

Big Data in Versorgung und Forschung – Herausforderungen und Perspektiven für die Digitalisierung im Gesundheitswesen

Big Data Konferenz | Berlin | 13.12.2016

Sebastian Claudius Semler

TMF – Technologie- und Methodenplattform
für die vernetzte medizinische Forschung e.V., Berlin

„Big Data“?



“**Big Data** is high-**volume**, high-**velocity** and/or high-**variety** information assets that demand cost-effective, **innovative forms of information processing** that **enable enhanced insight**, decision making, and process automation.”

Gartner IT Glossary

→ Nicht jede große Datenmenge ist also „Big Data“ im eigentlichen Sinne !

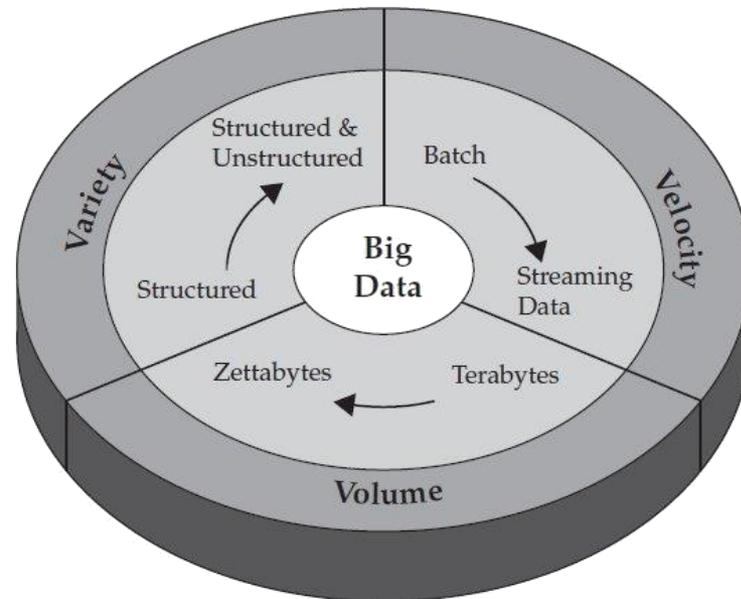


Figure 1-1 IBM characterizes Big Data by its volume, velocity, and variety—or simply, V³.

„Big Data“: „The Big V’s“ and the „Wanna Vs“ Debate



Doug Laney, META Group/Gartner, 2001:

1. „Volume“ (Datenmenge)
2. „Velocity“ (Datenfluss)
3. „Variety“ (Datenvielfalt)

IBM, 2012:

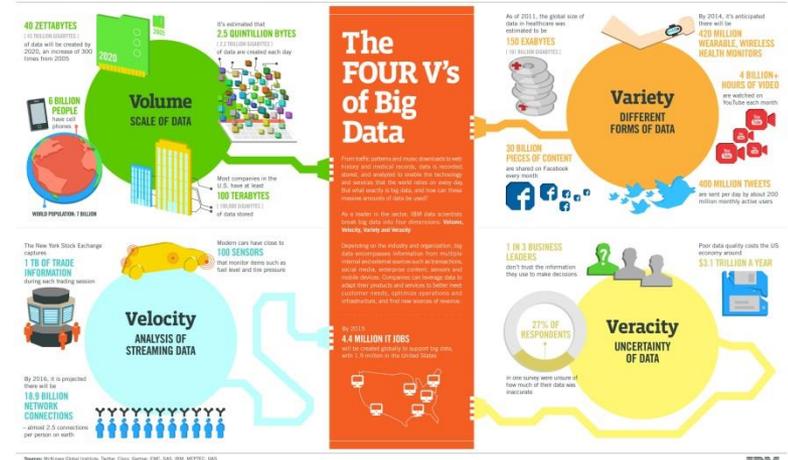
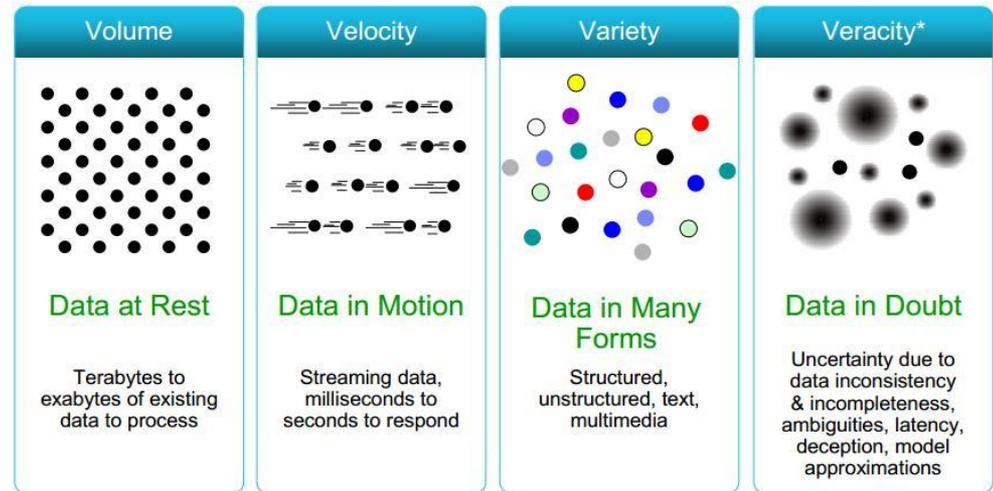
4. „Veracity“ (Richtigkeit)

PROS, Biehn 2013:

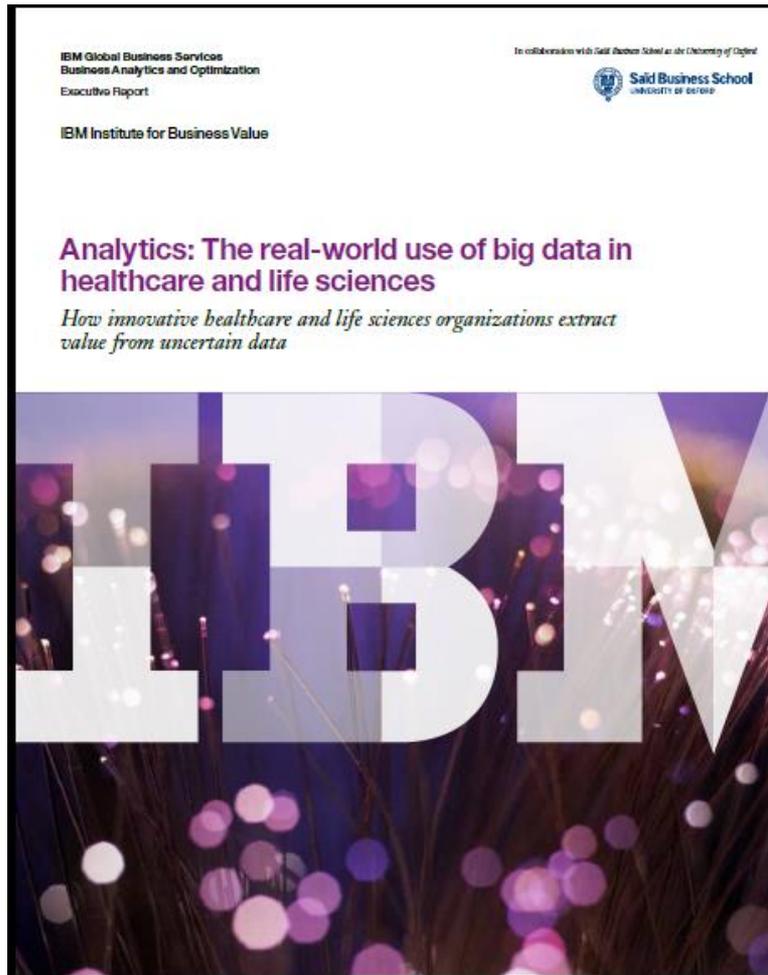
5. „Viability“ (Brauchbarkeit)
6. „Validity“ (Gültigkeit)
7. „Value“ (Wert/Nutzen)

Weiterhin (2012 ff.):

„Volatility“, „Variability“,
„Validity“, „Virality“,
„Vagueness“, „Visualization“,
„Vocabulary“, „Venue“ [...]



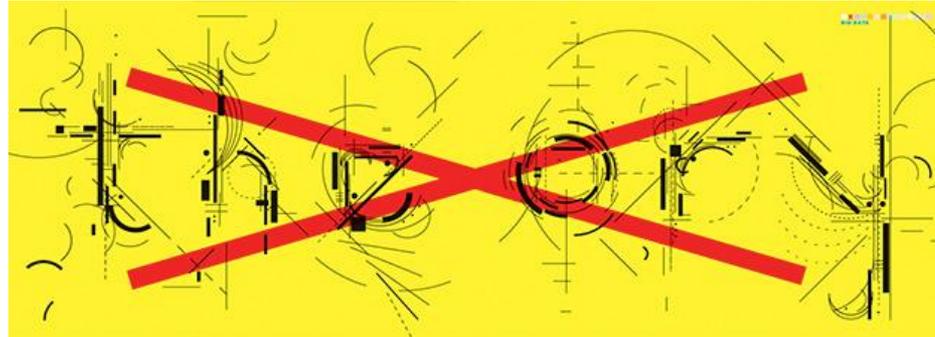
„Big Data“ im Gesundheitswesen?!



Big Data – Zukunft der Wissenschaft (?)



entnommen aus:
Prof. M.Krawczak, „Big
Data – Hype or Hope“,
conhIT 2016



© Wired Magazine

“All models are wrong, but some are useful.”

George Box, Statistiker, University of Wisconsin

“All models are wrong, and increasingly you can succeed without them.”

Peter Norvig, Director of Research, Google Inc.

“Correlation supersedes causation, and science can advance even without coherent models, unified theories, or really any mechanistic explanation at all.”

Chris Anderson, Chief Editor, Wired Magazine

„Big Data“ im Gesundheitswesen – um welche Daten geht es?

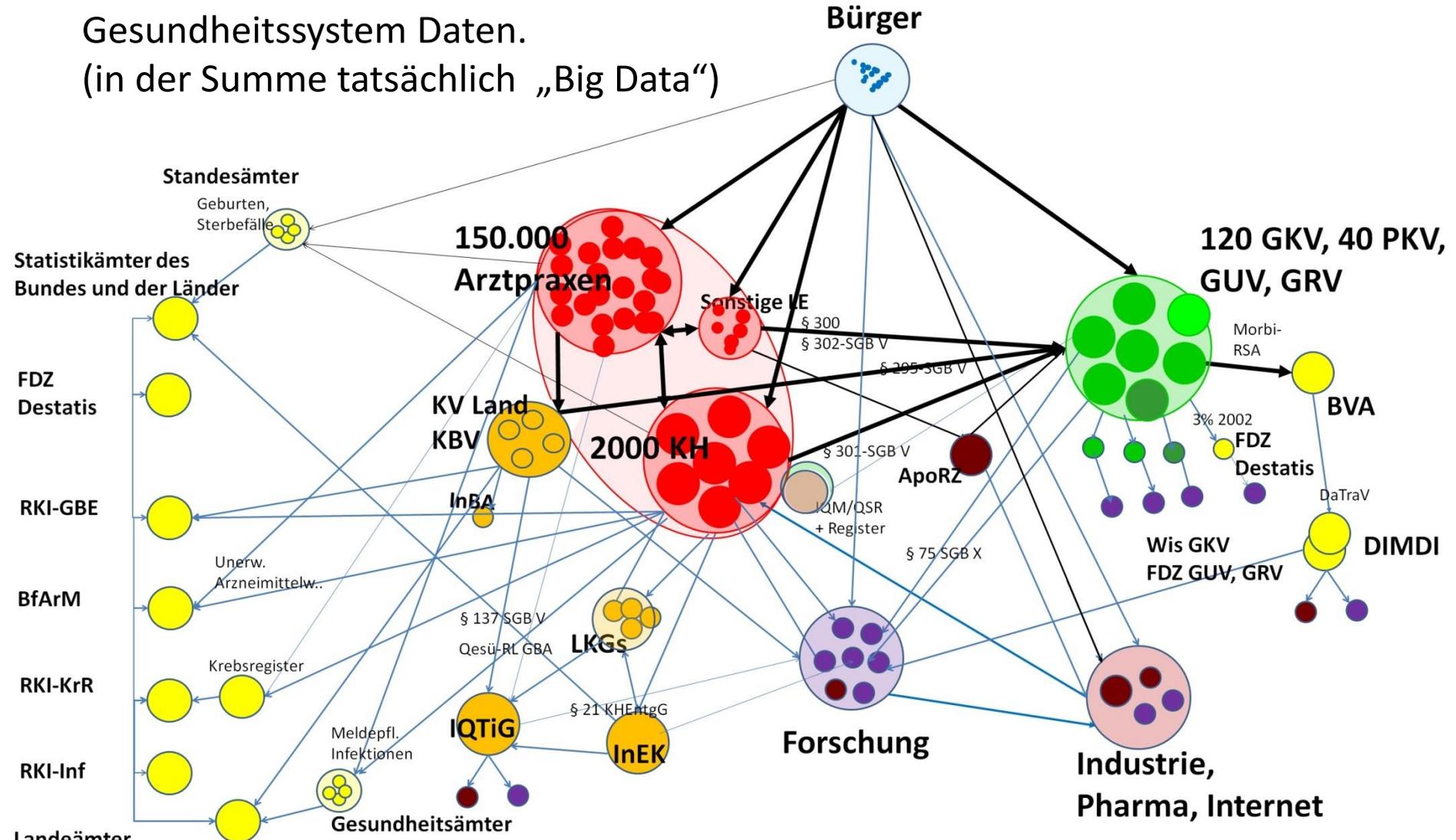
- ▶ Administrative Massendaten (Routinedaten, Sozialdaten)
- ▶ Öffentliche Melde- und Statistikdaten
- ▶ Behandlungen(massen)daten in den elektronischen Dokumentationssystemen der Leistungserbringer?
 - ▶ Auch: strukturierte phänotypischer Daten
 - ▶ Stammen z.T. auch aus der klinischen und epidemiologischen Forschung
 - ▶ („*unused big data*“ nach Kroemer)
- ▶ advanced analytical technology - Analysedaten (OMICs, Sequencing, Imaging)
 - ▶ (= „*conventional Big Data*“ nach Kroemer)
- ▶ Daten und Profile aus Sensoren / Apps / Wearables bei den Patienten / Bürgern
 - ▶ (= „*private Big Data*“ nach Kroemer)
- ▶ gesundheitsbezogene gewerbliche Nutzerdaten im Internet
- ▶ ... Kombination und Integration von all diesem?

Gesundheitsdaten (und E-Health-Infrastrukturen) sind die „stille Reserve“ der medizinischen Forschung und Gesundheitssteuerung.

Dies hängt nicht an der Frage, wie „Big“ - im Sinne der Definition - die „Data“ sind!

Defizite bei Datenzugang und Datennutzung

- ▶ Es existieren an vielen Stellen im Gesundheitssystem Daten. (in der Summe tatsächlich „Big Data“)



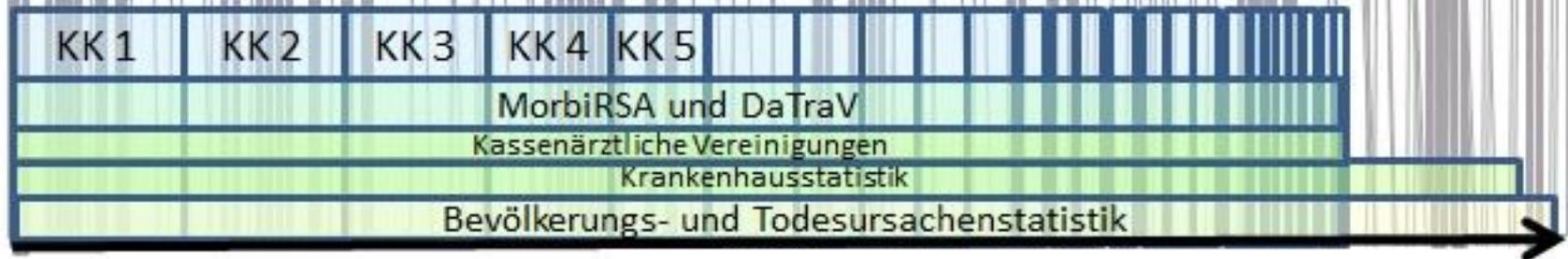
Defizite bei Datenzugang und Datennutzung



► Diese Daten weisen sehr unterschiedliche Datenbreite und -tiefe auf.

Informationenfülle' je Bürger

Dokumentationen der Leistungserbringer
sowie Register, Kohortendokumentationen und Surveys



Bürger/Versicherte

Defizite bei Datenzugang und Datennutzung

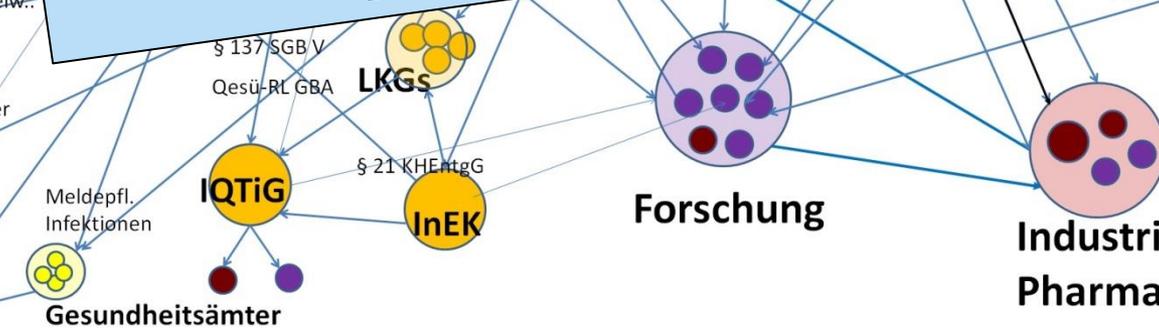
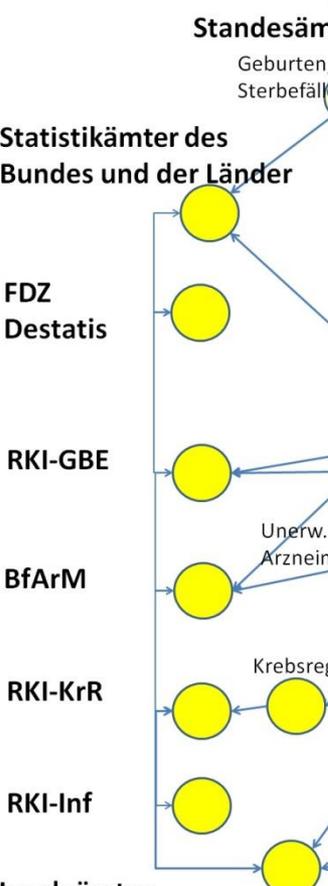
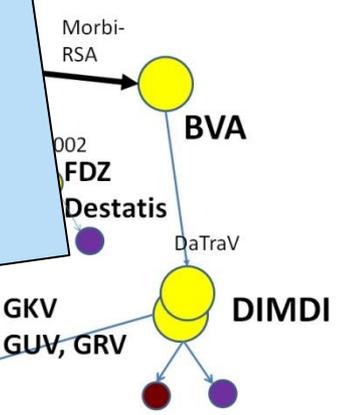
- ▶ Es existieren an vielen Stellen im Gesundheitssystem Daten.
(in der Summe tatsächlich „Big Data“)

Bürger

siehe Gutachten der TMF
(und Fraunhofer Big Data Allianz et al.)
für das TAB-Büro beim Dt. Bundestag 2015,
Veröffentlichung im TAB-Bericht Anfang 2017

(Auszug im TAB-Brief Nr. 47, Juli 2016, S.34 ff.,
K. Gerlinger: „Data Mining“,
<https://www.tab-beim-bundestag.de/de/pdf/publikationen/tab-brief/TAB-Brief-047.pdf>)

120 GKV, 40 PKV, 1000 GKV, 1000 PKV, 1000 GKV, 1000 PKV



- ▶ Allein bereits vorhandene Daten (besser) nutzbar zu machen, böte große Chancen für Versorgungsforschung und Gesundheitssteuerung / Gesundheitssystementwicklung.
- ▶ Aber:

Verfügbarkeit



Verwertbarkeit

Verknüpfbarkeit

entnommen aus: Prof. M.Krawczak, „Big Data – Hype or Hope“, conhIT 2016

- ▶ **Verfügbarkeit**
 - ▶ Praktisch – organisatorisch („Wissen um“ ...)
 - ▶ Technisch
 - ▶ Rechtlich (!)

- ▶ **Verknüpfbarkeit**
 - ▶ semantisch
 - ▶ technisch
 - ▶ (datenschutz)rechtlich (!)

- ▶ **Verwertbarkeit**
 - ▶ Datenqualität
 - ▶ ... und deren Beurteilbarkeit / Nachvollziehbarkeit
 - ▶ semantisch
 - ▶ (auch rechtlich)

(Versorgungs)Forschung anhand von „REAL WORLD DATA“ aus der Patientenversorgung



Methoden der empirischen Sozialforschung: [nach Pfaff 2003]

- ▶ Befragungsmethoden
- ▶ Beobachtungsmethoden
- ▶ Methoden der *Dokumentenanalyse*



obligat: (neben der Erhebung zusätzlicher Daten)

Analyse von Sekundärdaten

= aus der Routineversorgung erhobene Daten!

(Routinedaten i.w.S.)

3 Möglichkeiten:

- ▶ Analyse von (elektronischen) **Krankenakten**
- ▶ Auswertung von **Daten der Krankenkassen und Kven („Sozialdaten“)**
- ▶ Analyse von Daten aus (versorgungsbezogenen) **Registern**

Hürden und Herausforderungen in der Datennutzung für die Versorgungsforschung (1/2)



1. Methodische Fragen

- ▶ z.B. Beurteilbarkeit von Datenqualität
- ▶ z.B. „Verfallszeit“ von Daten und Ergebnissen
- ▶ siehe Arbeiten des DNVF!

auch organisatorische Aspekte
(z.B. Conflict of Interest in der VF)

2. Incentives für qualitative strukturierte Datenerhebungen ?!

3. Finanzierung einer kontinuierlichen Versorgungsforschungsinfrastruktur ?

Hürden und Herausforderungen in der Datennutzung für die Versorgungsforschung (2/2)



4. Datenstandardisierung

- ▶ (mehr als die Frage einer SNOMED-Lizenz!)
- ▶ Reale Nutzung im Versorgungsalltag über die Verpflichtungen gemäß SGB V vorantreiben!
- ▶ Internationale Standards nutzen aus Versorgung (HL7/LOINC/IHE/SNOMED?) und klinischer Forschung (CDISC, MedDra)

5. Rechtliche und datenschützerische Hürden

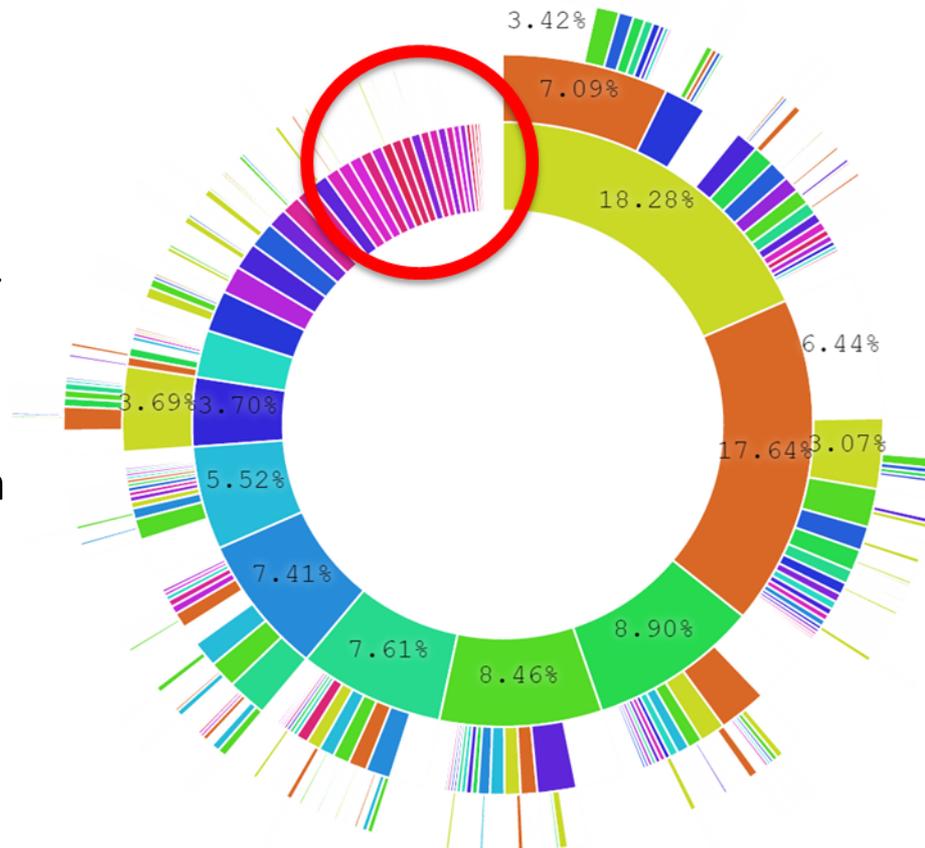
- ▶ Erste gesetzliche Regelungen (DaTraV, Krebsregistergesetz)
- ▶ z.B. Unterschiedliche Forschungsklauseln in den Landeskrankengesetzen
- ▶ praktische Bedeutung des BSG-Urteils 2008 zur Anwendung des SGB im Bereich Datenschutz
- ▶ Orientierungshilfe der Landesdatenschutzbeauftragten zu KIS
- ▶ § 291a SGB V, Abs. 8 als pauschale Hürde zur Nutzung der Sicherheitsinfrastruktur für Forschungsfragen !
- ▶ EU-Datenschutzgrundverordnung – Umsetzung und gesetzliches Zusammenwirken mit Bundes- und Landesgesetzen

Welche Erkenntnisse die Nutzung von Real-Versorgungsdaten bringen?



Beispiel entnommen aus: Prof. G.Hripcsak (Columbia Univ., NY) „Observational Health Data Sciences and Informatics (OHDSI)“, GMDS-TMF-WS Juli 2016

- ▶ Beispiel: Reale Patienten-Heterogenität in der Bluthochdruck-Therapie
- ▶ Datenbasis: Patientenakten von 2 Mio. Patienten



Lisinopril	18.28%
Hydrochlorothiazide	17.64%
Amlodipine	8.90%
Metoprolol	8.46%
Atenolol	7.61%
Furosemide	7.41%
Ramipril	7.09%
Bendroflumethiazide	6.44%
Losartan	5.52%
valsartan	3.70%
Triamterene	3.69%
olmesartan	3.42%
benazepril	3.07%
Diltiazem	3.07%
carvedilol	3.07%
Bisoprolol	3.07%
Doxazosin	3.07%
Enalapril	3.07%

25% of HTN patients (10% of others) have a unique path despite 250M pop

Dank

Weitere Informationen unter www.tmf-ev.de

Kontakt

sebastian.semmler@tmf-ev.de

Geschäftsstelle TMF e.V.

Charlottenstraße 42/Dorotheenstraße
10117 Berlin

+49 (30) 22 00 24 70

info@tmf-ev.de

www.tmf-ev.de | [@TMF_eV](https://www.instagram.com/TMF_eV)