

Vergleichbare Auswertungsergebnisse über Änderungen in Klassifikationen hinweg ohne den Einsatz von Approximationen

C. Lüpkes, OFFIS - Institut für Informatik, Oldenburg

Wie sich Datenanalysen über geänderte Klassifikationen wie ICD einfach mit dem Tool GrAHD¹ korrekt durchführen lassen

I Motivation Datenanalyse

- Auswertungen von über einen längeren Zeitraum erfasster Daten sind oft schwierig, da sich Klassifikationen (z.B. ICD oder Gemeindegemeinschaften) ändern (z.B. Einfügen zusätzlicher Codes, Änderung der Semantik, Restrukturierung).
- Beispiel ICD-10-GM: In der Version des Jahres 2010 wurde die Semantik des Codes C80 geändert. Es wurden zwei Sub-Codes (C80.0 & C80.9) hinzugefügt und zu codierende Inhalte in C79.9 verschoben.
- Auswirkung: Eine Analyseanfrage zu dem ICD-Code C80, z.B. der Jahre 2006 bis 2010, würde falsche Ergebnisse liefern, da je nach verwendetem Analysesystem entweder sowohl C80.0 und C80.9, aber in jedem Fall C79.9 unberücksichtigt bliebe.

- C80 2009: Bösartige Neubildung ohne Angabe der Lokalisation: Inkl. {Generalisierter: Krebs & maligner Tumor, Karzinom, Karzinose, Krebs, Maligner Tumor, Multipler Krebs, nicht näher bezeichneter Lokalisation (primär) (sekundär)}, {Bösartige Neubildung ohne Angabe der Lokalisation}
- C80 2010: Bösartige Neubildung ohne Angabe der Lokalisation; C80.0 Bösartige Neubildung, primäre Lokalisation unbekannt, so bezeichnet; C80.9 Bösartige Neubildung, nicht näher bezeichnet {Karzinom o.n.A., Krebs o.n.A., Maligner Tumor o.n.A.}
- C79.9 2010: Sekundäre bösartige Neubildung nicht näher bezeichneter Lokalisation: Generalisiert (sekundär): {Krebs o.n.A., Maligner Tumor o.n.A.}, {Karzinose (sekundär)}, {Multipler Krebs (sekundär) o.n.A., Exkl. Disseminierte maligne Neoplasie, ohne Angabe eines Primärtumors (C80.-)}

II Graphenbasierter Ansatz

- Klassifikationen lassen sich als Graphen beschreiben und Änderungen (sowohl syntaktische als auch semantische) als verbindene Kanten zwischen diesen Graphen.
- Durch geeignete Auswertung der Kanten lassen sich zusammenhängende Mengen von Klassifikationsknoten (z.B. ICD Codes bei ICD Klassifikationen) identifizieren, die über verschiedene Klassifikationen hinweg die gleiche Semantik repräsentieren. Wenn es mehrere Interpretationen eines Codes in verschiedenen Klassifikationen gibt, werden alle identifiziert.
- Beispiel ICD-10-GM: Für eine Anfrage nach C80 im Zeitraum 2008 bis 2011 gibt es zwei Semantiken. Die bis 2009 gültige und die ab 2010 gültige Semantik von C80, die sich durch C79.9 Inklusion/Exklusion inhaltlich unterscheiden.

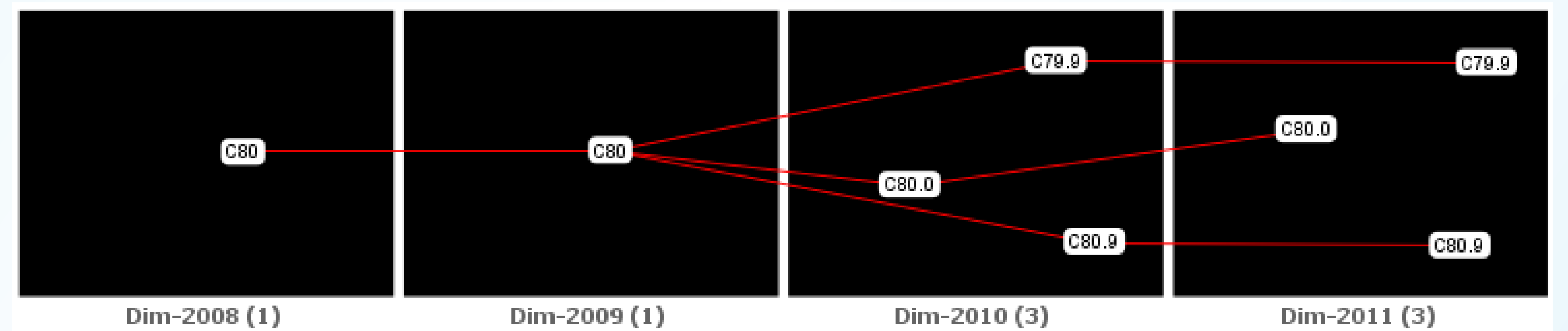


Abbildung 1: Darstellung der Entwicklung des ICD-10-GM Klassifikationsknoten C80 in der Bedeutung des Jahres 2008 im Zeitraum bis 2011

III GrAHD¹

- Das unter dem Namen GrAHD¹ erforschte und als Demonstrator umgesetzte Verfahren wurde u. A. mit den ICD-10-GM Klassifikationen der Jahre 2006 bis 2012 getestet. Hierbei wurden die vom DIMDI herausgegebenen Überleitungen interpretiert. Als Datenbasis standen die maschinenlesbaren strukturierten Qualitätsberichte (SQB) der Jahre 2006, 2008 und 2010 in einem DataWarehouse zur Verfügung.
- GrAHD visualisierte die Änderungen und generierte SQL-Statements zur Anwendung in Analysefragen. Die Nützlichkeit wurde in mehreren Evaluationen durch Anwender mit Expertise in medizinischer Dokumentation und Datenanalyse bestätigt. Bereits die Identifizierung verschiedener Bedeutungen von ICD Codes wurde als hilfreich angesehen.

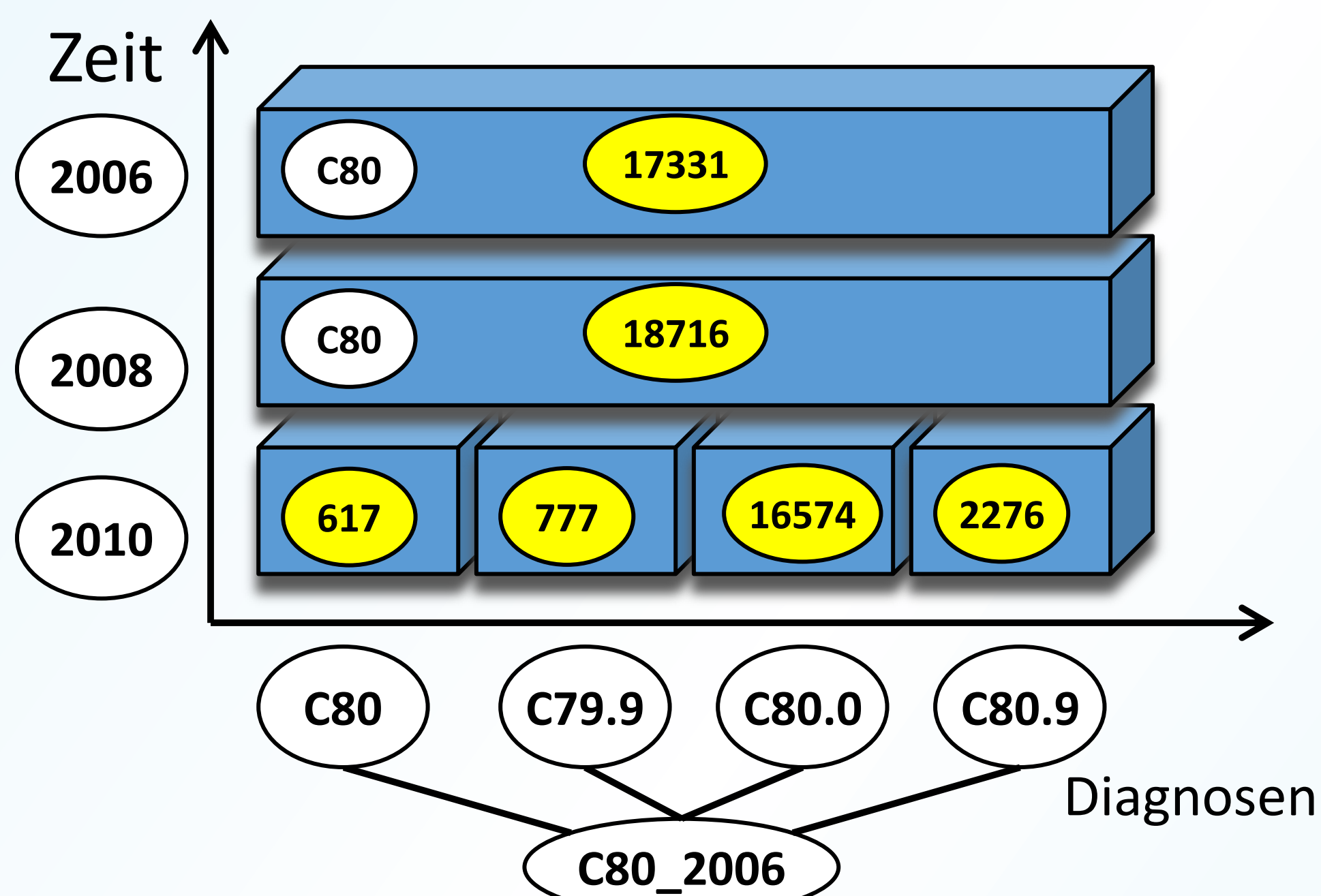


Abbildung 2: Summierten Fallzahlen (Hauptdiagnosen) der SQB aller im jeweiligen Berichtsjahr aktiven Krankenhäuser in DWH Visualisierung. Datenschutzwerte wurden auf Empfehlung des IGES mit 3,5 in die Berechnung einbezogen.

IV Diskussion & Ausblick

- Bei der Anwendung von GrAHD adaptierten Anfragen auf den SQB-Bestand ließen sich Indizien für bessere Ergebnisse über geänderte Klassifikationen hinweg finden.
- GrAHD ist für Datengrundlagen geeignet, bei denen bekannt ist nach welcher Version einer Klassifikation die Daten erfasst wurden. Eine kennzahlabhängige Approximation zur Umcodierung einzelner Werte wird nicht benötigt. Da GrAHD bereits für ICD Überleitungen getestet wurde, ist eine entsprechende Anwendung einfach möglich, weil die Überleitungen bereits existieren und der Datenbestand lediglich als Zusatzinformation die jeweilige Version enthalten muss. Es wurden bereits ICD-9-CM auf ICD-10-CM wie die Entwicklung der ICD-10-GM getestet.
- Da der Ansatz aus der IT heraus entwickelt wurde, besteht Bedarf an einer weitergehenden inhaltlichen Überprüfung durch die Epidemiologie, bzw. Versorgungsforschung.

Fallzahlen C80 \ Jahr	2006	2008	2010
ohne GrAHD (Veränderung zum vorherigen SQB)	17331	18716 (+ 8%)	617 (-96,7%)
mit GrAHD (Bedeutung 2004-2009) (Veränderung zum vorherigen SQB)	17331	18716 (+8%)	20244 (+8%)

Basierend auf den summierten Fallzahlen der SQB aller im jeweiligen Berichtsjahr aktiven Krankenhäuser. Datenschutzwerte wurden auf Empfehlung des IGES Instituts mit 3,5 in die Berechnung einbezogen.

Abbildung 3: Entwicklung der Fallzahlen bei summierten Hauptdiagnosen des Codes C80 der SQB in den Jahren 2006 bis 2010

Autor und Kontakt:

Dr.-Ing. Christian Lüpkes
Christian.Luepkes@offis.de

¹ GrAHD (Graphenbasierte Ad-Hoc-Datentransformation) wurde im Rahmen des Promotionsverfahrens von C. Lüpkes erforscht und als Anwendung umgesetzt.

Lüpkes, Christian (2014) (Dissertation): Ad-hoc-Datentransformationen für Analytische Informationssysteme. 29th ed. Edewecht: OIWR Verlag für Wirtschaft, Informatik und Recht, ISBN 978-3-95599-005-3