

Kodierübungen im Medizinstudium

Wie gut kodieren Anfänger Diagnosen mit der ICD-10?

Coding training for medical students

How good is diagnoses coding with ICD-10 by novices?

• Jürgen Stausberg¹ • Nils Lehmann¹

Die Vermittlung von Kenntnissen und Fähigkeiten der Dokumentation und der Kodierung gehört zu den Aufgaben des Medizinstudiums. Im Rahmen des Querschnittsbereiches Epidemiologie, medizinische Biometrie und medizinische Informatik wurden daher Kodieraufgaben mit den Studierenden durchgeführt. Hieraus lassen sich Rückschlüsse auf die Kodierqualität bei Anfängern ziehen. In einer Hausarbeit kodierten 118 Studierende Diagnosen von 15 Fällen aus der Nephrologie. Dies wurde mit einer Referenzlösung zur Bestimmung der Validität und untereinander zur Bestimmung der Interrater-Reliabilität verglichen. Auf Ebene der endständigen Codes waren 59,3% der Angaben richtig, die Vollständigkeit betrug 58,0%. Diese Werte stiegen auf 91,5% bzw. 87,7% bei Betrachtung der Kapitel. Das neu eingeführte, einfache Maß zur Bestimmung der Reliabilität unter den Bedingungen der Diagnosenkodierung ergab Werte von 0,46 für endständige Codes und 0,87 auf Ebene der Kapitel für die Übereinstimmung der Studierenden untereinander bei vergleichbarem Wert zur Referenzlösung. Demgegenüber liegen Werte aus der Krankenversorgung bei 0,34 bzw. 0,63. Sowohl für die Validität als auch für die Interrater-Reliabilität sind die Ergebnisse der Anfänger somit denjenigen von Experten, also den in der Praxis kodierenden Ärzten, vergleichbar. Der fehlende Vorsprung der Ärzte kann einerseits durch den hohen Dokumentationsaufwand und eine negative Einstellung gegenüber der Kodierung erklärt werden. Andererseits wird die Kodierung unter DRG-Gesichtspunkten durch viele Sonderregeln erschwert, die möglicherweise nicht zu einheitlichen sondern im Gegenteil zu falschen und zufälligen Kodierergebnissen führen. Die Studenten zeigen sich im 1. Semester des 2. Abschnitts der ärztlichen Ausbildung für die Kodierung jedenfalls gut gerüstet.

Schlüsselwörter: Diagnosen, Klassifikation, Kodierung, Medizinstudenten, Reliabilität und Validität

Teaching of knowledge and competence in documentation and coding is an essential part of medical education. Therefore, coding training had been placed within the course of epidemiology, medical biometry, and medical informatics. From this, we can draw conclusions about the quality of coding by novices. One hundred and eighteen students coded diagnoses from 15 nephrological cases in homework. In

¹ Institut für Medizinische Informatik, Biometrie und Epidemiologie, Universitätsklinikum Essen, Deutschland

addition to interrater reliability, validity was calculated by comparison with a reference coding. On the level of terminal codes, 59.3% of the students' results were correct. The completeness was calculated as 58.0%. The results on the chapter level increased up to 91.5% and 87.7% respectively. For the calculation of reliability a new, simple measure was developed that leads to values of 0.46 on the level of terminal codes and 0.87 on the chapter level for interrater reliability. The figures of concordance with the reference coding are quite similar. In contrary, routine data show considerably lower results with 0.34 and 0.63 respectively. Interrater reliability and validity of coding by novices is as good as coding by experts. The missing advantage of experts could be explained by the workload of documentation and a negative attitude to coding on the one hand. On the other hand, coding in a DRG-system is handicapped by a large number of detailed coding rules, which do not end in uniform results but rather lead to wrong and random codes. Anyway, students left the course well prepared for coding.

Keywords: classification, coding, diagnoses, medical students, reliability and validity

Einleitung

Die Dokumentation nimmt einen großen Teil der täglichen Arbeitszeit von Ärztinnen und Ärzten ein. Der Aufwand im Krankenhaus wird für die Innere Medizin auf 3 Stunden und 15 Minuten täglich pro Arzt geschätzt, für die Chirurgie auf 2 Stunden und 42 Minuten [2]. Hiervon entfallen 12 Minuten auf die Kodierung von Diagnosen und Prozeduren für Abrechnung, Qualitätssicherung und Gesundheitsberichterstattung. Die nach der neuen Approbationsordnung vom 27. Juni 2002 [1] geforderte Befähigung zur selbstständigen ärztlichen Berufsausübung umfasst daher auch die Vermittlung von Kenntnissen und Fähigkeiten der Dokumentation sowie der Kodierung von Diagnosen und Prozeduren mit den gesetzlich hierzu vorgeschriebenen Klassifikationen. Bei der Konzeption des Querschnittsbereiches Epidemiologie, medizinische Biometrie und medizinische Informatik an der Medizinischen Fakultät der Universität Duisburg-Essen wurde daher das Thema Dokumentation sowohl in die begleitende Vorlesung als auch in die scheinpflichtigen Übungen aufgenommen.

Der Querschnittsbereich ist in Essen geteilt (s. Tabelle 1). Von den 14 Vorlesungsstunden im 1. Semester des 2. Abschnitts der ärztlichen Ausbildung entfallen 4 auf die Medizinische Informatik, eine Übungsstunde ist für die Kodierung reserviert. Der neue Querschnittsbereich wurde erstmalig im Wintersemester 2003/2004 angeboten. Wir wollen die Ergebnisse der Studierenden bei der Kodierung von Diagnosen mit der ICD-10-SGB-V 2.0 [4] vorstellen und diese mit Angaben zur Kodierqualität von Ärzten vergleichen.

Tabelle 1: Stundenzahl des Querschnittsbereiches Epidemiologie, medizinische Biometrie und medizinische Informatik im 2. Abschnitt der ärztlichen Ausbildung

	Vorlesung	Übungen
1. Semester	14 Stunden	8 Stunden
6. Semester	24 Stunden	12 Stunden

Material und Methoden

In der freiwilligen Begleitvorlesung im Wintersemester 2003/2004 wurden in einer Stunde allgemeine Aspekte der Dokumentation, in einer weiteren Stunde Klassifikationen behandelt. Eine Stunde Kodierübungen an Computern unter Nutzung des Kodiertools ID DIACOS[®] der Firma ID GmbH, Berlin, war für alle Studierenden verpflichtend. Geleitet wurden die Übungen teilweise durch den Erstautor, teilweise auch durch vorher geschulte Tutoren, also Studierende höheren Semesters. Nach einer kurzen theoretischen Einführung wurden vorgegebene Diagnostexte durch die Studierenden mit ID DIACOS[®] in die ICD-10-SGB-V 2.0 kodiert, mit der Musterlösung verglichen und Unterschiede diskutiert. Als Hausarbeit erhielten die Studierenden ein Arbeitsblatt mit einer Liste von Diagnosen aus einem Arztbrief. Insgesamt standen 15 Fälle mit einer unterschiedlichen Anzahl von Diagnostexten zur Verfügung. Es handelte sich ausschließlich um nephrologische Krankheitsbilder. Abbildung 1 zeigt das allen Studierenden zur Verfügung gestellte Beispiel.

Der jeweilige Fall sollte innerhalb von 5-14 Tagen in die ICD-10-SGB-V 2.0 kodiert werden. Als Hilfsmittel wurden die Studierenden auf die HTML-Fassung der ICD-10 beim Deutschen Institut für Medizinische Dokumentation und Information verwiesen, nicht jedoch

Sehr geehrter Herr Kollege,

wir berichten Ihnen über Ihren Patienten, Herrn [Name], geb. [Geburtsdatum], wohnhaft [Adresse], [Geburtsort], der sich vom [Datum] bis zum [Datum] in unserer stationären Behandlung befand.

Diagnose .: Hypertensive Entgleisung bei Verdacht auf Nierenarterienstenose.

Abbildung 1: Beispielfall aus der Hausarbeit des Wintersemesters 2003/2004

auf die Deutschen Kodierrichtlinien. Teilweise konnte ID DIACOS® genutzt werden. Die Ergebnisse waren in der in Tabelle 2 dargestellten Form abzugeben. In der schriftlichen Prüfung befasste sich eine von fünf Aufgaben mit der Kodierung. Hier waren fünf vorgegebene Diagnosetexte aus dem Bereich der Hernien unter Zuhilfenahme eines entsprechenden Ausschnittes der ICD-10-GM 2004 zu kodieren: Hernia diaphragmatica (Zwerchfellhernie) (richtige Lösung K44.9), Hernia inguinalis, doppelseitig (beidseitige Leistenhernie) (K40.20), Hernia femoralis, irreponibel, ohne Gangrän (nicht reponierbare Femoralhernie ohne Gangrän) (K41.3), Hernia lumbalis (Lendenhernie) (K45.8), Hernia abdominalis (abdominelle Hernie) (K46.9).

Tabelle 2: Darstellung der Ergebnisse zum Beispielfall

Nummer der Diagnose:	Kode der ICD-10-SGB-V 2.0:					
1	1	1	5	•	0	
2	1	7	0	•	1	
3				•		
...				•		
10				•		

Die von den Studierenden zurückgegebenen Bögen wurden anonym in eine Datenbank unter Microsoft® Access 2000 eingegeben. Die Auswertung erfolgte mit Access, Microsoft® Excel 2000 und SAS Version 8.2. Die verwendeten Kenngrößen sind bei den Ergebnissen dargestellt, um den jeweiligen Kontext zu verdeutlichen. Als Vergleichskollektiv für die Ergebnisse zur Interrater-Reliabilität steht eine Doppelkodierung von 244 chirurgischen Fällen von September 2001 zur Verfügung. Diese Daten waren im Rahmen eines Vergleiches zwischen konventioneller Papierakte und rechnergestützter Basisdokumentation erhoben worden [13]. Ziel des Vergleiches war es, Stärken und Schwächen beider Dokumentationsformen hinsichtlich verschiedener Qualitätskriterien der Diagnosen- und Prozedurenkodierung herauszuarbeiten. Das unten eingeführte Reliabilitätsmaß wurde für die Daten dieser Untersuchung neu berechnet.

Von 129 Studierenden wurden 118 Bögen ausgewertet (91,5%), je Fall zwischen 6 und 11 Bögen. Von 519

Kodes waren 13 weder als kodierbarer Endpunkt noch als Knoten zulässig. Diese wurden bis auf zwei korrigiert, indem z. B. eine überflüssige Stelle im Kode gestrichen wurde. Da ein Kode doppelt eingegeben war, verblieben 516 für die weitere Analyse. Hierunter befanden sich noch 27 nicht endständige Kodes. Die Referenzlösung wurde sowohl bei der Hausarbeit als auch in der Klausur durch den Erstautor vorgegeben.

Ergebnisse

• Referenzlösung der Hausarbeiten

Für 15 Fälle wurden 66 Kodes benötigt. Im Mittel wurden 4,4 Kodes, im Median 4 Kodes (Spannweite 2-9) pro Fall verwendet. In einem Fall wurde eine Diagnose mit einer Kombination aus einem Kreuz-Kode und einem Stern-Kode abgebildet. Bei 46 unterschiedlichen Kodes kamen 14 nur bei einem Fall vor. Am häufigsten waren die *Essentielle (primäre) Hypertonie* (I10) mit 5, die *Sonstige chronische Niereninsuffizienz* (N18.8) mit 4 sowie die *Fettleber [fettige Degeneration]*, *anderorts nicht klassifiziert* (K76.0) und die *Tubulointerstitielle Nephritis, nicht als akut oder chronisch bezeichnet* (N12) mit je 3 Fällen. Häufigste Gruppen waren mit 7 Fällen *Hypertonie* (I10-I15) und *Niereninsuffizienz* (N17-N19) sowie mit 6 Fällen *Tubulointerstitielle Nierenkrankheiten* (N10-N16).

• Validität

Die Validität der Kodierung kann über die Vollständigkeit und Korrektheit bestimmt werden [7]. Für die Vollständigkeit wird die Anzahl richtiger Kodes im Nenner als Summe des Produkts der Anzahl abgegebener Bögen mit der Anzahl richtiger Diagnosen je Fall berechnet. Die Korrektheit wird als Quotient aus der Anzahl übereinstimmender Kodes zur Gesamtzahl abgegebener Kodes gebildet. Die von Logan [11] vorgeschlagene Unterscheidung zwischen falsch kodierten Diagnosen und nicht kodierten Diagnosen bei der Berechnung der Vollständigkeit war auf Grund der fehlenden eindeutigen Identifikation jeder Diagnose nicht möglich.

Die Berechnung der Vollständigkeit erfolgt analog zur Sensitivität, der Korrektheit analog zum positiven Vorhersagewert. Auf Grund der geringen Fallzahl ($N=15$) im Verhältnis zu den Klassen der ICD-10-SGB-V 2.0 ($N=12\ 401$) ist die Zahl der zu Recht nicht verwendeten Codes sehr groß. Damit liegen Spezifität und positiven Vorhersagewert automatisch sehr hoch und sind für diese Betrachtung von geringem Interesse.

Tabelle 3 stellt Vollständigkeit und Korrektheit dar. Für die Aggregationsebenen der ICD-10 werden auf jeder Ebene nur eindeutige Codes gewertet. Die Studierenden haben somit auf der Ebene der Klasse 58,0% der Diagnosen richtig kodiert, 59,3% der von den Studierenden ermittelten Ergebnisse sind richtig. Die Korrektheit schwankt zwischen 20,0% und 89,2% je nach Fall. Bei 9 Fällen sind mehr als 50% der Codes richtig, bei 6 Fällen weniger als 50%. Die Studierenden verwendeten im Mittel ebenfalls 4,4 Diagnosen pro Fall.

In der Klausur zeigte sich folgendes Ergebnis: 73 Studierende hatten alle 5 Diagnosen richtig kodiert, 35 Studierende 4 Diagnosen, 13 Studierende 3 Diagnosen, 3 Studierende 2 Diagnosen und 1 Studierender 1 Diagnose. Keiner der Studierenden hatte keine Diagnose richtig.

• Interrater-Reliabilität

Bei unserer Untersuchung treten zwei Sonderbedingungen für die Bestimmung der Interrater-Reliabilität auf. Zum einen ist die Zahl möglicher Kategorien (12 401 Klassen der ICD-10-SGB-V 2.0) im Verhältnis zur Zahl verwendeter Kategorien (123 unterschiedliche Codes) sehr hoch. Zum anderen werden Mengen mit einer oder mehreren Diagnosen verglichen. Zudem kann auch die Anzahl von Diagnosecodes je Kodierer unterschiedlich sein.

Das κ -Maß dient dem Vergleich einer durch zwei Personen kodierten Hauptdiagnose,

$\kappa = (p_o - p_e) / (1 - p_e)$. Hier ist p_o die Anzahl der Übereinstimmungen bezogen auf die Anzahl der Fälle, und p_e die sich aus den Randsummen der Mehrfeldertafel ergebende erwartete Übereinstimmung. Da die Zahl möglicher Codes der ICD-10 sehr hoch ist, ist die Mehrfeldertafel groß und typischerweise spärlich so besetzt, dass $p_e \ll p_o (< 1)$ gilt. Damit ist $\kappa \approx p_o$.

Für den Vergleich mehrerer Diagnosecodes bei 2 Kodierern berechnen wir je Fall $P = n_o / (n_1 + n_o + n_2)$. Hierbei ist n_o die Anzahl übereinstimmender Codes im gegebenen Fall, n_1 die Anzahl der nicht mit Kodierer 2

übereinstimmenden Codes von Kodierer 1 und n_2 die Anzahl der nicht mit Kodierer 1 übereinstimmenden Codes von Kodierer 2. In Abbildung 2 ist $n_o=3$, $n_1=2$, $n_2=4$ und $P=3/9$. Wir erhalten als Maßzahl p_{om} für die Übereinstimmung den Mittelwert der P über die Fälle. Bei einer Diagnose je Fall ist P entweder 0/2 oder 1/1 und es folgt $p_{om} = p_o$. Die Maßzahl p_{om} ist demnach eine mögliche Verallgemeinerung von p_o , gegeben mehrere Diagnosen je Fall.

Zum Vergleich der studentischen Diagnosen (mehrere Bögen je Fall) mit der Referenzlösung berechnen wir je Fall den Mittelwert der P und mitteln anschließend über die Fälle, um p_{om} zu erhalten. Zur Ermittlung der Interrater-Reliabilität bilden wir je Fall alle Paare von Studenten, die diesen kodiert haben (z. B. 5 Bögen bei einem Fall: $5 \cdot 4 / 2$ Paare), und berechnen das mittlere P . Die Berechnung von p_{om} erfolgt wieder durch anschließende Mittelung über die Fälle.

Tabelle 4 zeigt die Reliabilität der Studierenden untereinander, die Übereinstimmung mit der Referenzlösung sowie die Ergebnisse aus den Vergleichsdaten der Krankenversorgung. Für die Aggregationsebenen der ICD-10 wurden auf jeder Ebene ebenfalls nur eindeutige Codes gewertet. Auf Ebene der Klasse liegt p_{om} je Fall zwischen 0,24 und 0,65. Die mittlere Anzahl von Diagnosen je Fall war in den Vergleichsdaten aus der Krankenversorgung 3,9 bzw. 3,7, bei den Studierenden sowie der Referenzlösung 4,4.

Diskussion

Zur Übereinstimmung von Ärzten bei der Kodierung von Diagnosen liegen nur wenige kontrollierte Studien vor (z.B. [12]). Hierbei zeigen sich mäßige Ergebnisse bei der Betrachtung einer einzelnen Diagnose auf dem Niveau des endständigen Codes (=Klasse). Die Ursachen hierfür sind vielfältig.

- Die ICD ist fehlerhaft, mehrdeutig, missverständlich und widersprüchlich [15].
- Bei einer Kodierung auf der Basis von Befunden kann es zu unterschiedlichen Einschätzungen hinsichtlich der vorliegenden Erkrankung kommen [6]. Dieser Umstand entfällt bei der Vorgabe eines Diagnosentextes.
- Die Regeln zur Kodierung sind mit derzeit 193 Deutschen Kodierrichtlinien für den einzelnen nicht zu überblicken [3].
- Ökonomische Gesichtspunkte beeinflussen die Validität der Dokumentation [8].

Tabelle 3: Vollständigkeit und Korrektheit der Hausarbeit für die Aggregationsebenen der ICD-10

	richtige Codes	Kodes insgesamt	Korrektheit	richtige Lösungen	Vollständigkeit
Klasse	306	516	59,3%	527	58,0%
Dreisteller	337	506	73,7%	527	63,9%
Gruppe	407	472	86,2%	520	76,3%
Kapitel	322	352	91,5%	367	87,7%

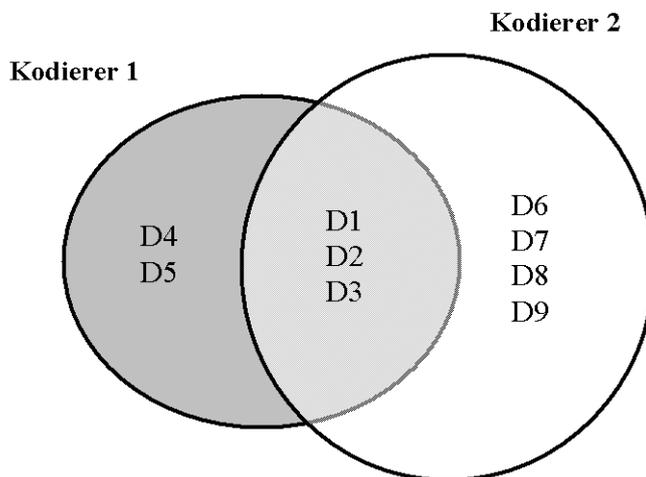


Abbildung 2: Berechnung des Maßes der Übereinstimmung bei mehreren Diagnosecodes als $P = n_o / (n_1 + n_o + n_2)$

Tabelle 4: Interrater-Reliabilität bei der Kodierung von mehreren Diagnosen, Verfahren zur Berechnung von p_{om} siehe Text

	Studierende		Studierende zu Referenzlösung		Vergleichsdaten aus der Krankenversorgung	
	p_{om}	95%-Konfidenzintervall	p_{om}	95%-Konfidenzintervall	p_{om}	95%-Konfidenzintervall
Klasse	0,46	0,39-0,53	0,47	0,35-0,59	0,34	0,30-0,38
Dreisteller	0,58	0,52-0,64	0,60	0,50-0,69	0,43	0,40-0,47
Gruppe	0,73	0,67-0,79	0,73	0,63-0,82	0,53	0,49-0,56
Kapitel	0,87	0,80-0,94	0,85	0,75-0,94	0,63	0,59-0,66

e) Der Aufwand zur Dokumentation ist hoch und in der täglichen Routine kaum zu leisten [2].

f) Das Thema Kodierung ist negativ besetzt. Der Kodierpflicht wird nur widerwillig gefolgt [5].

Von diesen Faktoren können wir in unserer Untersuchung von Studierenden des 1. Semesters im 2. Abschnitt der ärztlichen Ausbildung alle bis auf a) ausschließen. Hingegen war für die Studierenden die Kodierung durch den bislang fehlenden Kontakt mit dem Thema, den Umgang mit einer unbekanntem Software sowie das noch unvollständige medizinische Wissen erschwert.

Die Studierenden haben die Codes der Musterlösungen zu 58,0% richtig kodiert. Dieser Wert liegt höher, als er unter DRG-Gesichtspunkten für die Dokumenta-

tion der Krankenhäuser angenommen wird. Bei einer Untersuchung des Medizinischen Dienstes der Krankenversicherung Rheinland-Pfalz wurden im 1. Quartal der Prüfung nur 53,6% der Hauptdiagnosen aus einer Abteilung für Innere Medizin eines Krankenhauses der Grundversorgung bestätigt [14]. Nach einer Lernphase stieg dieser Wert im 3. Quartal auf 69,0%. Für Nebendiagnosen betrug die Vollständigkeit 31,1% im 1. Quartal und im 3. Quartal 44%. Eine Untersuchung von 1974, also noch vor der Entwicklung von DRGs, hatte einen Wert von 65,2% für die Übereinstimmung von Hauptdiagnosen ergeben [9].

In einer schwedischen Studie wurde der mittlere Kappa-Wert aus 3 Paaren von Kodierern gebildet, die jeweils 50 oder 51 unterschiedliche Befunde mit einer schwedischen Version der ICD-10 kodiert hatten [12].

Unter Zugrundlegung der Einteilung von Landis [10] (s. Tabelle 5) konnte bestenfalls eine moderate Übereinstimmung mit einem Kappa von 0,58 für den endständigen Kode erreicht werden. Die auch bei uns sichtbare Verbesserung auf höher aggregierten Ebenen der ICD-10 zeigte sich hier mit einem Kappa von 0,82 bei Betrachtung der Kapitel. Ähnliche Werte wurden für die Hauptdiagnose mit 0,49 bzw. 0,84 bei einem Vergleich von papiergestützter Krankenakte und rechnergestützter Basisdokumentation ermittelt [13].

Tabelle 5: Einteilung der Kappa-Werte nach Landis (1977) [10]

Kappa	Bewertung
<0,00	schlecht (poor)
0,00-0,20	gering (slight)
0,21-0,40	mäßig (fair)
0,41-0,60	moderat (moderate)
0,61-0,80	substantiell (substantial)
0,81-1,00	fast perfekt (almost perfect)

Zur Bewertung der Interrater-Reliabilität bei Kodierung von Studierenden wurde in der vorgelegten Untersuchung einerseits die Übereinstimmung mit der Referenzlösung herangezogen. Hierbei zeigen sich kaum Unterschiede. Im Vergleich zu den andererseits nachkalkulierten Daten aus [13] ist die Übereinstimmung jedoch deutlich besser mit einem p_{om} von 0,46 (95%-Konfidenzintervall 0,39-0,53) zu 0,34 (95%-Konfidenzintervall 0,30-0,38). Dieser Unterschied wird auf den höheren Aggregationsebenen der ICD noch deutlicher.

Schlussfolgerungen

Die Validität und Reliabilität der Kodierung von Diagnosen durch Studierende in die ICD-10-SGB-V 2.0 ist mit Ergebnissen von Ärzten vergleichbar. Dieser Vergleich wird allerdings durch das Design der publizierten Studien (zumeist fehlende Kontrolle der Kodierung in der täglichen Praxis) sowie das uneinheitliche Vorgehen zur Bestimmung von Validität und Reliabilität erschwert. Dennoch gibt eine gleich gute Qualität von „Anfängern“ (Studierenden) und Experten („Ärzten“) zu denken. Unserer Ansicht nach ist die fehlende Überlegenheit der Experten vor allem durch zwei der oben angeführten Ursachen begründet, die wir in unserer Untersuchung ausschließen konnten:

1. Die Deutschen Kodierrichtlinien tragen nicht zu einer richtigen Kodierung bei. Hingegen verwirren sie den Anwender, indem nicht klare Prinzipien, sondern vielfache Ausnahmen in den Vordergrund gestellt werden.

2. Der hohe Aufwand zur Dokumentation beeinflusst in Verbindung mit einer ablehnenden Einstellung in der Ärzteschaft die Kodierung negativ, wodurch auch die Qualität der gewonnenen Daten nicht das mögliche Maß erreicht.

Es ist daher eine radikale Vereinfachung der Kodierrichtlinien zu fordern. Damit wird gleichzeitig ein Beitrag zur Verringerung des Aufwandes geleistet. Zusätzlich sollte das Bewusstsein für die Bedeutung von Dokumentation und Kodierung sowie die hierzu notwendigen Kenntnisse, Fähigkeiten und Fertigkeiten schon im Studium vermittelt werden. Die hier vorgelegte Untersuchung zeigt, dass die Studierenden rasch einen zielführenden Zugang zur Kodierung finden.

Gleichzeitig wurde in diesem Beitrag ein einfaches Maß zur Berechnung der Reliabilität von Diagnosenmengen vorgestellt. Die Verwendung eines einheitlichen Maßes würde den kodierenden Bereichen eine sichere Beurteilung ihre Kodierqualität ermöglichen. Weitere Untersuchungen sind allerdings notwendig, um eine qualitative Einordnung von Zahlenwerten der Maßzahl p_{om} entsprechend der o. a. Einteilung von Kappa-Werten zu etablieren.

Danksagung

Die Autoren sind den Studierenden zu Dank verpflichtet, die sich sehr ernsthaft mit dem Thema Kodierung auseinandergesetzt haben. Für die Zurverfügungstellung der Referenzdaten bedanken wir uns bei Herrn Prof. Dr. med. Michael Betzler, Leitender Arzt der Klinik für Allgemeine Chirurgie, Unfall- und Gefäßchirurgie des Alfried Krupp Krankenhauses in Essen.

Korrespondenzadresse:

• Priv.-Doz. Dr. med. Jürgen Stausberg, Institut für Medizinische Informatik, Biometrie und Epidemiologie, Universitätsklinikum Essen, Hufelandstr. 55, 45122 Essen, Tel.: 0201/7234512, Fax: 0201/7235933
stausberg@uni-essen.de

Literatur:

[1] Approbationsordnung für Ärzte vom 27. Juni 2002. Bundesgesetzblatt. 2002;(Teil 1 Nr. 44 vom 3. Juli 2002):2405-35.

[2] Blum K, Müller U. Dokumentationsaufwand im Ärztlichen Dienst der Krankenhäuser. Bestandsaufnahme und Verbesserungsvorschläge. Düsseldorf: Deutsche Krankenhausverlagsgesellschaft; 2003.

[3] Deutsche Krankenhausgesellschaft, Spitzenverbände der Krankenkassen, Verband der Privaten Krankenversicherung, Institut für das Entgeltsystem im Krankenhaus. Deutsche Kodierrichtlinien, Version 2004.

Düsseldorf: Deutsche Krankenhausverlagsgesellschaft; 2004.

[4] Deutsches Institut für Medizinische Dokumentation und Information. Internationale statistische Klassifikation der Krankheiten und verwandter Gesundheitsprobleme 10. Revision. Systematisches Verzeichnis für die stationäre Versorgung - Version 2.0, Stand November 2000. Köln: Deutscher Ärzteverlag; 2001.

[5] Fietzek M. ICD: Verrat an der ärztlichen Kunst. Dtsch Ärztebl. 2004;101:A1084-A1086.

[6] Friedmann CP, Gatti GG, Murphy GC, Franz TM, Fine PL, Heckerling PS, Miller TM. Exploring the boundaries of plausibility: empirical study of a key problem in the design of computer-based clinical simulations. In: Kohane I, ed. Proceedings AMIA Annual Symposium. Philadelphia: Hanley & Belfus; 2002. p. 275-9.

[7] Hogan WR, Wagner MM. Accuracy of data in computer-based patient records. J Am Med Inform Assoc. 1997;4:342-55.

[8] Hsia DC, Krushat WM, Fagan AB, Tebbutt JA, Kusserow RP. Accuracy of diagnostic coding for Medicare patients under the prospective-payment system. N Engl J Med. 1988;318:352-5.

[9] Institute of Medicine. Reliability of Hospital Discharge Abstracts. Report of a Study. Washington: National Academy of Sciences; 1997.

[10] Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33:159-74.

[11] Logan JR, Gormann PN, Middleton B. Measuring the quality of medical records: A method for comparing completeness and correctness of clinical encounter data. In: Bakken S, ed. A medical odyssey: Visions of the future and lessons from the past. Proceedings AMIA 2001 Annual Symposium. Philadelphia: Hanley & Belfus; 2001. p. 408-12.

[12] Nilsson G, Petersson H, Ahlfeld H, Strender LE. Evaluation of three Swedish ICD-10 primary care visions: Reliability and ease of use in diagnostic coding. Method Inform Med. 2000;39:325-31.

[13] Stausberg J, Koch D, Ingenerf J, Betzler M. Comparing paper-based to electronic patient records: Lessons learned during a study on diagnosis and procedure codes. J Am Med Inform Assoc. 2003;10:470-7.

[14] Stockdreher K, Kuls G, Modrack M, Schlösser A, Walter S, Weibler-Villabos U, Nordhoff M. Parallelkodierung von Krankenhausfällen. Krankenhaus. 2004;96:437-43.

[15] Surján G. Questions on validity of International Classification of diseases-coded diagnoses. Int J Med Inform. 1999;54:77-95.