

# Automatisierte Berechnung der UICC-Stadien anhand von Krebsregisterdaten unter Berücksichtigung von verschiedenen TNM-Versionen

## Automated calculation of UICC stages based on cancer registry data with respect to different TNM classifications

### Abstract

Determination of UICC stages represents a key component of the diagnosis of solid tumors. This information is essential for therapy decision and prognosis of cancer patients. UICC staging is based on TNM status. The T status describes the size and extension of a tumor, N status indicates whether cancer has spread to lymph nodes, and the M status indicates whether distant metastases are present. The guidelines for calculating UICC stages vary among tumor types. These guidelines are subject to regular revision. Each version has a specific period of validity. Comprehensive cancer registration in Germany has a long history and thus encompasses various versions of UICC staging. The herein presented work used the Rhineland-Palatinate Cancer Registry as an example to describe automated calculation of UICC stages for all solid tumors, taking various staging guidelines of different versions into account. This methodology is applicable for other cancer registries as well as datasets with similar structures.

**Keywords:** cancer registration, UICC stages, TNM version, automation

### Zusammenfassung

Die Bestimmung von UICC-Stadien ist wesentlicher Bestandteil der Diagnosestellung von soliden Tumoren. Diese Angabe ist essentiell für die Therapieplanung und die Prognose von Krebspatientinnen und -patienten. Die Berechnung des UICC beruht auf Angaben zum TNM-Status. Der T-Status entspricht der Ausdehnung des Tumors, der N-Status gibt an, ob Lymphknoten befallen sind, und der M-Status gibt an, ob Metastasen vorliegen. Die Vorgaben zur Berechnung der UICC-Stadien sind je nach Tumorentität unterschiedlich, zudem gibt es regelmäßige Überarbeitungen der Berechnungsvorgaben. Diese Versionen haben jeweils einen bestimmten Gültigkeitszeitraum. Die flächendeckende Krebsregistrierung in Deutschland umfasst bereits viele Jahre und somit auch verschiedene Versionen der UICC-Berechnung. In der hier vorgestellten Arbeit wird am Beispiel des Krebsregisters Rheinland-Pfalz die automatisierte Berechnung der UICC-Stadien für sämtliche solide Tumoren beschrieben, die die verschiedenen Berechnungsvorgaben der unterschiedlichen Versionen berücksichtigt. Diese Methodik kann – neben anderen Krebsregistern – auch auf anderweitige Daten mit ähnlicher Struktur Anwendung finden.

**Schlüsselwörter:** Krebsregistrierung, UICC-Stadien, TNM-Version, Automatisierung

Benjamin Rieger<sup>1</sup>  
Cäcilie Mussel<sup>1</sup>  
Fabian Reinwald<sup>1</sup>  
Sinaida Branscheidt<sup>1</sup>  
Christina Justenhoven<sup>1</sup>  
Ruben Hamann<sup>1</sup>

1. Krebsregister Rheinland-Pfalz  
im Institut für digitale  
Gesundheitsdaten RLP,  
Mainz, Deutschland

## Grundlagen der Krebsregistrierung

Eine von fünf Personen weltweit erkrankt in ihrem Leben an Krebs, daher ist es wichtig, die Entwicklung dieser Erkrankung genau zu beobachten [1]. In Deutschland wurde aus diesem Grund eine gesetzliche Meldepflicht für onkologische Erkrankungen eingeführt. Jede medizinische Einrichtung, unterliegt dieser Meldepflicht [2]. Gemeldet werden Diagnosen, Therapien und die Ergebnisse von Nachsorgeuntersuchungen von Krebserkrankungen. Die Meldungen werden von der Einrichtung, die die jeweilige Leistung erbracht hat, an das im jeweiligen Bundesland befindliche Krebsregister übermittelt [3]. Um eine Vergleichbarkeit und eine deutschlandweite Zusammenführung der Daten zu ermöglichen, erfolgt die Krebsregistrierung bundesweit einheitlich entsprechend dem onkologischen Basisdatensatz [4]. Die auf dieser Basis entstandene flächendeckende Datensammlung bildet eine einzigartige Grundlage für epidemiologische und klinische Auswertungen, die, neben der Beobachtung des Krebsgeschehens und der Beantwortung von wissenschaftlichen Fragestellungen, insbesondere der Qualitätssicherung der onkologischen Versorgung dienen [2].

Um solche Auswertungen zu ermöglichen, müssen Krebsregister zunächst eine wichtige Aufgabe in der Qualitätsprüfung leisten. So erreichen das Krebsregister zu einem Meldeanlass, wie beispielsweise der Diagnose eines Tumors, oft zu unterschiedlichen Zeitpunkten von verschiedenen Einrichtungen mehrere Meldungen [5]. Diese Meldungen können sich inhaltlich voneinander unterscheiden, da auch die Diagnosestellung einem komplexen Prozess unterworfen sein kann. So kann beispielsweise ein Mammakarzinom im Rahmen der Früherkennung eine Erstdiagnose erhalten, die finale Beurteilung des Tumors erfolgt jedoch erst nach der operativen Entfernung. Die Informationen dieser Meldungen müssen dann sorgfältig geprüft und im Krebsregister zu einem sogenannten Best-of zusammengeführt werden [2]. Im Krebsregister Rheinland-Pfalz im Institut für digitale Gesundheitsdaten erfolgte dieser Prozess der Prüfung der Meldungen und der Zusammenführung der besten Informationen in den ersten Jahren manuell. Inzwischen erfolgt die Bildung des Best-of zu großen Teilen automatisiert. Bei der Implementierung von Automatisierungsschritten ist die Einhaltung von Qualitätsstandards unabdingbar, denn ausschließlich über eine qualitätsgeprüfte Datengrundlage sind verlässliche Auswertungen, z.B. im Rahmen der Versorgungsforschung, möglich. Grundlage aller Auswertungen ist eine korrekte Zuordnung eines Tumors. Neben dem ICD-10-GM-Code (International Classification of Diseases, 10<sup>th</sup> Revision, German Modification), über den die genaue Tumorentität verschlüsselt ist [6], gehören im Wesentlichen Informationen zur Histologie, der Differenzierung und der TNM-Status zur Diagnose eines Tumors [7], [8]. Das TNM-System umfasst den T-Status, d.h. die Ausdehnung des Tumors, den N-Status, d.h. ein möglicher Befall von Lymphknoten, und den M-Status, d.h. das mögliche Vorliegen von Metastasen [9]. Aus diesen TNM-Angaben kann das für Therapieentscheidungen und Prognose

wichtige UICC (Union für Internationale Krebsbekämpfung)-Stadium berechnet werden [9]. Bei einigen wenigen Entitäten werden neben dem TNM-Status auch andere Angaben zur Bestimmung des UICC-Stadiums hinzugezogen, so wird beispielsweise bei Appendix das Grading einbezogen [9]. Das UICC-Stadium ist ein wesentliches Kriterium der Diagnosestellung von soliden Tumoren. Es ist nicht nur ein wichtiger Faktor im Rahmen der Prognose, sondern auch essentiell für Therapieentscheidungen [9], [10].

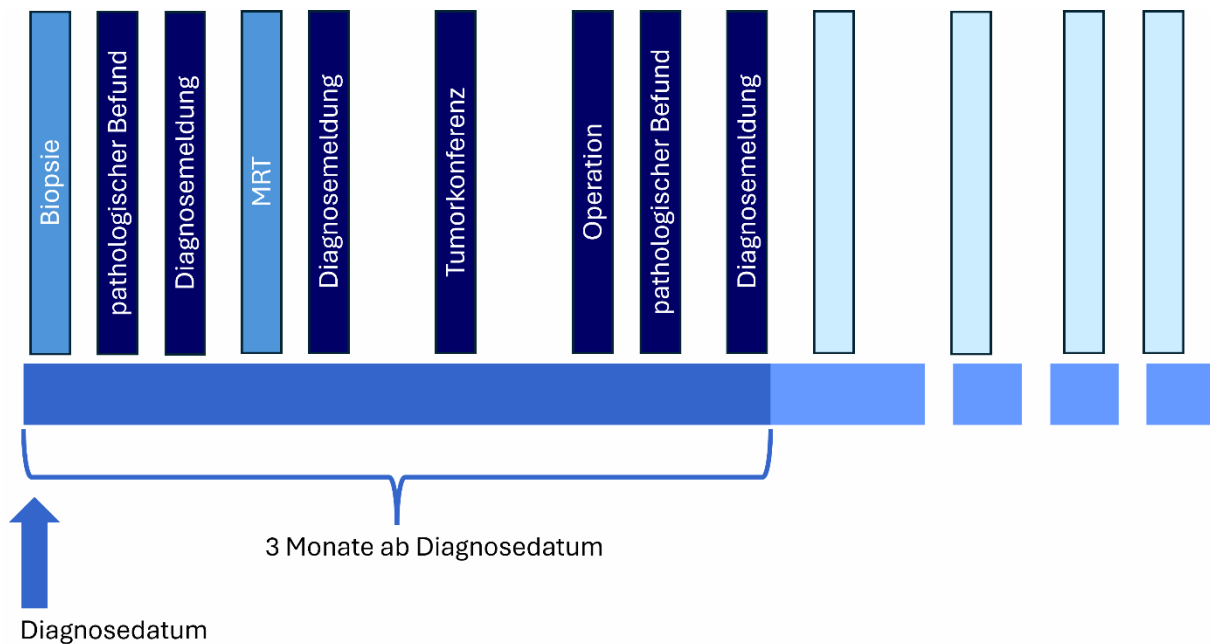
Die vorliegende Arbeit zeigt eine technische Möglichkeit zur korrekten Bestimmung der UICC-Stadien, die zum einen manuelle Schritte obsolet macht und zum anderen die sich über die Zeit ergebenden Veränderungen, durch Aktualisierungen der Vorgaben zur UICC-Berechnung, berücksichtigt.

## Schlüsselrolle des UICC

Ursprünglich wurde das TNM-System in den Jahren 1943–1952 entwickelt [11]. Diese Klassifikation von Tumoren wird seit dem regelmäßig aktualisiert und spielt eine wesentliche Rolle bei der Erfassung der Art des Tumors und der Abschätzung der Prognose [9]. Diese Informationen sind wesentlich für die Therapieplanung [10]. Die für Deutschland aktuell gültigen Empfehlungen finden sich vor allem in S3-Leitlinien des Leitlinienprogramms Onkologie [12]. Darüber hinaus ist zu beachten, dass Krebs eine sehr komplexe Erkrankung ist, die in vielen Fällen von Fachärztinnen und -ärzten verschiedener Disziplinen behandelt wird. Diese finden sich in der Regel zu sogenannten Tumorkonferenzen zusammen, um die optimale Therapieplanung einzelner Patientinnen und Patienten gemeinsam zu entscheiden [13]. Insbesondere im Hinblick auf die multidisziplinäre Betreuung von Krebserkrankungen ist es essentiell, eine übersichtliche und einheitliche Klassifikation von Tumoren zu nutzen, die einen reibungslosen Austausch zwischen den Behandelnden ermöglicht [9]. Neben der Behandlung von Krebspatientinnen und -patienten ist eine einheitliche und nachvollziehbare Dokumentation von Tumoren wesentliche Grundlage der Krebsforschung. Dieses Forschungsgebiet widmet sich vor allem der Aufdeckung des Optimierungspotenzials hinsichtlich der Vorbeugung, der Diagnose und Behandlung von Krebserkrankungen [14]. In diesem Bereich leisten die Krebsregister einen wesentlichen Beitrag [2].

## Herausforderungen der korrekten Bestimmung des UICC-Stadiums

Das UICC-Stadium wird im Wesentlichen über die Angaben zum TNM-Status berechnet [9]. Anmerkung: Für gynäkologische Tumoren wird kein UICC-Stadium berechnet, sondern ein FIGO (Fédération Internationale de Gynécologie et d'Obstétrique)-Stadium [9]. Im Folgenden wird aus Gründen der Lesbarkeit nur das UICC-Stadium er-



**Abbildung 1:** Im Rahmen der Krebsregistrierung werden innerhalb der ersten drei Monate nach Diagnosedatum sämtliche Informationen zu einem Tumor-Best-of zusammengefasst (dunkelblau). Diese Informationen stammen häufig aus unterschiedlichen Meldungen zu verschiedenen Meldeanlässen unterschiedlicher Einrichtungen.

Anmerkung: Biopsien oder MRT-Untersuchungen stellen keinen eigenen Meldeanlass dar. Die Ergebnisse dieser Untersuchungen werden im Rahmen einer Diagnosemeldung oder im späteren Verlauf als Verlaufsmeldung an das Krebsregister übermittelt.

wähnt, FIGO-Stadien sind aber in sämtliche Aussagen eingeschlossen. Hierbei stehen die Krebsregister vor verschiedenen Herausforderungen. In der Krebsregistrierung werden Informationen zur Diagnose von Tumoren aus allen Meldungen, die in den ersten drei Monaten ab dem Diagnosedatum eingehen, zusammengefasst [15]. Dies bedeutet, dass Informationen zum Tumor, darunter auch Angaben zum TNM-Status, nicht nur zu unterschiedlichen Zeitpunkten beim Krebsregister eingehen, sondern auch von verschiedenen Einrichtungen im Rahmen unterschiedlicher Meldeanlässe (Abbildung 1).

So kann eine Diagnose zunächst über bildgebende Verfahren und eine Biopsie gestellt und das Ergebnis an das Krebsregister übermittelt werden. Im Rahmen einer Tumorkonferenz wird die Diagnose besprochen und weitere Informationen zum Tumor werden an das Krebsregister gemeldet. Eine nachfolgende Operation und der zugehörige pathologische Befund liefern weitere Details. Häufig enthalten Meldungen über diesen Zeitraum unterschiedliche Angaben, da beispielsweise ein bildgebendes Verfahren einen Tumor kleiner erscheinen lassen kann als später bei der Operation festgestellt. Zudem ist es möglich, dass zunächst von Metastasenfreiheit ausgegangen wird, nach einigen Wochen aber doch eine Metastase nachgewiesen wird. Um dieser Fülle an Informationen gerecht zu werden und einer bundesweiten Einheitlichkeit so nah wie möglich zu kommen, wurden von den Krebsregistern Vorgehensweisen erarbeitet, die sowohl in einem Buch veröffentlicht wurden als auch in Arbeitsgruppen stetig weiterentwickelt werden [15], [16]. So wird in der Regel immer die höchste Angabe zum TNM-Status übernommen. Wurde beispielsweise zunächst davon ausgegangen, dass keine Metastasen vorliegen (M0) und etwas

später, aber innerhalb der ersten drei Monate nach Diagnose, doch eine Metastase nachgewiesen, wird im Best-of der Status M1 übernommen. Zusätzlich gilt, dass Angaben aus pathologischen Untersuchungen gegenüber Angaben aus einer klinischen Untersuchung vorgezogen werden. Liegen im Anschluss basierend auf diesen Regeln, die besten Informationen zum TNM-Status vor, erfolgt die Berechnung des UICC-Stadiums. Hierbei ist zu beachten, dass die Vorgaben zur Berechnung über die Zeit regelmäßig aktualisiert werden. Die erste Version der TNM-Klassifikation wurde in den Jahren 1943–1952 entwickelt und regelmäßig überarbeitet. Im Jahr 2010 erschien die 7. Version der TNM-Klassifikation [17]. Ab 2017 bis Ende 2025 richtete sich die Berechnung der UICC-Stadien nach Version 8 [9]. Die 9. Version wurde 2025 in der englischen Fassung veröffentlicht [18]. In den Krebsregistern liegen Tumoren mit TNM-Angaben aus mehreren Jahrzehnten vor. Entsprechend der immer wieder erfolgten Aktualisierungen müssen UICC-Stadien für Tumoren aus unterschiedlichen Jahren verschieden berechnet werden. In Tabelle 1 ist dies am Beispiel des Bronchialkarzinoms gezeigt.

In der Regel geben Meldende an, welche aktuell geltende Version der TNM-Klassifikation bei Diagnosestellung verwendet wurde. Fehlen diese Angaben zur TNM-Version, so erfolgt eine Zuordnung über das Leistungsdatum, der in der Meldung berichteten onkologischen Leistung. Liegen beispielsweise sämtliche Meldungen zu Diagnosen, Operationen und pathologischen Befunden im Zeitraum 2010 bis 2016, wird das UICC-Stadium nach Version 7 berechnet, die in diesem Zeitraum gültig war. Gehen TNM-Meldungen zu einem Tumor mit verschiedenen Versionen ein, so wird die aktuellere Version bevorzugt. Ausschlag-

**Tabelle 1: Unterschiede der Berechnung des UICC-Stadiums UICC II in den Versionen 6–8 am Beispiel des Bronchialkarzinoms (ICD-10-GM 34)**

Version der TNM-Klassifikation	T-Status	N-Status	M-Status	UICC-Stadium
Version 6 veröffentlicht 2002	T1	N1	M0	IIA
	T2	N1	M0	IIB
	T3	N0	M0	IIB
Version 7 veröffentlicht 2010	T2b	N0	M0	IIA
	T1a, T1b, T2a	N1	M0	IIA
	T2b	N1	M0	IIB
Version 8 veröffentlicht 2017	T2b	N0	M0	IIA
	T1a-c, T2a, T2b	N1	M0	IIB
	T3	N0	M0	IIB

**Tabelle 2: Vereinfachtes Beispiel zur Bildung des Best-of anhand eines Darmtumors. Die für das Best-of ausgewählten Felder sind grau unterlegt.**

Meldung	ICD-10-GM	Histo-morphologie	T-Status	N-Status	M-Status
Diagnose 1	C18.2	8010/3	x	x	x
Diagnose 2	C18.2	8140/3	1	0	0
Operation	C18.2	8140/3	2	1	0
Pathologie	C18.2	8480/3	2	1	x
<b>Best-of</b>	<b>C18.2</b>	<b>8480/3</b>	<b>2</b>	<b>1</b>	<b>0</b>

**Tabelle 3: Weg vom Datenbankeintrag zum UICC-Stadium an vier Beispielen zum Bronchial- (C34) und Darmkrebs (C18)**

Best-of-Datengrundlage						Keysting	Abgleich über Hash-Tabelle (Abbildung 2)	UICC-Stadium
ICD-10-GM	Histo-morphologie	TNM-Version	T	N	M			
C34.1	8041/3	8	4	3	1c	C34.1_8_8041/3_4_3_1c	→	IVB
C34.2	8072/3	8	2	3	1a	C34.2_8_8072/3_2_3_1a	→	IVA
C18.0	8140/3	8	3	1	0	C18.0_8_8140/3_3_1_0	→	IIIB
C18.6	8140/3	8	3	0	0	C18.6_8_8140/3_3_0_0	→	II

gebend für die Zuordnung ist dabei jeweils die Angabe zum T-Status.

Die im Krebsregister Rheinland-Pfalz im Institut für digitale Gesundheitsdaten etablierte automatisierte Berechnung der UICC-Stadien, die auf diese Weise verschiedene TNM-Versionen berücksichtigt, stellt eine Besonderheit dar, die bisher wenig Verbreitung gefunden hat. Das genaue Vorgehen wird im Folgenden vorgestellt.

## Automatisierte Berechnung der UICC-Stadien

Um das UICC-Stadium korrekt berechnen zu können, müssen, neben dem Best-of zum TNM-Status, auch das Best-of des zur Tumorentität gehörigen ICD-10-GM-Codes und der Histomorphologie berücksichtigt werden [9]. Die Bildung des Best-of erfolgt nach unter den Krebsregistern abgestimmten Regeln in dem genauere bzw. höhere Angaben geringeren bzw. ungenaueren Angaben gegenüber bevorzugt werden [15], [16]. In Tabelle 2 findet sich ein vereinfachtes Beispiel.

Nach der Best-of-Bildung liegt zu jeder Tumordiagnose eine Zeile vor (Best-of-Datengrundlage). In dieser Datengrundlage befinden sich die Informationen zu den einzelnen Variablen in getrennten Feldern (Tabelle 3). Die Übersetzung dieser einzelnen Ausprägungen in einen UICC-Code erfolgt über sogenannte Hashtabellen, für deren Hashfunktion die zusammengesetzten Strings der einzelnen Ausprägungen der Variablen als Schlüssel dienen. So kann jede denkbare Stringkombination einem exakten UICC-Wert zugeordnet werden. In den nachfolgenden Schritten werden zuerst der ICD-10-GM, die TNM-Version sowie der histomorphologische Code zu einem Keysting zusammengesetzt. Dieser String wird übersetzt in einen Gruppen-TAG-String, welcher wiederum zusammen mit den Ausprägungen der T-, N- und M-Attribute zu dem jeweiligen UICC-Wert führt. Die Skripte zur Best-of-Bildung und zur Bestimmung des UICC-Stadiums wurden mit R [19] in RStudio erstellt [20].

Für den Abgleich über Hashtabellen wurden die Informationen aus den TMN-Büchern in digitale Form übertragen (Abbildung 2). Hierbei ist zu beachten, dass es im Buch Angaben wie „all“ gibt [9]. Dies bedeutet zum Beispiel,

1	ALL	keine Einschränkung	8000-9992
2	KA	Karzinome (HG: 1,3,4,5)	8051-8084, 8120-8131, 8140-8149, 8160-8162, 8190-8221
3	ME	Melanome	8720-8790
4	PL	Plattenepithel- und Ubergangsepithelkarzinome	8051-8084, 8120-8131
5	AD	Adenokarzinome	8140-8149, 8160-8162, 8190-8221, 8260-8337, 8350-8551, 8570-8576
6	MT	Mesotheliome	9050-9055
7	TY	Epithelale Neoplasien des Thymus	8560-8589
8	SW	Sarkome und Weichteiltumore	8680-8713, 8800-8921, 8990-8991, 9040-9044, 9120-9138
9	AN	Andere naeher bezeichnete Krebsarten (HG: 16)	8720-8730, 8930-8936, 8950-9983, 9000-9030, 904
10	UZ	Ubergangszellpapillome und -karzinome	8120-8131
11	KAM	Karzinome ohne Merkezzell (HG: 1,3,4,5)	8051-8084, 8120-8131, 8140-8149, 8160-8162
12	KAK	Karzinome ohne Kiatskin und D. Cysticus (HG: 1,3,4,5)	8051-8084, 8120-8131, 8140-8149
13	O1	schilddruese_v8_papilaer_foelkulaer	8050/2, 8050/3, 8052/2, 8052/3, 8130/1, 8130/3
14	O2	schilddruese_v8_medullaer	8345/3, 8346/3, 8347/3, 8510/3, 8512/3, 8513/3, 8581/1
15	O3	schilddruese_v8_anaplastisch	8021/3
16	O4	leber_v8_hepatozellulaer	8170/0, 8170/3, 8172/3, 8173/3, 8174/3, 8175/3
17	O5	intrahept_gallengang_v8	8160/0, 8160/3, 8180/3

Umsetzer der Histomorphologien (Ausschnitt)

```

627 my %BUFF;
628 my %BUFF_EDEP;
629 my %BUFF_OMIT_TNM;
630 my %BUFF_DIFF;
631 my %BUFF_DIFF_MIN_SCORE;
632
633 my %TMN2UICC;
634 my %EV2GROUP;
635
636 for my $p (sort {$a <=> $b} keys (%HEAD_POS))
637 {
638     my $header = $UICC{$p};
639     #next unless($header =~ /_v8_/); ## TEST TEST WICHTIG !!!!
640     my $stable_group = $HEAD_POS2GROUP{$p};
641     print STDERR "verarbeitete Tabellenposition '$p' \n";
642     # Keine erweiterten Abhängigkeiten für diese Tabelle
643     unless(exists($HEAD_POS2EXTEND{$p}))
644     {
645         # Iteriere über jeweilige Tabelle (ohne Header -> i = p + 1)
646         for (my $i = $p + 1; $i <= $HEAD_POS{$p}; $i++)
647         {
648             my @COL = split(/./, $UICC{$i});
649             my $uicc = $COL[0];
650             # splittle Spalten an semikolons
651             $COL[1] = "...";
652             $COL[2] = "...";
653             $COL[3] = "...";
654             $COL[4] = "...";
655             $COL[5] = "...";
656             $COL[6] = "...";
657             $COL[7] = "...";
658             $COL[8] = "...";
659             $COL[9] = "...";
660             # splittle TMN-Spalten an Kommas & kombiniere alle Möglichkeiten
661             foreach my $t (split(/./, $COL[1]))
662             {
663                 foreach my $m (split(/./, $COL[2]))
664                 {
665                     foreach my $n (split(/./, $COL[3]))
666                     {
667                         # TMN(E) Strings sollten alle eindeutig einem UICC zugeordnet werden
668                         $BUFF{"$t_$m_$n"} = $uicc;

```

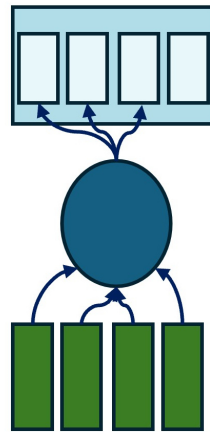
PERL-Code ca. 1200 Zeilen (Ausschnitt)

1593	IIIC	1a,1b,2a,2b : 2c,3	0
1594	IIIC	3b,4a	1,2,3
1595	IIIC	4b	1,2
1596	IIID	4b	3
1597	IV	all	1
1598			
1599	#Haut_v8_10_44.0.44.1.44.2.44.3.44.4.44.5.44.6.44.7.44.8.44.9.63.2		
1600		0 is	0
1601	I	1	0
1602	IIA	2,3	0
1603	IIB	4	0
1604	IIIA	0,1b	0
1605	IIIA	all	1a,1a_sn
1606	IIB	all	1,2,3
1607	IV	all	all
1608			
1609	#Mamma_v8_KA_50.0.50.1.50.2.50.3.50.4.50.5.50.6.50.8.50.9		
1610		0 is	0
1611	IA	1,1mi	0
1612	IB	0,1	1mi
1613	IIA	0,1,1mi	1
1614	IIA	2	0
1615	IIB	2	1
1616	IIB	3	0
1617	IIIA	0,1,1mi,2	2
1618	IIIA	3	1,2
1619	IIB	4, 0,1,2	0
1620	IIIC	all	3
1621	IV	all	1
1622	IV	all	1

Informationen der TNM-Bücher in digitaler Form, 156 Tabellen (Ausschnitt)



Informationen aus den TNM-Büchern



Hashtabelle implementiert in R-Code-Datenstruktur Environments

Abbildung 2: Erstellung einer Hashtabelle für den Abgleich der Keystrings. Die Hashtabelle in R-Code Datenstruktur Environments wurde mit PERL, unter Verwendung digital vorliegender Daten zu den TNM-Klassifikationen und der Zuordnung der Histomorphologien, erstellt.

dass für ein bestimmtes UICC-Stadium die Angaben zum T-Status irrelevant sind. Um der Menge der variablen Ausprägungen in den Meldungen gerecht zu werden, muss jedoch für jede mögliche Ausprägung des T-Status eine Zeile in die Tabelle geschrieben werden. Zudem wurde eine digitale Zuordnung der Histomorphologie-Code-Gruppen erstellt, welche innerhalb der digitalen Listen nur als Schlagworte vorliegen, wie beispielsweise „Adenokarzinom“, welches bei der Erstellung der Tabellen in einen oder mehrere Histomorphologie-Codes übersetzt wird. Diese digitalen Tabellen wurden mit PERL (Practical Extraction and Reporting Language) ausgelesen und in entsprechenden R-Code umgewandelt (Abbildung 2). Innerhalb PERL wurde dies mithilfe von Hashtabellen mit dynamischer Tiefe sowie der Zuhilfenahme von regulären Ausdrücken erreicht.

Durch verschiedene Permutationsstrategien wurden diese Hashtabellen durch künstliche Schlüssel-Werte Paare erweitert, durch welche fehlerhafte Eingabewerte abgefangen werden können, sofern diese nicht mit real existierenden Strings kollidieren. Wird z.B. für den T-Status T1c gemeldet, diese Angabe jedoch für diese Entität nicht im TNM-Buch enthalten ist, kann auf der Funktionsseite automatisch getestet werden, ob an der gleichen Position T1 möglich ist. Wenn ja, wird T1 weiterverwendet, die Anpassung jedoch mit einem Stern (\*) im zugehörigen UICC-Ausgabewert markiert (Abbildung 3, rechte Hashtabelle). Diese Anpassungen betreffen nur die T-, N- und M-Werte, mehr als zwei Anpassungen werden nicht berücksichtigt. Des Weiteren werden bei der Erzeugung der Stringschlüssel Möglichkeiten implementiert, Attributausprägungen mit hoher Variabilität durch Platzhalter („-“) zu ersetzen und auf Funktionsseite diese bei der Übersetzung optional auszulassen. So kann z.B. eine Übersetzung ganz ohne Angabe des histomorphologischen Codes versucht werden.

Das Ergebnis beinhaltet den R-Code für zwei ineinandergreifende Hashtabellen (Abbildung 3). Die erste legt gleiche UICC-Tabellen zusammen, welche in ICD-10-GM, TNM-Version und Histomorphologie-Code übereinstimmen. Diese kodiert auf einen bestimmten Gruppen-TAG (dreistellig), welcher dann zusammen mit den einzelnen TNM-Ausprägungen in einer zweiten Hashtabelle den spezifischen UICC-Wert kodiert. Diese Vorgehensweise der Zusammenlegung ist nötig, da die einzelnen Ausprägungen von ICD-10-GM, TNM-Status und Histomorphologie-Code zu mehr als 22 Millionen Kombinationsmöglichkeiten führen würde. Die daraus resultierenden Hashtabellen in R-Code bilden die Grundlage für alle weiterverarbeitenden R-Funktionen des hier vorgestellten R-Packages.

Neben den bisher genannten Angaben zur Version der verwendeten TNM-Klassifikation, dem ICD-10-GM, dem TNM-Status und der Histomorphologie können in Einzelfällen auch weitere Faktoren eine Rolle bei der Berechnung des UICC-Stadiums spielen [9]. Die Faktoren können beispielsweise Alter, bestimmte Serummarker, Mitosemarker usw. sein. Diese Marker können im verfügbaren Datensatz, wie beispielsweise dem hier verwendeten des

Krebsregisters Rheinland-Pfalz im Institut für digitale Gesundheitsdaten, unterschiedlich vollständig vorliegen. Das Alter von Patientinnen und Patienten liegt beispielsweise immer vor, Serummarker sind jedoch nicht Bestandteil der klinischen Krebsregistrierung [4] und Mitosemarker können, müssen aber bisher nicht gemeldet werden. Können UICC-Werte dadurch nicht eindeutig bestimmt werden, werden sie durch ‚|‘ separiert nebeneinander dargestellt (Abbildung 4). Kam es zu Einträgen mit Stern (\*) (Abbildung 3) durch uneindeutige TNM-Angaben und somit zu nicht eindeutigen UICC-Stadien, werden diese mit ‚./‘ separiert nebeneinander dargestellt (Abbildung 4). Fehlt die Angabe zur Histomorphologie werden die möglichen UICC-Stadien durch ‚#‘ separiert. Diese Markierungen können auch in Kombination auftreten.

Die für diesen Prozess erstellten PERL- und R-Skripte sind unter folgendem Link zugänglich: <https://github.com/IDGRLP/CanDi2>

Das hier beschriebene Vorgehen wurde für den Datensatz des Krebsregisters Rheinland-Pfalz im Institut für digitale Gesundheitsdaten entwickelt und optimiert. Auf diese Weise ist eine automatisierte Berechnung der UICC-Stadien im Krebsregister Rheinland-Pfalz möglich und Bestandteil der monatlichen Erstellung von klinischen und epidemiologischen Datengrundlagen, die aus der Produktivdatenbank generiert werden. Diese Datengrundlagen sind die Basis für sämtliche Auswertungen und Datensätze des Krebsregisters.

Der hier beschriebene Prozess kann sowohl in anderen Krebsregistern als auch bei ähnlich strukturierten Datensammlungen Anwendung finden. Hierbei kann die UICC-Berechnung für sich allein genommen werden, oder als Teil anderer Prozesse fungieren.

## Ausblick

Die Etablierung der hier vorgestellten Berechnungsmethode ist nie vollständig abgeschlossen. Derzeit wird an der Implementierung der TNM-Version 9 gearbeitet. Zudem laufen Optimierungen für einige wenige Tumorentitäten bei denen die UICC-Berechnung davon abhängt, ob der TNM-Status klinisch oder pathologisch erhoben wurde. Dies betrifft beispielsweise Ösophagus und Magen. Bisher erfolgte hierfür die UICC-Berechnung entsprechend den Regeln für den pathologischen TNM-Status.

Insgesamt zeigt die hier vorgestellte technische Lösung des Krebsregisters Rheinland-Pfalz im Institut für digitale Gesundheitsdaten, dass eine automatisierte Berechnung von UICC-Stadien unter Berücksichtigung der besten Informationen zu ICD-10-GM, Histomorphologie und TNM-Status unter Einbeziehung der unterschiedlichen Versionen der TNM-Klassifikation möglich ist. Diese Vorarbeiten können zukünftig in anderen Krebsregistern, die bisher über keine automatisierte Berechnungsmethode verfügen, aber auch für andere Datensammlungen genutzt werden, die über Angaben zum TNM-Status verfügen, um daraus UICC-Stadien zu berechnen.

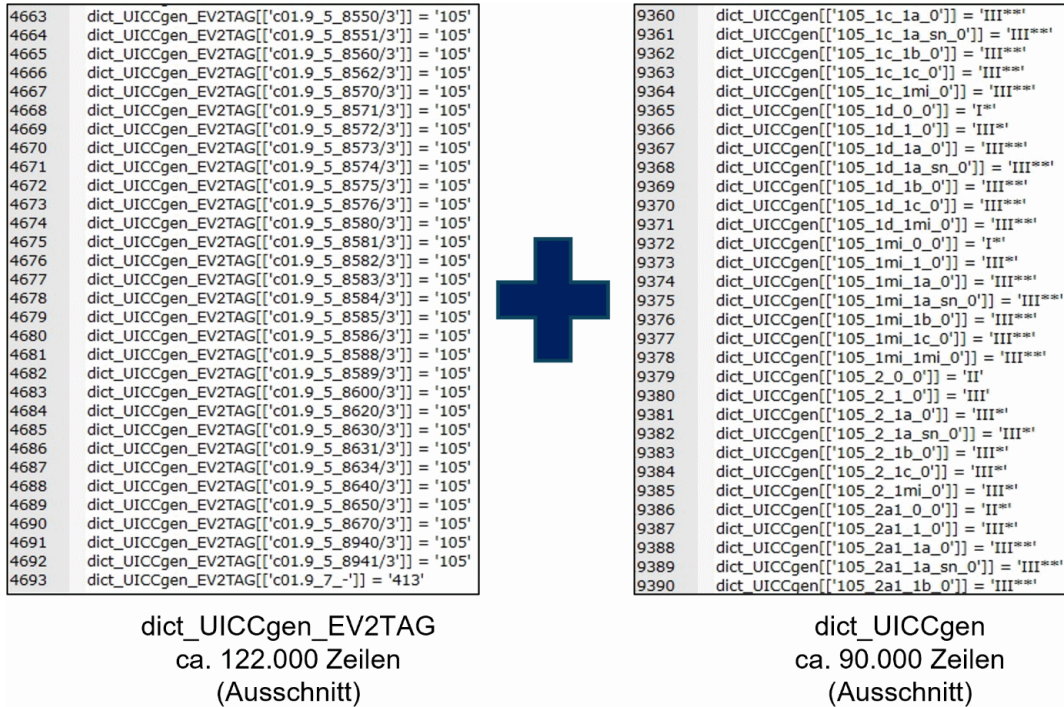


Abbildung 3: Zwei ineinandergreifende Hashtabellen. In der ersten Hashtabelle (links) werden gleiche UICC-Tabellen zusammengelegt. In der zweiten Tabelle findet sich dann der entsprechende UICC-Code, der einem spezifischen UICC-Stadium zugeordnet wird.

```

[[1]]
                                UICC
1: III|IIIA|IIIB|IIIC (Serumtumormarker x|0,1|2|3)
2:  II|IIB|IIIB|IIIC (Serumtumormarker x|0,1|2|3)
3:                                IIIB|IIIC (Serumtumormarker 2|3)
4:  II|IIB|IIIB|IIIC (Serumtumormarker x|0,1|2|3)*
5:                                I|IB (Serumtumormarker x|0)

[[2]]
                                UICC
1:                                I* # I|IIIA (Mitoserate n|h)* (Histo)
2:                                II|IIIB (Mitoserate n|h) # IIA (Histo)
3:                                I|IIIA (Mitoserate n|h) # I (Histo)
4:  I|IIIA (Mitoserate n|h)* # IA* # IA|IB (Grading 1,2,x|3)* (Histo)
5:  I|IIIB (Mitoserate n|h) # IB # IB|IIA (Grading 1,2,x|3) (Histo)

[[3]]
                                UICC
1:                                IA|IIB (Grading 1,2|3,4)*
2:  I|IIIA (Mitoserate n|h) # IA # IA|IB (Grading 1,2,x|3) (Histo)
3:  I|IIIA (Mitoserate n|h)* # IA|IB (Grading 1,2,x|3)* # IA* (Histo)
4:                                IB|IIC (Grading 1,2|3,4)/IIA|III (Grading 1,2|3,4)*
5:                                IB|IIB (Grading 1,2,x|3,4)/IB|III (Grading 1,2,x|3,4)*

[[4]]
                                UICC
1:  II|IVC (Altersgrenze 0|1) # IVC (Histo)
2:  II|IV (Altersgrenze 0|1) # IV # IVB (Histo)
3:  II|IVB (Altersgrenze 0|1) # IVC (Histo)

[[5]]
                                UICC
1:  IV|IVA|IVB*
2:  IB|IIA**
3:  III*
4:  IIA**
5:  IVB* # IV* (Histo)

[[6]]
                                UICC
1:  IV # IIIA (Histo)
2:  IVC # IV (Histo)
3:  II|IIIB (Mitoserate n|h) # II (Histo)
4:  IVB* # IV* (Histo)
5:  IIA* # II* (Histo)
                    
```

Abbildung 4: Listung und Markierung von UICC-Stadien, die durch uneindeutige oder fehlende Angaben nicht eindeutig zugeordnet werden können

## Anmerkung

## Interessenkonflikte

Die Autor:innen erklären, dass sie keine Interessenkonflikte in Zusammenhang mit diesem Artikel haben.

## Literatur

1. International Agency for Research on Cancer. Cancer Topics. [letzter Zugriff: 08.01.2026]. Verfügbar unter: <https://www.iarc.who.int/cancer-topics/#country>
2. Katalinic A, Halber M, Meyer M, Pflüger M, Eberle A, Nennecke A, Kim-Wanner SZ, Hartz T, Weitmann K, Stang A, Justenhoven C, Hollecsek B, Piontek D, Wittenberg I, Heßmer A, Kraywinkel K, Spix C, Pritzke R. Population-Based Clinical Cancer Registration in Germany. *Cancers (Basel)*. 2023 Aug 2;15(15):3934. DOI: 10.3390/cancers15153934
3. Kropf S, Burger E, Radinski I, Ridwelski K, Lippert H, Altendorf-Hofmann A, Bernarding J. Vollständigkeit und Qualität der Basisdaten und der Nachbeobachtung im Krebsregister – Eine Untersuchung am Beispiel des kolorektalen Karzinoms [Completeness and quality of baseline data and follow-up in cancer registry – an analysis on the example of colorectal cancer]. *Dtsch Med Wochenschr*. 2015 May;140(11):e106-13. DOI: 10.1055/s-0041-102171
4. Arbeitsgemeinschaft Deutscher Tumorzentren e.V. (ADT); Deutsche Krebsregister e.V. (DKR); Plattform 65c. Einheitlicher onkologischer Basisdatensatz. [letzter Zugriff: 08.01.2026]. Verfügbar unter: <https://www.basisdatensatz.de/basisdatensatz>
5. Plachky P, Mussel C, Justenhoven C. Krebsregister erhalten umfassende Informationen pro Meldeanlass. *Forum*. 2025;6:404-7.
6. Bundesinstitut für Arzneimittel und Medizinprodukte. ICD-10-GM. [letzter Zugriff: 08.01.2026]. Verfügbar unter: [https://www.bfarm.de/DE/Kodiersysteme/Klassifikationen/ICD/ICD-10-GM/\\_node.html](https://www.bfarm.de/DE/Kodiersysteme/Klassifikationen/ICD/ICD-10-GM/_node.html)
7. Krebsinformationsdienst. Histologische und zytologische Diagnostik in der Krebsmedizin. [letzter Zugriff: 08.01.2026]. Verfügbar unter: <https://www.krebsinformationsdienst.de/untersuchungen-bei-krebs/histologische-diagnostik>
8. Onko Internetportal. Klassifikation von Tumoren (TNM-System & Grading). [letzter Zugriff: 08.01.2026]. Verfügbar unter: <https://www.krebsgesellschaft.de/onko-internetportal/basisinformationen-krebs/basisinformationen-krebs-allgemeine-informationen/klassifikation-von-tumoren-tnm.html>
9. Wittekind C. *TNM Klassifikation maligner Tumoren*. 8. Aufl. Wiley-VCH; 2020. ISBN: 978-3-527-34772-8
10. O'Sullivan B, Brierley JD, D'Cruz AK, Fey MF, Pollock R, Vermorken JB, Huang SH. *UICC Manual of Clinical Oncology*. Wiley-VCH; 2015. ISBN: 9781444332445
11. Gupta S, Aitken JF, Bartels U, Brierley J, Dolendo M, Friedrich P, Fuentes-Alabi S, Garrido CP, Gatta G, Gospodarowicz M, Gross T, Howard SC, Molyneux E, Moreno F, Pole JD, Pritchard-Jones K, Ramirez O, Ries LAG, Rodriguez-Galindo C, Shin HY, Steliarova-Foucher E, Sung L, Supriyadi E, Swaminathan R, Torode J, Vora T, Kutluk T, Frazier AL. Paediatric cancer stage in population-based cancer registries: the Toronto consensus principles and guidelines. *Lancet Oncol*. 2016 Apr;17(4):e163-e172. DOI: 10.1016/S1470-2045(15)00539-2
12. Deutsche Krebsgesellschaft e.V. Leitlinienprogramm Onkologie. [letzter Zugriff: 08.01.2026]. Verfügbar unter: <https://www.leitlinienprogramm-onkologie.de/home>
13. Hermes-Moll K, Baumann W, Kowalski C, Ohlmeier C, Gothe H, Heidt V. Multidisziplinäre Tumorkonferenzen in Deutschland. *Monitor Versorgungsforschung*. 2021;5:57-61. DOI: 10.24945/MVF.05.21.1866-0533.2346
14. Krebsinformationsdienst. *Forschen zu Krebs*. [letzter Zugriff: 08.01.2026]. Verfügbar unter: <https://www.krebsinformationsdienst.de/forschung>
15. Arbeitsgemeinschaft Deutscher Tumorzentren e.V. (ADT); Gesellschaft der epidemiologischen Krebsregister in Deutschland (GEKID). *Manual der klinischen und epidemiologischen Krebsregistrierung 2018*. 2018. ISBN: 978-3-86371-165-8
16. Plattform 65c. *Arbeitsgruppen*. [letzter Zugriff: 08.01.2026]. Verfügbar unter: <https://plattform65c.de/wer-wir-sind/arbeitsgruppen/>
17. Wittekind C. *TNM 2010. Was ist neu? [TNM 2010. What's new?]*. *Pathologe*. 2010 Oct;31 Suppl 2:153-60. DOI: 10.1007/s00292-010-1341-y
18. Brierley JD, Giuliani M, O'Sullivan B, Rous B, Van Eycken E, editors. *TNM Classification of Malignant Tumours*. John Wiley & Sons; 2025. ISBN: 978-1-394-21687-1
19. R Core Team. *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing; [letzter Zugriff: 20.11.2025]. Verfügbar unter: <https://www.R-project.org/>
20. RStudio Team. *RStudio: Integrated Development for R*. Boston: RStudio; [letzter Zugriff: 20.11.2025]. Verfügbar unter: <https://www.rstudio.com>

### Korrespondenzadresse:

Benjamin Rieger  
Krebsregister Rheinland-Pfalz im Institut für digitale Gesundheitsdaten gGmbH, Hechtsheimer Str. 37, 55131 Mainz, Deutschland  
rieger@idg-rlp.de

### Bitte zitieren als

Rieger B, Mussel C, Reinwald F, Branscheidt S, Justenhoven C, Hamann R. *Automatisierte Berechnung der UICC-Stadien anhand von Krebsregisterdaten unter Berücksichtigung von verschiedenen TNM-Versionen*. *GMS Med Inform Biom Epidemiol*. 2026;22:Doc09. DOI: 10.3205/mibe000307, URN: urn:nbn:de:0183-mibe0003072

### Artikel online frei zugänglich unter

<https://doi.org/10.3205/mibe000307>

Veröffentlicht: 12.05.2026

### Copyright

©2026 Rieger et al. Dieser Artikel ist ein Open-Access-Artikel und steht unter den Lizenzbedingungen der Creative Commons Attribution 4.0 License (Namensnennung). Lizenz-Angaben siehe <http://creativecommons.org/licenses/by/4.0/>.