@eHealthSummitAT

eHEALTH SUMMIT AUSTRIA

HIMSS Europe

In Kooperation mit

eHEALTH 2017 www.eHealth2017.at

23. – 24. MAI 2017

Schloß Schönbrunn, Wien
Apothekertrakt und Orangerie

Big Data for Public Health – Public Data for Big Health: Datengrundlagen, Datennutzung und Datenlimitationen

Holger Gothe, Dr. med. Dipl.-Komm.wirt, IGES Institut, Berlin, Deutschland

holger.gothe@iges.com

Präsentiert von

HIMSS CHIME INTERNATIONAL

AIT AUSTRIAN INSTITUTE OF TECHNOLOGY TOMORROW TODAY

Österreichische Gesellschaft für Biomedizinische Technik
Austrian Society for Biomedical Engineering

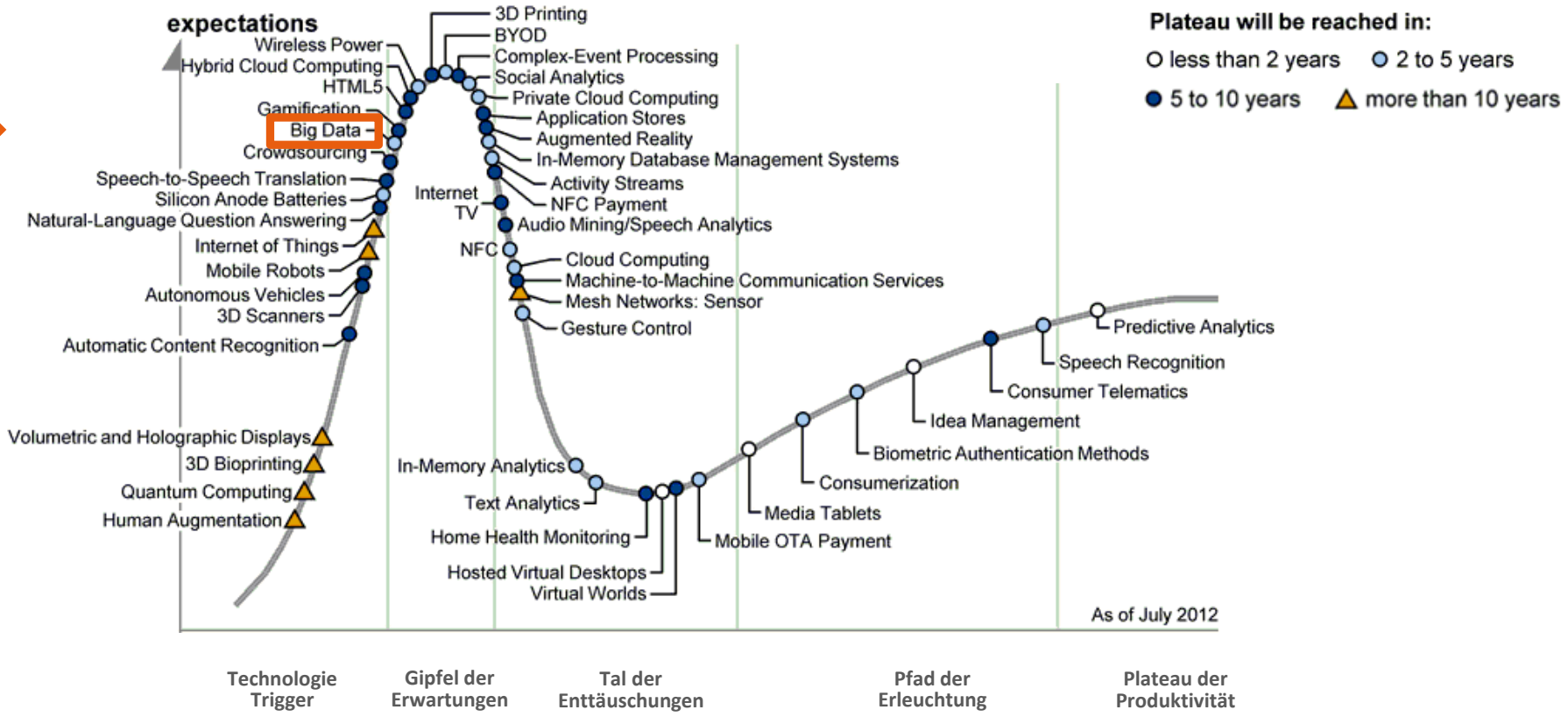
OESTERREICHISCHE COMPUTER GESELLSCHAFT
AUSTRIAN COMPUTER SOCIETY

UMIT

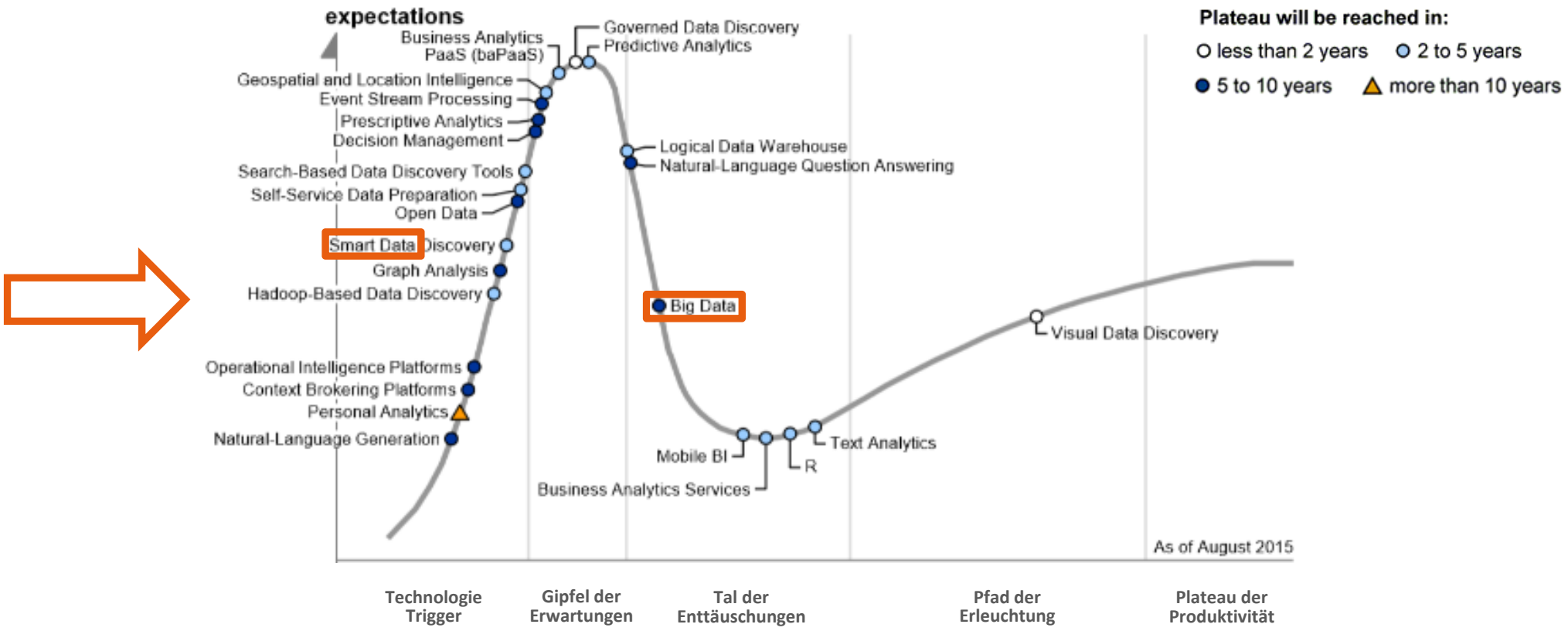
AGENDA

- Worüber sprechen wir, wenn **»Big Data«** im Gesundheitswesen Thema sind?
- Welche **Datengrundlagen** für Public Health gibt es heute?
- Wofür können diese **Daten genutzt** werden?
- Welche **Aussagekraft** haben die Daten?
- Welche **Limitationen** weisen die Daten auf?

Gartner‘s Hype Cycle for Business Intelligence and Analytics (2012)



Gartner‘s Hype Cycle for Business Intelligence and Analytics (2015)



»Big Data« ???

- ≈ 40 Mio. Personen
- aus allen Regionen (West-)Deutschlands
- aus allen Altersgruppen
- aus allen Berufsgruppen
- mit allen ambulanten medikamentösen Verordnungen, die über Krankenkassen abgerechnet werden

frühe Arzneiverordnungsreporte aus dem deutschen GKV-System

„Routinedaten“ und „Sekundärdaten“

- „***Routinedaten*** sind prozessproduzierte, umfangreiche Informationssammlungen, die im Rahmen der Verwaltung, Leistungserbringung bzw. Kostenerstattung (z. B. bei der gesetzlichen Kranken-, Renten- oder Unfallversicherung) anfallen. …
  Sie werden … unter dem Begriff ***Sekundärdaten*** zusammengefasst.“ (Hoffmann & Glaeske 2011)

- ***Sekundärdaten*** sind Daten, die nicht direkt erhoben wurden, sondern aus ***Primärdaten*** durch Modellierungs- oder Verarbeitungsschritte hervorgehen.

Erfüllen Sekundärdaten die konstitutiven Dimensionen von »Big Data«?

• ***Volume***	Umfang, Datenvolumen	**≤ 70 Mio. Personen**
• ***Velocity***	Geschwindigkeit, mit der die Daten generiert und transferiert werden	**„sofort“ verfügbar**
• ***Variety***	Bandbreite der Datentypen und Datenquellen	**Makro-/Meso-/Mikro-Ebene**
• ***Value***	Wert, Aussagekraft	**Populationsbezug, lange Zeitreihen**
• ***Validity***	Gültigkeit	**Real World (externe Validität)**
• ***Veracity***	Unvollständigkeit, Inkonsistenzen	**Limitationen**

Quelle: Laney (2001); Bachmann, Gerzer, Kemper (2014)

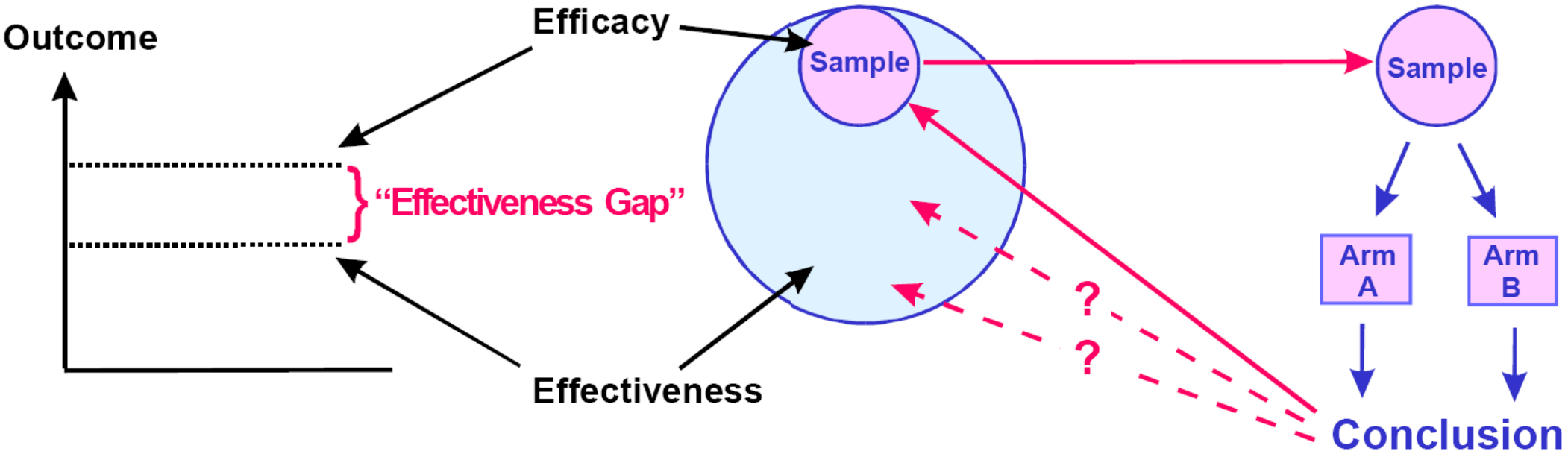
Welche Datengrundlagen für Public Health gibt es heute? ⇨ *Variety*

- ***Makroebene***
  - Bundesversicherungsamt (BVA), Datentransparenzverordnung (DaTraV)
  - Unfall- und Krankenversicherungen, gesetzlich und privat (GKV/PKV)
  - Amtliche Statistiken und krankheitsbezogene Register
- ***Mesoebene***
  - Kassenärztliche Vereinigungen (KVen/KBV), Zentralinstitut der kassenärztlichen Versorgung (ZI)
  - Apothekenrechenzentren (ARZ)
  - Institut für das Entgeltsystem im Krankenhaus (InEK)
  - Spezialisierte Panels von Marktforschungsinstituten (IMS, Insight Health)
- ***Mikroebene***
  - Arztpraxen und medizinische Versorgungszentren (MVZ)
  - Krankenhäuser und rehabilitative Einrichtungen
  - Medizinische Labore

Wofür können die Daten genutzt werden? ⇨ *Value*

- ***Übergeordnete Ziele***
  - Steigerung der „Disease Awareness“
  - Schließung von Lücken in der jeweils aktuellen Evidenzlage
  - Informed Decision Making für Politik, Kostenträger und Leistungserbringer
- ***Spezifische Ziele von Analysen großer (Sekundär-)Datensätze***
  - Charakterisierung und Typisierung von Patienten (Soziodemographie, Komorbidität)
  - Schätzung der Krankheitshäufigkeit (Prävalenz, Inzidenz)
  - Darstellung von Inanspruchnahmemustern und Inanspruchnahmeintensität
  - Gesundheitsökonomische Aspekte (z.B. Kosten der Therapie)
  - Beschreibung von Versorgungstrajektorien
  - Identiifikation potentieller „Fehldiagnosen“
  - Analyse der „diagnostischen Odyssee“
  - …

Welche Aussagekraft haben die Daten? ⇨ *Validity*



Welche Limitationen weisen die Daten auf? ⇨ ***Veracity***

- ***Vorteil*** der ***Provenienz der Daten*** (Behandlungsalltag, Versorgungs-routine) ist zugleich ***Nachteil*** (Inanspruchnahme des Versorgungssystems, Datenkollekte zu Prozesszwecken, z. B. Leistungsabrechnung)
- Daten sind weitestgehend frei von ***Untersuchungs-Bias***, im Zusammen-hang mit dem sekundären Verwendungszweck (wiss. Analyse und Inter-pretation) allerdings nicht gänzlich ohne Verzerrungen, wie beispielsweise ***Missclassification*** oder ***Confounding***.
- ***Datenhandling*** aufwändig (Hardware-/Software-Ausstattung)
- Expertise bzgl. ***Datenschutz*** und ***Datenbereinigung / Plausibilisierung***
- Methodische Kompetenz in der ***Auswertung*** und ***Interpretation***

»Big Data«  »Smart Data« ???