL	BJAHR
CzSUmwj37jMYLp3aUHb	615404492	20194	1	90	0	10499	0	0	3	2019
DKX3KbRYBhRhDgYd0Ln	079902716	20191	2	90	0	10010	0	0	3	2019
mXT6vMsR39YQn7kpWvx	534751948	20192	1	92	0	10010	0	0	3	2019
6axCZCorQBUvR5DXyqS	714434665	20191	1	90	0	10499	0	0	3	2019
sHxeQYgzlbCkNWuO7sY	651263745	20193	2	90	0	10010	0	0	3	2019
AozlZ0Dk2PVFEYyXxsF	653954607	20193	1	92	0	11010	0	0	3	2019
dkJ75IJZqudYBiYGqfJ	429496691	20192	1	91	0	10010	0	0	3	2019
ky2LSl23kOkAlw1G48T	357097647	20193	1	92	0	11010	0	0	3	2019
nBhpi0pa3SpjWCffAbQ	019235054	20193	1	91	0	11010	0	0	3	2019
fw1DsQHjzO7H90hf0n5	621133343	20191	1	91	0	10010	0	0	3	2019

Verordnungen von Arzneimitteln vs. Therapie

Daten stammen aus dem Public Use File.

REZ (Auszug)

ZNR	PZNREZ	VODAT	BSNRVOPSEUDO	BSNRVOVB	BSNRVOREGKNZ	LENRVOPSEUDO	LENRVOFG	LENRVOPRUEF	LENRVOPSEUDOARZTNR	ABGABEDAT	BEGF
250	3279505	20191031	302823623151		72.0	311512974003	15.0	1.0	0	20190401	
57	00232236	20190218	656465865205		93.0	0218					
7292	01798000	20190514	924501508166		66.0	8038					
78846	00811709	20191111	985774377218		40.0	7687					
9058	01995930	20190617	113444572283		66.0	0908					

Warnung

1. Verordnung $\neq$ Einnahme
2. Nicht eingelöste Rezepte fehlen.
→ Das ändert sich mit den Daten der ePA (eML)

Was Arzneimittel-Daten leisten können:

- Verordnungen von Leistungserbringern
- Abgabedaten
- Wirkstoffebenen via ATC-PZN-Mapping

Was Arzneimittel-Daten **nicht** leisten können:

- Einnahmetreue
- Tatsächliche Dosierung
- Klinische Effekte

Besonderheiten bei Stationären Daten

Analyse-Hinweise

- Entlassdatum nicht direkt vorhanden
- Über Fachabteilungen rekonstruieren!

Analyse-Hinweise

KHFALL
↓
KHFA (Fachabteilungen)
↓
Entlasszeit rekonstruieren

KHFALL

	FALLIDKH	KHPSEUDO	KHKCLASS	KHREGKZ	KHPRUEF	AUFNDAT	AUFNGR
2	52320439	015860752329	26	5	1	20190121	107
7	231401	990751214688	26	15	1	20190416	107
2	3197416	897260728342	26	3	1	20190210	107
4	86548249	579547623677	26	12	1	20190901	201
7	8261702	171639947278	26	5	1	20190717	101
8	81466405	431472586282	26	16	1	20191114	107
5	7555242	241233110421	26	9	1	20190415	101
3	8256450	128973789282	26	5	1	20190213	107
5	840498	004144055804	26	3	1	20191220	107
9	29561062	301056276107	26	3	1	20190125	507

KHFA

	PSID	VSID	FALLIDKH	FA	ENTLASSDAT	ENTLASSZEIT	DATENMODELL	BJAHR
	CzSUmwj37jMYLp3aUHb	615404492	52320439	1500	20190509	1015.0	3	2019
	BLKymIagLQa2gZsbw5b	520635847	7265007	1000	20191104	1119.0	3	2019
	xeupUKX2A61qnDGI8SA	903731698	5120387	2400	20190825	1833.0	3	2019
	rOlV3vPPvSAXNdKT96P	819842843	188876	1500	20190222	832.0	3	2019
	7D6qDLgBpDYVTAzISSr	810548243	4774071	0	20190204	1001.0	3	2019
	r3wLXu3G20MjBvZQ6vP	125303756	85975524	1600	20190604	1510.0	3	2019
	Eo9faNIDRsqh09hMO4a	967856307	7067805	2400	20190304	1400.0	3	2019
	cChm6il0rFnAffbfqTq	763207875	240472	2	20191223	1300.0	3	2019
	bi4WsnOfYTskGya9BiT	457335642	731784	2400	20190206	1340.0	3	2019
	4YbGneADMT09MMY9mh3	123376998	66094266	0	20190412	0.0	3	2019

	FALLIDKH	KHPSEUDO	KHKLASS	KHREGKZ	KHPRUEF	AUFNDAT	AUFNGR
2	52320439	015860752329	26	5	1	20190121	107
7	231401	990751214688	26	15	1	20190416	

KHFA

			PSID	VSID	FALLIDKH	FA	ENTLASSDAT	ENTLASSZEIT	DATENMODELL	BJAHR
2	3197416	897260								
4	86548249	579547	CzSUmwj37jMYLp3aUhb	615404492	52320439	1500	20190509	1015.0	3	2019
7	8261702	171639	BLKymIaGLQa2gZsbw5b	520635847	7265007	1000	20191104	1119.0	3	2019
8	81466405	431472	xeupUKX2A61qnDGI8SA	903731698	5120387	2400	20190825	1833.0	3	2019
5	7555242	241233	rOlV3vPPvSAXNdKT96P	819842843	188876	1500	20190222	832.0	3	2019
3	8256450	128973	7D6qDLgBpDYVTazlSSr	810548243	4774071	0	20190204	1001.0	3	2019
5	840498	004144	r3wLXu3G20MjBvZQ6vP	125303756	85975524	1600	20190604	1510.0	3	2019
9	29561062	301056	Eo9faNIDRsqh09hMO4a	967856307	7067805	2400	20190304	1400.0	3	2019
			cChm6ilOrFnAffbfqTq	763207875	240472	2	20191223	1300.0	3	2019
			bi4WsnOfYTskGya9BiT	457335642	731784	2400	20190206	1340.0	3	2019
			4YbGneADMTO9MMY9mh3	123376998	66094266	0	20190412	0.0	3	2019

Mögliche Analysefehler und Fehlinterpretationen

✗ „Häufigkeit ambulanter Behandlung“

→ gezählt: **Leistungen**, interpretiert als **Personen**

✗ „Mehr Fälle = mehr erkrankte Menschen“

→ kann auch bedeuten: **mehr Kontakte pro Person**

✗ „Diagnose kodiert = Person ist krank“

→ Kodierung kann Verdacht (V), Ausschluss (A) oder Zustand nach (Z) sein

→ Auch gesichert (G) heißt nicht: klinisch schwer oder aktuell behandlungsbedürftig

✗ „Mehr Diagnosen in Region A = schlechtere Gesundheit“

→ Kann heißen: bessere Versorgung, höhere Arztinanspruchnahme, unterschiedliche Altersstruktur, urban vs. ländlich

Interpretationshinweis

Gleicher Datensatz – andere ID – andere Aussage

Warnung

Verdachtsdiagnosen
nicht versehentlich als
Prävalenz zählen.

Mögliche Analysefehler und Fehlinterpretationen

✗ „Keine Rezept-Verordnung = keine Erkrankung“

→ Kann heißen:

Selbstmedikation, Krankenhausbehandlung,

andere Therapiestrategie, Nicht-Inanspruchnahme

✗ „Diagnosedatum = Zeitpunkt des Ereignisses“

→ Häufig Dokumentations- oder Abrechnungsdatum

✗ „Fallzahlsteigerung über die Jahre = epidemiologischer Trend“

→ Kann heißen: bessere Erkennung, Screeningprogramme, veränderte Leitlinien, mehr Dokumentation

Warnung

Ganz generell:

✗ „Diese Daten zeigen Krankheit“

→ Sie zeigen **Inanspruchnahme von Versorgung**

Für die Planung vor dem Antrag ...



Informationsquellen des FDZ und Unterstützungsangebote



ÜBERBLICK INFOPORTAL SCHULUNG PROJEKTE AKTUELLES SUCHE 🔍

Links

<https://www.forschungsdatenzentrum-gesundheit.de/infoportal/externe-seiten-auf-einen-blick>

Antragsportal des FDZ Gesundheit

Reichen Sie Ihren Antrag beim FDZ Gesundheit einfach und sicher über unser digitales Antragsportal ein. Dort führen wir Sie Schritt für Schritt durch den gesamten Antragsprozess.

ZUM
ANTRAGSPORTAL

Datensatzbeschreibung

Nutzen Sie zur Vorbereitung Ihres Antrags die Datensatzbeschreibung. So erkennen Sie schnell, welche Variablen für Ihr Forschungsvorhaben relevant sind und können diese im Antrag gezielt und präzise benennen.

ZUR
DATENSATZBE-
SCHREIBUNG

Public Use File

Das Public Use File (PUF) bildet die Struktur des Echtdatensatzes ab. Es ermöglicht Ihnen, bereits vor der Antragstellung erste Auswertungsskripte zu entwickeln und methodische Abläufe zu testen.

ZUM PUBLIC USE FILE
DOWNLOAD

Statistikportal

Nutzen Sie das Statistikportal des FDZ Gesundheit, um die Machbarkeit Ihrer Forschungsfragen frühzeitig zu prüfen. So erhalten Sie transparente und verlässliche Hinweise zur Datengrundlage.

ZUM
STATISTIKPORTAL

Informationsquellen des FDZ und Unterstützungsangebote

Externe Angebote

Wenn Sie Ihr Angebot hier auflisten möchten, wenden Sie sich bitte an den Helpdesk.

Trainingsangebot

Wenn Sie die FDZ Forschungsdatenbasis genauer kennen lernen oder sich darauf vorbereiten wollen, effiziente Abfragen zur Ihrer Forschungsfrage zu erstellen, können Sie hier spezifische Trainingsangebote, Jupyter Notebooks und einen SQL CoPilot der DCC Risikoanalytik nutzen.

ZUM
TRAININGSANGEBOT

<https://www.forschungsdatenzentrum-gesundheit.de/infoportal/externe-seiten-auf-einen-blick>

Zeit für Ihre Fragen!



Kontakt

Bundesinstitut für Arzneimittel und Medizinprodukte
Forschungsdatenzentrum
Kurt-Georg-Kiesinger-Allee 3
53175 Bonn

Dr. Melanie Ludwig
helpdesk@forschungsdatenzentrum-gesundheit.de
www.bfarm.de
www.forschungsdatenzentrum-gesundheit.de

