



Elektronische Kodierung von Todesbescheinigungen

Einleitung

Die Verbesserung der Qualität und die Erhöhung des Grades der Digitalisierung sind zentrale Herausforderungen bei der Erstellung der deutschen Todesursachenstatistik. Ein großer Teil der deutschen Todesbescheinigungen wird noch immer manuell für die Statistik aufbereitet.

Die Todesursachenstatistik wird von den statistischen Ämtern auf Landesebene nach einem internationalen Standard ausgewertet. Dieser Standard wird von der Weltgesundheitsorganisation (WHO) in der Internationalen statistischen Klassifikation der Krankheiten und verwandter Gesundheitsprobleme (ICD-10) festgelegt und regelmäßig aktualisiert. In die Statistik fließt für jeden Sterbefall das sogenannte Grundleiden ein, das laut WHO definiert ist als [1]:

- die Krankheit oder Verletzung, die den Ablauf der direkt zum Tode führenden Krankheitszustände auslöste, oder
- der Umstand des Unfalls oder der Gewalteinwirkung, der den tödlichen Ausgang verursachte.

Da das Regelwerk der WHO, das von Signierfachkräften anzuwenden ist, äußerst komplex ist, benutzen weltweit immer mehr statistische Ämter das elektronische Kodiersystem Iris. Der Kodierkern MUSE (Multicausal and Unicausal Selection Engine) ist eine Softwarekomponente, die seit 2013 von einigen deutschen statistischen Ämtern und seit 2017 auch im internationalen Kodiersystem Iris verwendet wird [2, 3]. Die Signierentscheidungen von MUSE beruhen auf den sogenannten Entscheidungstabellen der Mortalitätsstatistik, die von der Iris Core Group (Entwicklergruppe des internatio-

nenal Kodiersystems) jährlich nach den Vorgaben der WHO aktualisiert werden.

In diesem Beitrag wird zunächst der Stand der Digitalisierung der deutschen Todesursachenstatistik dargestellt. Nach der Beschreibung der Methodik zur Kodierung der Todesursachen wird das Verfahren zur Texterkennung vorgestellt, das neu in der deutschsprachigen Variante des internationalen Kodiersystems (nachfolgend als Iris/MUSE bezeichnet) implementiert ist. Ergebnisse der automatischen Kodierung und die Möglichkeit zur strukturierten Nachbearbeitung werden exemplarisch dargestellt. Die letzten drei Abschnitte beschäftigen sich mit der Bewertung des neuen Kodiervorgangs, der zu verbessernden Datenqualität der Todesursachenstatistik und möglichen Weiterentwicklungen des Verfahrens.

Elektronische Todesursachenstatistik in Deutschland

Im Jahr 2017 nutzten sechs¹ von 14 deutschen statistischen Ämtern auf Landesebene das elektronische Kodiersystem Iris/MUSE im Produktivbetrieb für die Erstellung der amtlichen Todesursachenstatistik. Somit werden inzwischen mehr als 50 % der deutschen Todesbescheinigungen vollständig elektronisch kodiert [4].

Eine vollumfängliche elektronische Bearbeitung von Todesbescheinigungen beinhaltet derzeit folgende Schritte:

- Elektronische Erhebung der Todesbescheinigungen

- Erfassung der medizinischen Einträge der Todesbescheinigungen über das Informationssystem des Gesundheitsamtes oder über Iris/MUSE im Statistischen Amt
- Elektronische Meldung an den „CORE-Dateneingang“ der statistischen Ämter des Bundes und der Länder [5]
- Datenimport in das Kodiersystem Iris/MUSE
- Kodierung aller auf den Todesbescheinigungen angegebenen Krankheiten und Zustände
 - Weitestgehend automatische Kodierung durch die Texterkennung
 - Manuelle Nachbearbeitung nicht automatisch kodierbarer Texte
 - Erzeugung der multikausalen ICD-10-Kodes (Abb. 1) durch den Kodierkern MUSE, welcher bei ca. 20 % der Scheine weitere Umkodierungen vornimmt² [3]
- Unikausale Signierung der Todesbescheinigungen
 - Ermittlung des Grundleidens³, das den Tod zur Folge hatte, mit dem Kodierkern MUSE auf Basis der kodierten multikausalen Todesursachen
 - Export der ICD-10-kodierten Ergebnisdaten aus Iris/MUSE in das Aufbereitungssystem BASIS+ der Bevölkerungs- und Todesursachenstatistik

² Z. B. wird der ICD-10-Kode für Kachexie (R64) in die Verschlüsselung für Tumorkachexie (C809R64) geändert, wenn Kachexie als Folge eines Karzinoms angegeben ist.

³ Das Grundleiden ist nach ICD-10 kodiert (siehe auch Abb. 1) und umfasst bei äußeren Ursachen (bspw. Unfällen, Morden, Suiziden) zusätzlich einen zweiten Code für die Klassifizierung der Hauptverletzung.

¹ Statistische Ämter in Baden-Württemberg, Berlin-Brandenburg, Nordrhein-Westfalen, Rheinland-Pfalz, Sachsen, Sachsen-Anhalt.

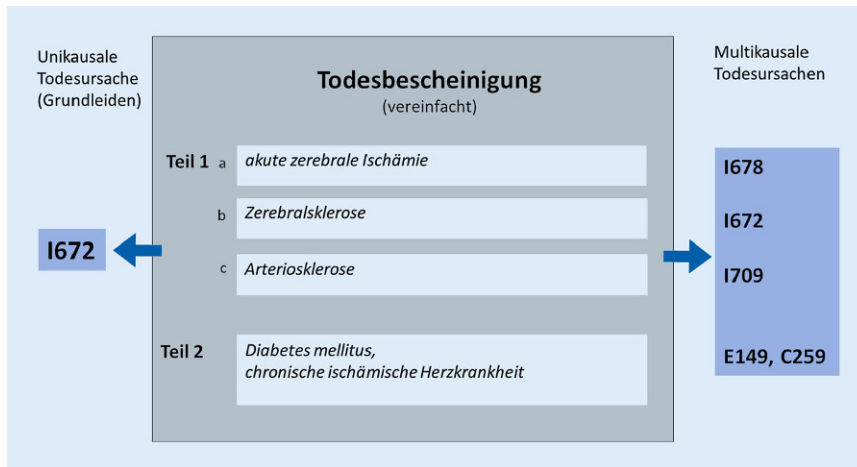


Abb. 1 ▲ Ableitung der unikausalen und der multikausalen ICD-10-kodierten Todesursachen (blaue Blöcke links und rechts) aus einer Todesbescheinigung (Mitte), wobei die Zeilen a bis c die Kausalkette des Teils 1 repräsentieren

- Erstellung der amtlichen Todesursachenstatistik
 - Vollzähligkeitsabgleich mit den Sterbefällen der Bevölkerungsstatistik inkl. Datenaustausch mit anderen Bundesländern [6]
 - Jährliche Publikation der amtlichen Todesursachenstatistiken auf Landes- und Bundesebene (bspw. auf www.destatis.de oder www.gbe-bund.de)

Kodierung multikausaler Todesursachen

■ **Abb. 1** verdeutlicht am Beispiel einer vereinfachten Todesbescheinigung den Unterschied zwischen der unikausalen Todesursache und den multikausalen Todesursachen. Teil 1 enthält die Kausalkette mit der direkt zum Tode führenden Krankheit („akute zerebrale Ischämie“) in Zeile 1a und der zugrunde liegenden Krankheit in Zeile 1c („Arteriosklerose“). In Teil 2 sind weitere Krankheiten angegeben, die zum Tod beigetragen haben. Die *multikausalen Todesursachen* sind die ICD-10-Kodes aller medizinischen Terme, die ärztlicherseits in Teil 1 und Teil 2 dokumentiert sind. Die *unikausale Todesursache* ist dagegen das Grundleiden, das nach Anwendung des Regelwerks der WHO ausgewählt wird. In diesem Beispiel wird nicht der Startpunkt der Kausalkette I709 (generalisierte und nicht näher bezeichnete Atherosklerose), sondern die spezifischere Kodierung I672

(zerebrale Atherosklerose) als Grundleiden ausgewählt.

Die Kodierung eines medizinischen Terms nach ICD-10 bedeutet grundsätzlich einen Verlust an Information, da ähnlichen Krankheiten (A), synonymen Krankheitsbezeichnungen (B) oder Krankheitsbezeichnungen mit zusätzlichen Detailangaben (C) derselbe ICD-10-Kode zugeordnet wird:

- A. Synonyme von Krankheiten (K65.9): „Peritonitis“ und „Bauchfellentzündung“
- B. Ähnliche Krankheiten (K65.8): „gallige Peritonitis“ und „kotige Peritonitis“
- C. Term mit bzw. ohne zusätzliche Details (C34.9): „Adenokarzinom der Lunge“ bzw. „Karzinom der Lunge“

Beispielhaft wird die multikausale Kodierung anhand der Todesbescheinigung einer 65-jährigen Frau erläutert:

- 1a: Septischer Schock
- 1b: Ischämische Kolitis und Ileitis
- 1c: Hochgradig stenosierende Arteriosklerose der Mesenterialarterien
- 2: Hochgradige allgemeine Arteriosklerose

Im alphabetischen Verzeichnis zur ICD-10 WHO 2019 [7] werden Einträge gesucht, die möglichst gut zu den angegebenen Diagnosetexten passen:

- „Schock, septisch“ (ICD-10-Kode **R572**) für Zeile 1a

- „Kolitis, ischämisch“ (ICD-10-Kode **K559**) für Zeile 1b (erster Term)
- „Ileitis, nichtinfektiös“ (ICD-10-Kode **K529**) für Zeile 1b (zweiter Term)
- „Arteriosklerose, Darmarterie“ (ICD-10-Kode **K551**) für Zeile 1c
- „Arteriosklerose, allgemein“ (ICD-10-Kode **I709**) für Zeile 2

Die Qualifizierungen „hochgradig stenosierend“ (1c) und „hochgradig“ (2) sind zusätzliche Details der ärztlichen Dokumentation und irrelevant für die Kodierung nach ICD-10.

Daraus resultiert die nachfolgende multikausale Kodierung:

R572 / K559 K529 / K551 * I709.

Erläuterung zur Notation der multikausalen Kodierung:

- Das Symbol / trennt die ICD-Kodes benachbarter Zeilen der Kausalkette in Teil 1.
- Kodes hinter dem Symbol * stehen für die zum Tod beitragenden Krankheiten oder Zustände aus Teil 2.
- Sind mehrere Codes einer Zeile zugeordnet, so sind diese durch Leerzeichen voneinander getrennt.

Verfahren der rechnergestützten Kodierung

Eine manuelle Verschlüsselung der multikausalen Todesursachen mittels der vom Deutschen Institut für Medizinische Dokumentation und Information (DIMDI) bereitgestellten Bücher ist zeitaufwendig und erfordert medizinisches Wissen sowie die korrekte Anwendung der komplexen Kodierregeln der WHO. Daher wurde ein Kodierverfahren entwickelt, das medizinische Texte möglichst gut automatisiert kodiert und lediglich unbekannte und fehlerbehaftete Texte zur manuellen Nachkodierung oder Überprüfung übrig lässt.

Exemplarisch wurden die Ergebnisse einer elektronischen Kodierung für $N=5687$ anonymisierte Todesbescheinigungen ausgewertet, die dem Statistischen Bundesamt für die Weiterentwicklung von Iris/MUSE bereitgestellt worden sind. Insgesamt sind 16.359 Text-

zeilen aus den Zeilen 1a, 1b, 1c und 2 in die Analyse eingeflossen:

- $N_{1a} = 5621$ (vorhanden bei 99,8 % der Scheine)
- $N_{1b} = 4032$ (70,9 %)
- $N_{1c} = 3268$ (57,5 %)
- $N_2 = 3438$ (60,5 %)

Im Durchschnitt sind 2,9 Zeilen pro Todesbescheinigung ausgefüllt.

Lexikalische Analyse medizinischer Wörter

Die Texterkennung der deutschen Version von Iris/MUSE enthält einen Scanner⁴, der die medizinischen Texte in elementare Zeichenketten (lexikalische Elemente oder Token) zerlegt. Jedem lexikalischen Element sind ein Typ, ein normierter Text und ggf. weitere Attribute wie Klassifikation, Synonym, Hyperonym (=Oberbegriff) und Grad der Kodierrelevanz zugeordnet. **Tab. 1** zeigt Beispiele für lexikalische Elemente.

Jedes lexikalische Element ist über einen sogenannten regulären Ausdruck [8] spezifiziert, sodass alle Deklinationsformen und Abkürzungsvarianten eines Nomens demselben Token zugeordnet werden. Beispielsweise werden der Normalform⁵ *Basalzellkarzinom* folgende Textvarianten zugeordnet:

- *Basalzellkarzinome*,
- *Basalcellcarcinoms*,
- *Basalzell-Ca*,
- *Basalzell-Ca*.

Zusätzlich führt der Scanner beim Einlesen von Texten eine Ähnlichkeitsanalyse⁶ durch, sodass auch Zeichenketten mit wenigen vertauschten, fehlenden oder zusätzlichen Zeichen der betreffenden Normalform zugewiesen werden. Beispiele hierfür sind:

⁴ Die technischen Werkzeuge eines Scanners und Parsers sind etablierte Methoden der Informatik [8].

⁵ In der Regel der Wortstamm.

⁶ Einen Überblick über Methoden der Ähnlichkeitsanalyse von Texten und Erkennung von Duplikaten sind im Artikel von Winkler zu finden [9]. Mit dem dort genannten Jaro-Distanzmaß werden hohe Trefferquoten bei gleichzeitiger Minimierung von Fehlentscheidungen erzielt.

Bundesgesundheitsbl 2019 · 62:1468–1475 <https://doi.org/10.1007/s00103-019-03045-2>
© Der/die Autor(en) 2019

O. Eckert

Elektronische Kodierung von Todesbescheinigungen

Zusammenfassung

Derzeit wird etwa die Hälfte der deutschen Todesbescheinigungen durch die statistischen Ämter auf Landesebene elektronisch kodiert, um das für einen Todesfall verantwortliche Grundleiden nach den Vorgaben der Weltgesundheitsorganisation (WHO) zu ermitteln. In diesem Beitrag wird der Kodiervorgang erklärt und seine Bedeutung für die Todesursachenstatistik erläutert. Der Kodierkern MUSE ist seit einigen Jahren als Softwarekomponente in das internationale Kodiersystem Iris integriert und wird von der internationalen Iris Core Group kontinuierlich weiterentwickelt. Ein neues Modul, das die syntaktischen Besonderheiten der deutschen Sprache berücksichtigt, weist soweit wie möglich den textlichen Todesursachen aus Teil 1 und Teil 2 der Todesbescheinigung automatisch ICD-10-Kodes zu. Hierbei wird erstmals auch eine automatische

Textkorrektur eingesetzt. Uncodierte Textteile werden von Signierfachkräften nachkodiert. Trotz der Bemühungen ist die Verbesserung der Datenqualität für die Todesursachenstatistik weiterhin die größte Herausforderung. Diese Aufgabe kann nur von den beteiligten Akteuren (Ärzten, Gesundheitsämtern und statistischen Ämtern) gemeinsam gemeistert werden. Aus dem Prozess der elektronischen Kodierung in den statistischen Ämtern ergeben sich wertvolle Hinweise, bei welchen Todesbescheinigungen Qualitätsdefizite bestehen. Diese sollten künftig für eine gezielte Rückmeldung an die gegenüber der Statistik berichtspflichtigen Gesundheitsämter genutzt werden.

Schlüsselwörter

Todesursachenstatistik · Iris/MUSE · Elektronische Kodierung · Multikausale Todesursachen · Elektronische Todesbescheinigung

Electronic coding of death certificates

Abstract

About half of all German death certificates are processed electronically by regional statistical offices to select the underlying cause of death in accordance with the instructions from the World Health Organization. This paper illustrates electronic coding and its importance for cause of death statistics. The electronic coding kernel MUSE was added a few years ago to the international coding system Iris, which is maintained by the Iris Core Group. A new module assigns, as far as possible, ICD-10 codes to medical terms documented in death certificates. It takes into account syntactical specifics of the German language. In addition, automatic text correction is implemented. Unrecognised text parts are highlighted and coded manually.

Despite these efforts, improvement of data quality is the greatest challenge of German cause-of-death statistics. All involved stakeholders (physicians, local health authorities, and regional statistical offices) can cope with this task by a common effort. The process of electronic coding provides valuable hints for improving the quality of death certificates. In future, the coding system could generate feedback to local health authorities indicating medical documentation problems.

Keywords

Cause of death statistics · Iris/MUSE · Electronic coding · Multiple causes of death · Electronic death certificate

- *Basalzellkrazinom*,
- *Baslzellkarzinom*,
- *Bassalzellkarzinom*.

Für 5687 Scheine wurden durch eine automatische lexikalische Analyse insgesamt 32.160 elementare Zeichenketten berechnet. Hiervon sind nur 21.771 verwertbare Token, die nicht aus Satzzeichen

oder Leerzeichen bestehen. Mehr als 57 % der lexikalischen Elemente sind Nomen oder Adjektive (**Tab. 2**).

Die Zeilentexte enthalten eine oder mehrere Krankheits- oder Zustandsbeschreibungen (hier kurz Diagnose genannt). Eine einzelne Diagnose kann ein einfaches Nomen (z. B. „Pneumonie“) oder eine Nominalphrase (z. B.

Tab. 1 Beispiele lexikalischer Elemente (Token), die durch elektronische Zerlegung der medizinischen Texte in den Todesbescheinigungen entstehen

Zeichenkette	Typ des Tokens	Normierter Text	Synonym	Kodierrelevanz ^a
A. carotis interna	Nomen	Arteria carotis interna	–	Hoch
Basalzell-Ca	Nomen	Basalzellkarzinom	–	Hoch
Ethyltoxisch	Adjektiv	Äthyltoxisch	Alkoholtoxisch	Hoch
Art. HT	Fachliche Abkürzungen	Arteriell Hypertonie	–	Hoch
COPD	Fachliche Abkürzungen	COPD	Chronisch obstruktive Lungenerkrankung	Hoch
Stadium Gold IV	Quantitative Angaben	Gold 4	–	Gering
Auf Grund	Präposition (PRAEP)	Aufgrund	Bei	Hoch

^aJedes Token mit hoher Kodierrelevanz hat ein textliches Äquivalent, das in einem oder mehreren Einträgen des alphabetischen Verzeichnisses enthalten ist

Tab. 2 Differenzierung von 21.771 signifikanten Token nach Wortkategorien

Lexikalische Kategorie/Wortkategorien	Anzahl	Anteil (%)
Adjektive und adjektivisch verwendete Partizipien	4192	19,3
Adverbien und adverbiale Bestimmungen	19	0,1
Artikel	162	0,7
Fachliche Abkürzungen und Akronyme	916	4,2
Konjunktionen und Trennzeichen (Kommata)	2771	12,7
Nomen	9788	45,0
Präpositionen	961	4,4
Quantitative, metrische, klassifikatorische und zeitliche Angaben	386	1,8
Unbekannte Zeichenfolgen (nicht automatisch erkannt)	2576	11,8

„bösartige Neubildung der Harnblase“) sein, wobei das zentrale Nomen eine Krankheit, Verletzung oder eine äußere Ursache ist.

Gleichberechtigte Diagnosen sind durch Konjunktionen („und“ oder „mit“) oder durch Kommata voneinander getrennt:

Aortenklappenstenose *und* Mitralklappeninsuffizienz.

Insgesamt wurden bei den analysierten Todesbescheinigungen 3069 Trennzeichen und Konjunktionen verwendet, die zur Trennung von Diagnosen genutzt werden.

Diagnosen können auch durch kausale Präpositionen („wegen“, „bei“, „aufgrund“, „nach“) voneinander getrennt sein:

- petrochantäre Femurfraktur *nach* Sturz,
- Infektion *nach* Hüft-TEP,

- Kolostomie *nach* Ileus *bei* Peritonealkarzinose.

Kodierparser

Der Kodierparser (kurz Parser [8]) ist eine Softwarekomponente, die für jeden Zeilentext des medizinischen Teils der Todesbescheinigung möglichst automatisch Vorschläge zur Kodierung nach ICD-10 berechnet. Der durch den Scanner zerlegte Text wird nach einem festen Regelwerk analysiert. Das dabei verwendete Wörterbuch wurde auf Basis des alphabetischen Verzeichnisses der ICD-10 WHO erstellt [7].

Der Algorithmus wird beispielhaft anhand des nachfolgenden Textes erläutert: *Mamma-Ca mit pulmonalen und zerebr. Metastasen und Hirnblutungen.*

Der Parser benutzt für die weiteren Schritte die Normalform des Textes:

Mammakarzinom mit pulmonal und zerebral Metastase und Hirnblutung.

In mehreren Kodierversuchen (=KV) sucht der Parser einen passenden Wörterbucheintrag für den gesamten Text oder für Textteile. Zusätzlich werden die Texte nach den vom DIMDI entwickelten Standardisierungsregeln umgeformt [2, 10].

KV NORMALFORM. Überführung aller Textteile in die Normalform.

Das Kodierergebnis (C509, C793) ist unvollständig, da „pulmonale Metastasen“ (C780) nicht kodiert wurden. Daher wurden weitere Kodierversuche unternommen.

KV SYNTAX. Regelbasierte Textumformung aufgrund der syntaktischen Struktur von Textteilen.

Der Teiltext „*pulmonal und zerebral Metastase*“ entspricht dem syntaktischen Muster „ADJ1 UND ADJ2 NN“. Für dieses Wortmuster ist eine Vorschrift hinterlegt, die nacheinander Kodierversuche mit umgeformten Texten durchführt:

- „*pulmonal zerebral Metastase*“⁷ → C793 (unvollständig kodiert),
- „*pulmonal Metastase, zerebral Metastase*“⁸ → C780, C793.

KV REST. Verbleibende unkodierte und zusammenhängende Textteile.

In diesem Beispiel wurden folgende Textteile erfolgreich kodiert:

- „Mammakarzinom“ → C509,
- „Hirnblutung“ → I619.

Der Parser baut in diesem Beispiel einen sogenannten vollständigen Kodierbaum auf, der unterhalb des Wurzelknotens⁹ aus Knoten von kodierten Textteilen oder Trennwörtern und -zeichen besteht (Abb. 2).

In den Parser sind zusätzlich die nachfolgend aufgelisteten Kodiervarianten integriert, die im obigen Beispiel nicht angewandt worden sind:

⁷ Schema für Textumformung „ADJ1 ADJ2 NN“.

⁸ Schema für Textumformung „ADJ1 NN, ADJ2 NN“.

⁹ Der Wurzelknoten repräsentiert den Text einer Zeile.

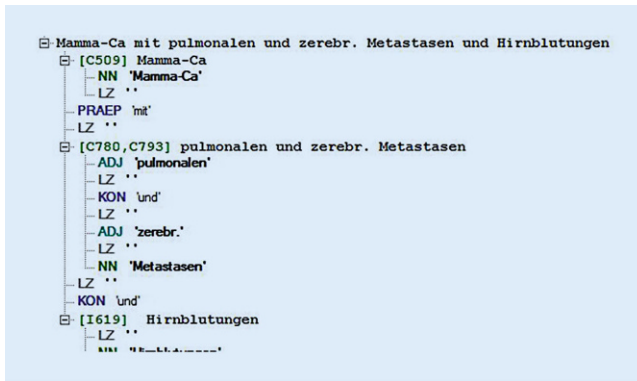


Abb. 2 ▲ Beispiel eines Kodierbaums, der durch die Softwarekomponente „Kodierparser“ erzeugt wird und den medizinischen Text (oben) in kodierbare Kindelemente (Texte hinter den ICD-10-Kodes in eckigen Klammern) und nichtkodierrelevante Kindelemente (Trennwörter) zerlegt. *NN* Nomen, *PRAEP* Präposition, *LZ* Leerzeichen, *ADJ* Adjektiv, *KON* Konjunktion

KV SYNONYM. Verwendung von Synonymen.

Beispielsweise ist „Krankheit“ für „Erkrankung“ als Synonym hinterlegt. Wenn die Kodierung mit dem ursprünglichen Begriff fehlschlägt, so ist ein Kodierversuch unter Verwendung des Synonyms ggf. erfolgreich (Beispiel: „Erkrankung des Gehirns“ wird in „Krankheit des Gehirns“ umgewandelt). Insbesondere bei lateinischen und deutschen Synonymen von Krankheiten oder Organen ist diese Methode sinnvoll.

KV HYPERONYM. Verwendung von Oberbegriffen zur Kodierung.

Wenn spezifische Lokalisationen wie Mediastromgebiet angegeben sind, die nicht Bestandteil des Wörterbuches sind, so kann ein übergeordnetes Organ (Gehirn) als Oberbegriff zugeordnet werden. Wenn der spezifische Term (Infarkt im Mediastromgebiet) nicht kodiert werden kann, so wird ein Kodierversuch mit dem Oberbegriff (*Infarkt im Gehirn* = I639) unternommen.

KV SIGNIFIKANT. Weglassen von irrelevanten Einzelwörtern.

In einem abgestuften Verfahren werden aus dem Standardtext Wörter von niedriger oder mittlerer Kodierrelevanz entfernt (z. B. wird in einem nachfolgenden Kodierversuch beim Term „COPD Gold IV“ die Gradeinteilung „GOLD IV“ weggelassen).

Ergebnisse der automatischen Kodierung

Bei einer automatischen Kodierung mit dem Wörterbuch und ohne Einsatz des Parsers wurden 10.051 Zeilen (61,44%) vollständig kodiert. Diese relativ hohe Rate hängt damit zusammen, dass die gebräuchlichsten Diagnosetexte im Wörterbuch hinterlegt sind.¹⁰

Bei Verwendung des Kodierparsers wurden dagegen für 15.200 Textzeilen (92,92%) die multikausalen Codes vollständig berechnet.

14.436 Textzeilen (88,25%) wurden bereits nach Überführung in die Normalform (siehe Kodiervariante **KV NORMALFORM**) vollständig kodiert. Ein wichtiger Grund hierfür ist, dass Abkürzungen ersetzt und Schreibfehler korrigiert werden.

Bei 764 Textzeilen (4,67%) mit komplexeren Diagnosetexten wurden die restlichen im vorherigen Abschnitt genannten Kodiermethoden eingesetzt.

866 Zeilen (5,29%) wurden unvollständig mit mindestens einem ICD-10-Kode kodiert. Lediglich 293 Zeilen (1,79%) wurde kein ICD-10-Kode zugewiesen. Neben Termen mit gravierenden Rechtschreibfehlern und unvollständigen Texten gab es aber auch fachlich

korrekte Bezeichnungen und Abkürzungen, die noch nicht in das Wörterbuch aufgenommen worden sind (z. B. „AEG-II-Ca“, „Aggrastat-Therapie“, „Akuter Verschluss d. A. mesenteria sup.“, „Hautläsionen, Kortison bedingt“).

Strukturierte Nachbearbeitung

Todesbescheinigungen, bei denen Teiltex-te nicht automatisch kodiert werden oder automatische Kodierungen überprüft werden sollen (bspw. bei der Verwendung von Oberbegriffen für die Kodierung), werden mithilfe des Systems in einer strukturierten Weise überprüft.

■ **Abb. 3** zeigt exemplarisch das Fenster für die Überprüfung von Kodierungen. In der linken Spalte sind die Textzeilen von Teil 1 und 2 dargestellt. Für Zeile 1a und 1b konnte jeweils eine automatische Kodierung ermittelt werden, was an dem gesetzten Häkchen erkennbar ist. Teil 2 der Todesbescheinigung wurde dagegen nicht vollständig kodiert. Für unkodierte Textteile (hier „Angiostenose“) wird ein Fragezeichen in das Feld mit den Codes eingefügt. In der rechten Spalte werden Aufgaben für die Kodierung eingeblendet. In diesem Beispiel kann die Signierfachkraft entweder die fehlende Kodierung akzeptieren oder über das integrierte Wörterbuch einen geeigneten ICD-10-Kode manuell aus-suchen.

Die Recherche nach geeigneten Kodierungen für nichterkannte Terme (vergl. Auflistung im vorherigen Abschnitt) erfordert gute Kenntnisse der medizinischen Terminologie und der ICD-10.

Unkodierte Terme werden von Iris/MUSE bei der Fallbearbeitung protokolliert und können bspw. nach Häufigkeit ihrer Verwendung ausgewertet werden. Diese Informationen werden genutzt, um die automatische Kodierung gezielt zu verbessern.

Bedeutung der rechner-gestützten Kodierung multikausaler Todesursachen

Die Ergebnisse der automatischen Kodierung zeigen, dass der Kodierparser den Automatisierungsgrad der Texter-

¹⁰ Die zehn häufigsten Zeilentexte (Herzinsuffizienz, Pneumonie, KHK, Hypertonie, Demenz, Herzversagen, respiratorische Insuffizienz, COPD, akutes Nierenversagen, arterielle Hypertonie) sind in 13,91% aller Zeilen zu finden.

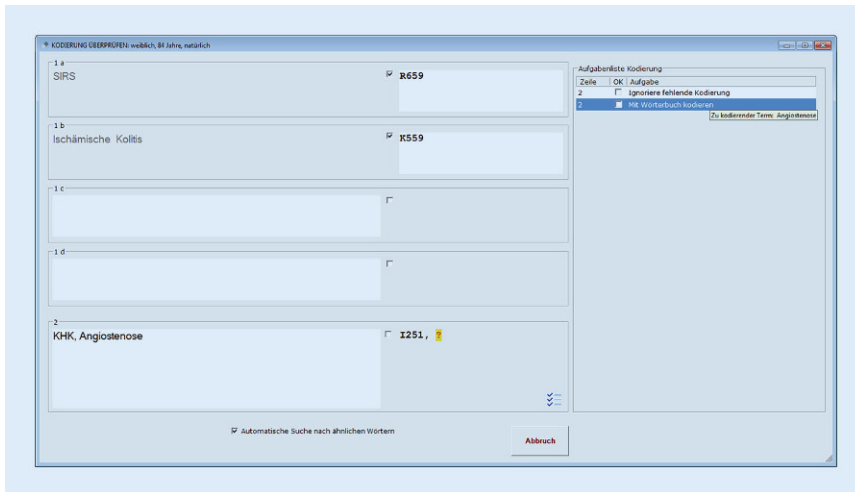


Abb. 3 ▲ Maske zur Nachbearbeitung eines nichtkodierten Textteils mithilfe einer Aufgabenliste. Das Fragezeichen markiert einen fehlenden ICD-Code

kennung um ca. 50 % erhöht, verglichen mit der automatischen Suche nach passenden Einträgen im Wörterbuch ohne den Parser. Da das verwendete Wörterbuch mit fast 400.000 Einträgen bereits gebräuchliche Diagnosetexte und ihre Varianten (z. B. Vertauschungen der Wortreihenfolge) enthält, werden die erweiterten Kodiervarianten des Kodierparsers (siehe oben) relativ selten genutzt.

Die vollständige Kodierung der auf einer Todesbescheinigung verzeichneten Krankheiten und Zustände ist ein unverzichtbarer Arbeitsschritt bei der Erstellung der amtlichen Todesursachenstatistik, da die WHO-konforme Auswahl des Grundleidens auf der vollständigen Kodierung aller auf dem Totenschein angegebenen Krankheiten und Zustände basiert.

Jedoch spiegeln multikausale Kodierungen ärztliche Dokumentationsfehler wider, wie das folgende Beispiel einer nichtplausiblen Kausalkette zeigt:

- 1a: Herzinsuffizienz (I509)
- 1b: Parkinsonsyndrom (G20)
- 1c: Diabetes mellitus (E149)
- 2: –

Diabetes mellitus ist hier als Ursache eines Parkinsonsyndroms dokumentiert. Die multikausale Kodierung (I509 / G20 / E149) übernimmt die falsche Kausalbeziehung. Bei Anwendung der WHO-Regeln wird aber nicht E149, sondern der Code G20 als Grundleiden ausgewählt, da die Kausalbeziehung

zwischen G20 und E149 als unzulässig klassifiziert ist. Der unikausale Algorithmus der WHO beinhaltet somit eine Strategie, Fehler zu erkennen und trotzdem ein möglichst aussagekräftiges Grundleiden zu selektieren. Aus statistischer Sicht wäre aber eine Korrektur der ausgestellten Todesbescheinigung mit einer nachvollziehbaren Kausalkette sinnvoll, um eine möglichst hohe Datenqualität zu erzielen.

Das multikausale Kodiersystem erfordert eine kontinuierliche Qualitätssicherung und Verbesserung. Die wichtigste Komponente ist die fortlaufende Pflege des amtlichen Wörterbuches und die Anpassung an die ICD-10-Updates der WHO durch das DIMDI.

Verbesserungsvorschläge und Fehlermeldungen der Signierfachkräfte, die sich aus der täglichen Arbeit mit dem Kodiersystem ergeben, werden regelmäßig vom Statistischen Bundesamt, von Experten der statistischen Ämter auf Landesebene und vom DIMDI geprüft und bei Änderungsbedarf umgesetzt. Zusätzlich werden regelmäßig unkodierte Texte analysiert und für häufig vorkommende Konstellationen Verbesserungen des Kodierparsers implementiert.

Die neue Texterkennung (Kodierparser) der deutschen Adaption von Iris/MUSE ist eine Weiterentwicklung gegenüber der Texterkennung (Standardisierung) der internationalen Iris-Version, welche zusätzlich die Syntax der medizinischen Fachsprache berücksichtigt und

ein Verfahren zur Fehlerkorrektur beinhaltet.

Herausforderung: Verbesserung der Datenqualität

Die Daten der deutschen Todesursachenstatistik sind nach Kriterien der WHO nur von mittlerer Qualität, da der Anteil ungenau bezeichneter Todesursachen zu hoch ist [4]. Verbesserungen sind nur durch koordinierte Anstrengungen der drei Akteure (Leichenschauärzte, Gesundheitsämter und statistische Ämter) möglich, die die gemeinsame Verantwortung für die Datenqualität tragen:

- Hauptsächlich hängt die Qualität der Todesursachenstatistik davon ab, wie sorgfältig Ärztinnen und Ärzte die Todesursachen feststellen und auf den Todesbescheinigungen dokumentieren. In einer vom Bundesministerium für Gesundheit beauftragten Studie werden Zuverlässigkeit, Vollständigkeit und Lesbarkeit der Angaben auf Todesbescheinigungen als entscheidende Einflussgrößen genannt [11]. Madea et al. [12] berichten von erheblichen Diskrepanzen (>40 %) zwischen den Angaben der ersten Leichenschau und späteren Obduktionsbefunden, die als Goldstandard für die Erstellung einer validen Todesbescheinigung angesehen werden. Schwedische Studien zeigen aber eine relativ hohe Übereinstimmung von 77 % zwischen den Angaben der Todesbescheinigung und den Todesursachen, die retrospektiv aus korrespondierenden Patientenakten abgeleitet werden [13]. Bei jüngeren Verstorbenen und einzelnen Krankheitsgruppen werden deutlich höhere Übereinstimmungen berichtet. Trotzdem enthalten Leichenschauscheine, die auf Basis klinischer Befunde in Patientenakten erstellt werden, wertvolle Informationen für die Todesursachenstatistik.
- Darüber hinaus ist sehr bedeutsam, ob die gegenüber den statistischen Ämtern berichtspflichtigen Gesundheitsämter die Lesbarkeit und Plausibilität sorgfältig prüfen und ggf. Korrekturen nach Rückfragen bei den Leichenschauärzten vornehmen.

Trotz einer in Deutschland niedrigen Obduktionsrate von unter 5 % [12] sollten die Gesundheitsämter immer die Obduktionsergebnisse durch Staatsanwaltschaft bzw. Kriminalpolizei und Krankenhäuser erhalten, damit korrigierte Todesbescheinigungen in die Statistik einfließen können. Aufgrund unterschiedlicher landesrechtlicher Regelungen sind die Prüfprozesse deutscher Gesundheitsämter aber heterogen.

- Nicht zuletzt wird die Qualität der Todesursachenstatistik durch die Kodierqualität der regionalen statistischen Ämter beeinflusst. Wegen der noch nicht vollständigen Verwendung des elektronischen Kodiersystems Iris/MUSE gibt es systematische Verzerrungen durch unterschiedliche Kodierpraktiken (manuell vs. elektronisch) in den statistischen Ämtern [4]. Die Nutzung des Kodierkerns MUSE hat mehrere Vorteile:
 - Das Grundleiden wird weitgehend automatisiert nach einem international abgestimmten Regelwerk ermittelt. Dadurch wird die internationale Vergleichbarkeit der deutschen Todesursachenstatistik gefördert.
 - Das in der Software hinterlegte Regelwerk wird jährlich von der Iris Core Group aktualisiert, sodass ICD-10-Updates der WHO zeitnah und bundeseinheitlich in Deutschland umgesetzt werden können.

Perspektiven für eine Weiterentwicklung

Basis für die Erstellung der Todesursachenstatistik sind diejenigen Krankheiten und Zustände, die auf Todesbescheinigungen in Textform aufgeschrieben sind. Die zentrale Kodierung durch die statistischen Ämter erfolgt in einem nachgelagerten Prozess, da die Klartextangaben in ausreichender Präzision und zeitnah von anderen Akteuren benötigt werden (bspw. bei kriminalpolizeilichen Ermittlungen nach Hinweisen auf nicht-natürlichen Tod oder bei einer zweiten Leichenschau durch das Gesundheitsamt vor Feuerbestattungen). Mit der

Kodierung sollen folgende Ziele erreicht werden:

- Umsetzung des weltweit gültigen Kodierstandards der ICD-10 WHO, da im deutschen Gesundheitswesen die nationale Klassifikation ICD-10 GM Standard ist.
- Entlastung der Ärzte von der Kodierung in der für sie ungewohnten Klassifikation ICD-10 WHO: Sie können Krankheiten, Verletzungen und äußere Ursachen in ihrer Fachsprache dokumentieren.¹¹
- Sicherstellung der Anwendung von Priorisierungsregeln, die auch nicht-medizinische Kriterien beinhalten (bspw. Relevanz für das Gesundheitswesen und die Gesundheitsberichterstattung).

In einer neueren Publikation weisen Zack et al. [14] nach, dass Todesbescheinigungen in einem nicht unerheblichen Maße fehlerhaft ausgefüllt sind. Hierdurch wird die Arbeit der statistischen Ämter deutlich erschwert.

Eine wichtige Verbesserungsmaßnahme für die Datenqualität soll die Pilotierung einer bundeseinheitlichen elektronischen Todesbescheinigung sein, die in einem mehrjährigen Projekt entwickelt werden soll, das vom Bundesministerium für Gesundheit gefördert wird.

Da wichtige Hinweise auf fehlende Lesbarkeit und Plausibilität während des statistischen Aufbereitungsprozesses mit Iris/MUSE anfallen, wäre eine gezielte einzelfallbezogene Rückmeldung an das berichtspflichtige Gesundheitsamt sinnvoll. Die inhaltlichen Kriterien für die Generierung eines Warnhinweises müssen noch entwickelt und abgestimmt werden. Beispielsweise könnten folgende Meldungen sinnvoll sein:

- „Todesursachen auf der Bescheinigung sind nach den Kriterien der WHO zu ungenau“ (Beispiel: „Herzstillstand“ als einzige Information).

¹¹ Vielfach ist eine hinreichend präzise Beschreibung von Krankheiten nur in Textform möglich, da sowohl die ICD-10 als auch die künftige ICD-11 Restkategorien von Krankheitsklassen enthält, bei denen unterschiedlichen Krankheiten derselbe Code zugewiesen wird.

- „Bitte überprüfen Sie die angegebene Kausalkette in Teil 1.“
- „Verwendete Abkürzung in Zeile 1a ist mehrdeutig“ (Beispiel: „HWI“ = „Hinterwandinfarkt“ oder „Harnwegsinfekt“).

Da bei Onlinemeldungen an den statistischen Verbund routinemäßig maschinenlesbare Prüfprotokolle [5] erstellt werden, welche die strengen Geheimhaltungsvorschriften nach § 16 des Bundesstatistikgesetzes beachten, könnte ein vergleichbares Verfahren auch in das elektronische Kodiersystem integriert werden. Damit können Signierfachkräfte gezielte Rückfragen anstoßen. Ein praktischer Nutzen kann sich dann ergeben, wenn die Prüfprotokolle in die Erfassungssoftware der Gesundheitsämter eingelesen werden können.

Korrespondenzadresse

Dr. rer. biol. hum. Olaf Eckert
Referat H101 Gesundheitsstatistiken,
Statistisches Bundesamt
Graurheindorfer Straße 198, 53117 Bonn,
Deutschland
olaf.eckert@destatis.de

Einhaltung ethischer Richtlinien

Interessenkonflikt. O. Eckert gibt an, dass kein Interessenkonflikt besteht.

Für diesen Beitrag wurden vom Autor keine Studien an Menschen oder Tieren durchgeführt. Für die aufgeführten Studien gelten die jeweils dort angegebenen ethischen Richtlinien.

Open Access Dieser Artikel wird unter der Creative Commons Namensnennung 4.0 International Lizenz (<http://creativecommons.org/licenses/by/4.0/deed.de>) veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Literatur

1. DIMDI (2018) ICD-10-WHO Version 2019. <https://www.dimdi.de/dynamic/downloads/klassifikationen/icd-10-who/version2019/icd10who2019regel-pdf.zip>. Zugegriffen: 11. Juni 2019

2. Iris Institute (2019) About Iris and Iris users. <https://www.dimdi.de/dynamic/en/classifications/iris-institute/>. Zugegriffen: 11. Juni 2019
3. Eckert O (2017) Verbesserte Qualität der nationalen und internationalen Todesursachenstatistik durch den Kodierkern MUSE. WISTA 4:118–130. https://www.destatis.de/DE/Methoden/WISTA-Wirtschaft-und-Statistik/_inhalt.html
4. Eckert O, Vogel U (2018) Todesursachenstatistik und ICD, quo vadis? Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 61:796. <https://doi.org/10.1007/s00103-018-2756-5>
5. Statistisches Bundesamt (2013) Flyer Online-Meldeverfahren eSTATISTIK.core. <https://erhebungsportal.estatistik.de/Erhebungsportal>. Zugegriffen: 28. Juni 2019
6. Häring F (2015) Von der Todesbescheinigung zur Todesursachenstatistik – Verfahrensablauf in Baden-Württemberg. Statistisches Monatsheft Baden-Württemberg 3:18–22. https://www.statistik-bw.de/Service/Veroeff/Monatshefte/PDF/Beitrag15_03_03.pdf
7. DIMDI (2018) ICD-10 WHO 2019 Alphabet PDF-Referenzfassung. <https://www.dimdi.de/dynamic/downloads/klassifikationen/icd-10-who/version2019/icd10who2019alpha-pdf.zip>. Zugegriffen: 11. Juni 2019
8. Levine JR, Mason T, Brown D (1999) lex & yacc. O'Reilly, USA
9. Statistical Research Division U.S. Census Bureau (2006) Overview of record linkage and current research directions by Winkler WE. <https://www.census.gov/srd/papers/pdf/rrs2006-02.pdf>. Zugegriffen: 28. Juni 2019
10. WHO (2006) Language standardization for mortality coding. A German approach by Weber S and Özer O. [http://apps.who.int/classifications/apps/icd/meetings/2006meeting/WHOFIC2006-R103 – Language Standardization for Mortality Coding.pdf](http://apps.who.int/classifications/apps/icd/meetings/2006meeting/WHOFIC2006-R103-LanguageStandardizationforMortalityCoding.pdf). Zugegriffen: 28. Juni 2019
11. Bundesministerium für Gesundheit (2013) Gutachten zur Aufwand-Nutzen-Abschätzung zum Aufbau und Betrieb eines nationalen Mortalitätsregisters. <https://www.bundesgesundheitsministerium.de/service/publikationen>. Zugegriffen: 28. Juni 2019
12. Madea B, Rothschild M (2010) Ärztliche Leichenschau – Feststellung der Todesursache und Qualifikation der Todesart. Dtsch Arztebl Int 107(33):575–588. <https://doi.org/10.3238/arztebl.2010.0575>
13. Brooke HL, Talbäck M, Hörnblad J, Johansson LA, Ludvigsson JF, Druid H, Feychting M, Ljung R (2017) The Swedish cause of death register. Eur J Epidemiol 32(9):765–773
14. Zack F, Kaden A, Riepenhausen S, Rentsch D, Kegler R, Büttner A (2017) Fehler bei der Ausstellung der Todesbescheinigung. Rechtsmedizin 27(6):516–527