



Evaluierung Datenanalyse und Analyse von Modellierungsmethoden für Krebserkrankungen und Erkrankungen bei Kindern und Jugendlichen

Teilprojekt: Krebserkrankungen

Autoren: DI Günther Zauner
DI Patrick Einzinger
DI Nikolas Popper
Dr. Gottfried Endel
Dr. Irmgard Schiller-Frühwirth, MPH

11. April 2011



Inhalt

A. Detailbetrachtung spezieller Krebserkrankungen auf Basis der MBDS Datensätze in GAP-DRG	
1 Aufgabenstellung	3
2 Vorgangsweise – Datenextraktion aus GAPDRG	3
3 Leberkrebs	5
4 Lungenkrebs.....	7
5 Brustkrebs	10
6 Darmkrebs.....	13
7 Bauchspeicheldrüsenkrebs	16
8 Speiseröhre.....	19
9 Niere.....	22
10 Prostata.....	25
11 Hoden.....	27
12 Gebärmutterhalskrebs (Zervix)	29
13 Eierstock.....	31
14 Resümee.....	33
15 Qualitätssicherung – Vergleich mit Krebsregister	34
B. Detailbetrachtung der ATC-ICD Ergebnisse speziell für Krebsdiagnosen	
1 Aufgabenstellung	36
2 Vorgangsweise bei den Analysen.....	36
3 Analyse und Interpretation für die ICD9 Gruppe 170-175	38
4 Analyse und Interpretation für die ICD9 Gruppe 150-159	39
5 Analyse und Interpretation für die ICD9 Gruppe 160-165	40
6 Analyse und Interpretation für die ICD9 Gruppe 230-234 (Carcinoma in situ)	42
7 Resümee.....	43
C. Kontrolle der Größenordnung von Todesfällen auf Grund von Krebs bei Kindern – Vergleich mit Statistik Austria Daten	
D. Krebserkrankungen – Literaturrecherche	45
1 Allgemeine Zielsetzung	45
2 Literaturrechercheergebnisse	45
E. Technische Realisierung der Abfragen zu ausgewählten Krebserkrankungen am Beispiel Magenkrebs	

A. Detailbetrachtung spezieller Krebserkrankungen auf Basis der MBDS Datensätze in GAP-DRG

1 Aufgabenstellung

Ziel ist es auf Basis der in GAP-DRG gesammelten anonymisierten Personendatensätze der Jahre 2006 und 2007 die Krebsneuerkrankungen des Kalenderjahres 2007 zu erkennen und nach Geschlecht und Alter zu klassifizieren. Anschließend werden die erhaltenen Fallzahlen mit Daten des Krebsregisters der Statistik Austria verglichen und neben den Absolutwerten die Inzidenzen auf Basis der Bevölkerungszahlen für das Jahr 2007 aus den Daten der Statistik Austria berechnet.

Da Krebserkrankungen im Allgemeinen einen stationären Aufenthalt nach sich ziehen wird zunächst festgelegt welche ICD10 Diagnosen(gruppen) in die Analysen aufgenommen werden. Neben den Diagnosecodes für bösartige Neubildungen (C00-C97) wird ein Teil der mit D bezeichneten Diagnosen berücksichtigt. Generell werden dabei folgende beiden Gruppen als notwendig identifiziert:

1. D00-D09 In-situ-Neubildungen
2. D37-D48 Neubildungen unsicheren oder unbekanntem Verhaltens

Die Gruppe der Gutartigen Neubildungen (D10-D36) wird ausgeschlossen. Zu beachten ist, dass in den Neubildungen unsicheren oder unbekanntem Verhaltens in mehreren Krebserkrankungen unterschiedlicher Lokationen zusammengefasst sind.

Eine detaillierte Beschreibung der Abfrageplanung wird im nächsten Kapitel vorgenommen. Die für die konkreten Krebserkrankungen durchgeführten SQL-Abfragen werden in einem gesonderten kommentierten Dokument dem Bericht beigelegt, sie finden sich im Anhang.

2 Vorgangsweise – Datenextraktion aus GAPDRG

Die Datenextraktion und die Filterung nach Neuerkrankungen erfolgt in mehreren Stufen und bedarf aus Qualitätssicherungsgründen einer Trennung in mehrere Schritte und mehrmaligen Kontrolle (zum Beispiel wird das Alter einer identifizierten Person und ihrer zugeordneten Datensätze überprüft; gleiches geschieht für das Geschlecht).

1. Im ersten Schritt werden zu den für eine spezielle Krebserkrankung vorgegebenen ICD10 – Diagnosen alle Krankenhausaufenthalte aus den MBDS – Daten erhoben.
2. Es werden alle Personen IDs mit einem Krankenhausaufenthalt mit entsprechender Diagnose für das Jahr 2007 gefiltert.
3. Diese Personenliste wird um die Personen IDs reduziert, zu denen im Kalenderjahr 2006 ein Krankenhausaufenthalt mit einschlägiger Diagnose existiert. Dies geschieht, da man Neuerkrankungen (Inzidenzfälle) erkennen will, nicht das gesamte im Jahr 2007 erbrachte Leistungsaufkommen.
4. Da i.A. nach einem Erstaufenthalt bei einer einschlägigen Diagnose ein weiterer therapeutischer Krankenhausaufenthalt folgt, werden im folgenden Schritt für alle bisher als neu klassifizierten Fälle um jene verringert, die in den Monaten Jänner bis Oktober 2007 einen Erstaufenthalt hatten und in den weiteren Daten (Datenspektrum reicht bis inkl. Dezember 2007) keinen Folgeaufenthalt hatten.

Die Einschränkung auf die ersten zehn Kalendermonate wird durchgeführt, da über das Jahr 2007 hinaus keine Daten vorliegen. Außerdem hat man für die Monate November bzw. Dezember bereits eine Zeitspanne von mindestens 22 Monaten in denen man alle möglichen vorhandenen Krankenhausaufenthalte erkennt. Daher verringert sich die



Wahrscheinlichkeit einen Nachbehandlungstermin bzw. eine Kontrolle als Erstaufenthalt zu erkennen stark.

5. Analyse der abgegebenen Medikamente zu den gefundenen Personen im Zeitbereich vor dem ersten gefundenen Krankenhausaufenthalt.
Filterung der Personen die keine Medikamente erhalten haben die eindeutig der Therapie der zu untersuchenden Krebserkrankung zuzuordnen sind.
→ Die potentiellen Medikamentengruppen (ATC-Gruppen) werden durch Zuhilfenahme der Ergebnisse des ATC-ICD Projekts analysiert. Die Resultate sind in Berichtteil B dargestellt.
6. Sobald das MetaHono - Matching in GAP-DRG und dessen Validierung abgeschlossen ist, kann als nächster Analyseschritt für die gefundenen Personengruppen nach Leistungen im niedergelassenen Bereich gefiltert werden.
Dabei kann die Qualitätssicherung in zwei Richtungen erfolgen:
 - a. Durch Analyse des Zeitraums vor dem ersten Krankenhausaufenthalt werden Fälle abgesichert. Das bedeutet, es wird die Wahrscheinlichkeit eines Codierungsfehlers in einer Krankenhaus gemindert.
 - b. Es können weitere Fälle, die keine Neuerkrankung darstellen, erkannt werden.

Die in Aufzählungspunkt 5 angeführte Integration von Medikamentendaten in die Ausschlusskriterien für neue Fälle wird ebenso wie Punkt 6 vorbereitet. In den aktuell dargestellten Ergebnissen sind sie nicht enthalten. Wie aus den folgenden Vergleichen mit Daten der Statistik Austria zu ersehen ist, wird auch ohne diese Kriterien bereits ein sehr hoher Übereinstimmungsgrad erreicht.

Dennoch können diese Ansätze falls erforderlich für weiterführende Analysen und die Berechnung ökonomischer Kennzahlen verwendet werden.

In den folgenden Kapiteln werden für ausgewählte Krebserkrankungen die Detailergebnisse dargestellt.

3 Leberkrebs

Basierend auf den in der Aufgabenstellung beschriebenen Fragestellungen werden Leberkarzinome im Detail analysiert. Dazu werden folgende Diagnosen herangezogen:

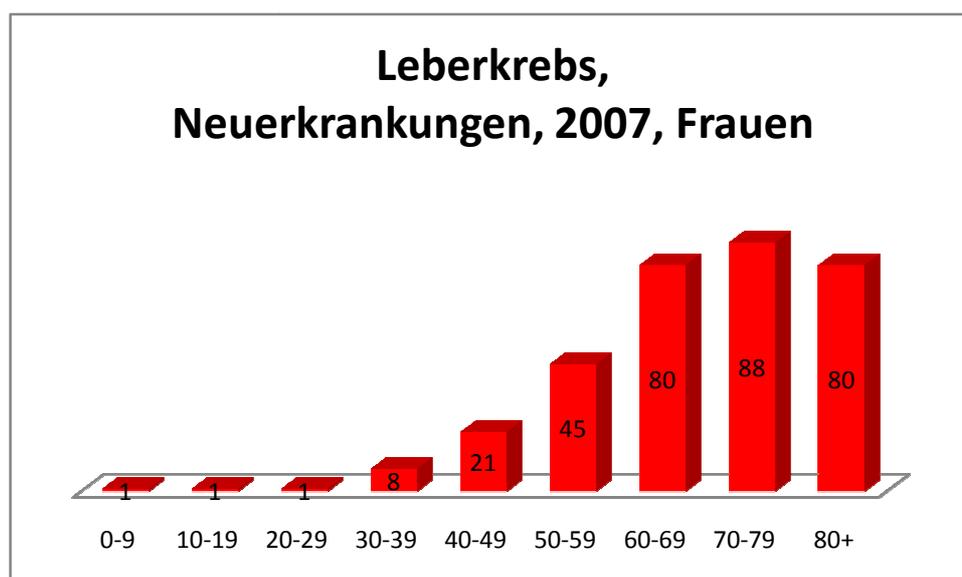
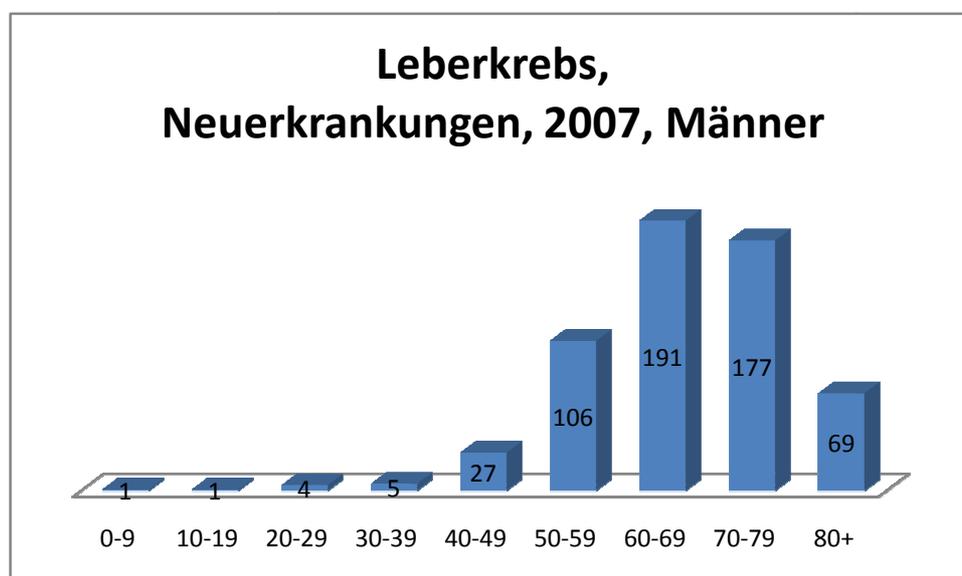
D01.5 Leber, Gallenblase und Gallengänge

Inkl.: Ampulla hepatopancreatica [Ampulla Vateri]

D37.6 Leber, Gallenblase und Gallengänge

Inkl.: Ampulla hepatopancreatica [Ampulla Vateri]

Neben der Gesamtanzahl an Neuerkrankungen von 906 Fällen für das Jahr 2007 ist die Darstellung der Verteilung nach Absolutzahlen und Altersklassen von besonderem Interesse. Die folgenden beiden Graphiken zeigen die Fallzahlen für Männer und Frauen getrennt.



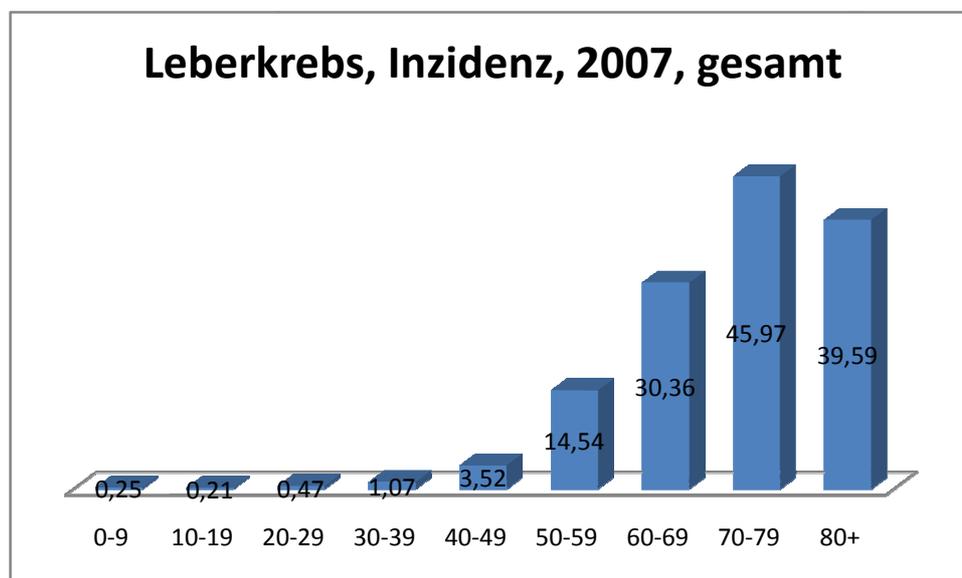
Aus diesen Graphiken erkennt man einen deutlichen Anstieg der Neuerkrankungsfälle ab einem Alter von 50-59 Jahren. Eine reduzierte Fallzahl für die Gruppe der über 80 Jährigen bedeutet nicht gesichert eine geringere Inzidenz da die Personenanzahl in dieser Altersgruppe geringer ist, als in der Altersklasse der 70-79 Jährigen. Diese muss noch gesondert analysiert werden.

Zur Berechnung der Inzidenz werden die Bevölkerungsdaten der Statistik Austria für das Jahr 2007 zu Grunde gelegt. Man erhält die in der nachstehenden Tabelle angeführten Inzidenzen in Fällen pro 100.000 Einwohner und Jahr für 2007:

Altersgruppe	männlich	weiblich	gesamt
0-9	0,24	0,25	0,25
10-19	0,20	0,21	0,21
20-29	0,75	0,19	0,47
30-39	0,82	1,32	1,07
40-49	3,91	3,11	3,52
50-59	20,71	8,55	14,54
60-69	44,95	17,11	30,36
70-79	71,46	26,76	45,97
80+	62,27	30,13	39,59

Klar ersichtlich ist aus der tabellarischen Darstellung eine stark erhöhte Inzidenz bei Männern im Vergleich zu Frauen für die Altersgruppen ab 40 Jahren. Für die Altersgruppe der 70-79 Jährigen erreicht die Neuerkrankungszahl von 71,46 Personen pro 100.000 Männern dieser Altersgruppe und Jahr ihr Maximum.

Die folgende Darstellung zeigt die Ergebnisse bezüglich der geschlechterneutralen Inzidenzen (in Fällen pro 100.000 Einwohner und Jahr für 2007) in Form eines Säulenplots:



Abschließend kann aus den berechneten Daten ersehen werden, dass die Anzahl an Neuerkrankungen bei den unter 40 Jährigen minimal ist. Darüber hinaus erkennt man, dass die aus der GAP-DRG Datenbank identifizierten Fälle (906), sehr genau mit den Daten der Statistik Austria

(http://www.statistik.at/web_de/statistiken/gesundheit/krebserkrankungen/leber/021756.html), die 892 Neuerkrankungsfälle berichten, übereinstimmt.

4 Lungenkrebs

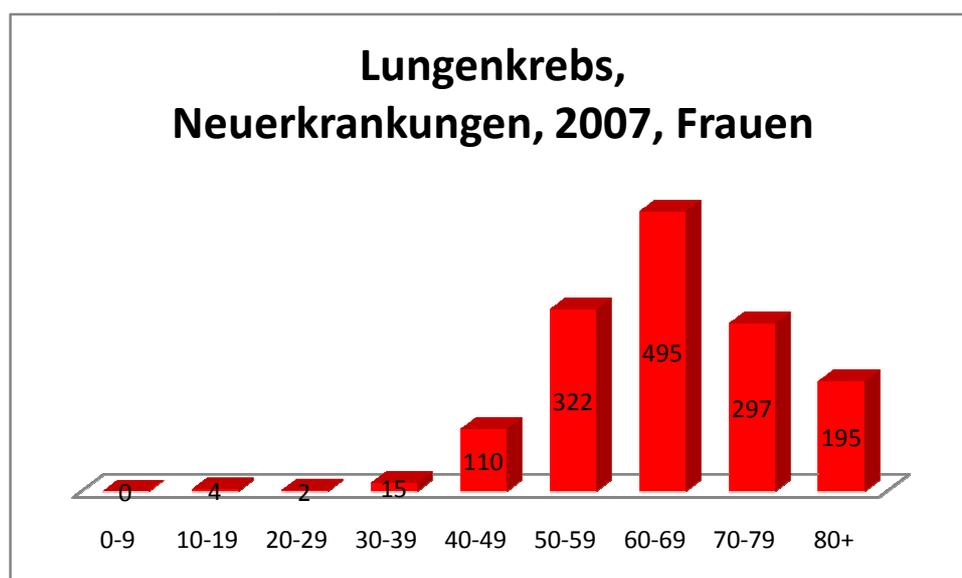
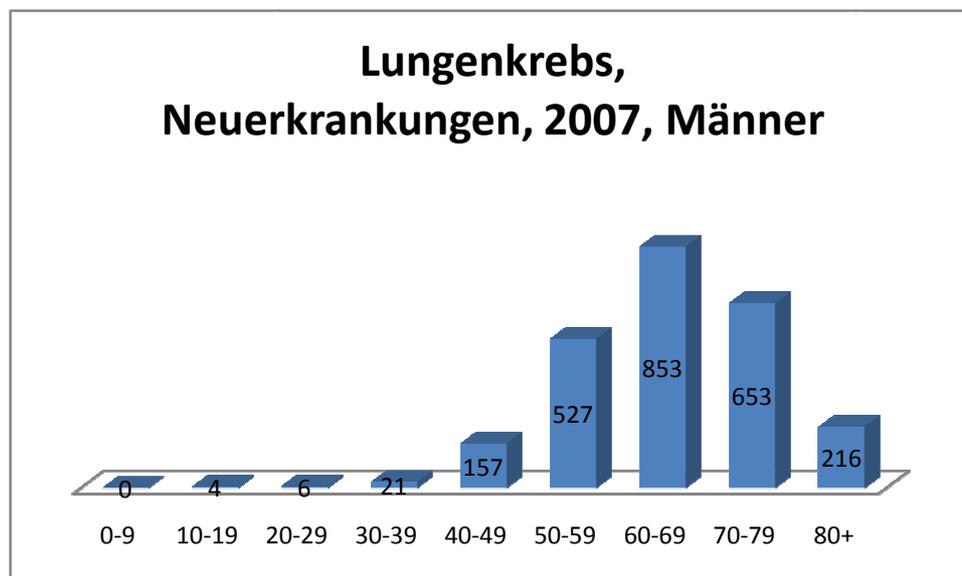
Basierend auf den in der Aufgabenstellung beschriebenen Fragestellungen werden Lungenkarzinome im Detail analysiert. Dazu werden folgende Diagnosen berücksichtigt und zugehörige Aufenthalte aus den MBDS – Datensätzen analysiert:

C34.X Bösartige Neubildung der Bronchien und der Lunge

D02.2 Bronchus und Lunge

D38.1 Trachea, Bronchus und Lunge

Neben der Gesamtanzahl an Neuerkrankungen von 3877 Fällen für das Jahr 2007 ist die Darstellung der Verteilung nach Absolutzahlen und Altersklassen von besonderem Interesse. Die folgenden beiden Graphiken zeigen die Fallzahlen für Männer und Frauen getrennt.





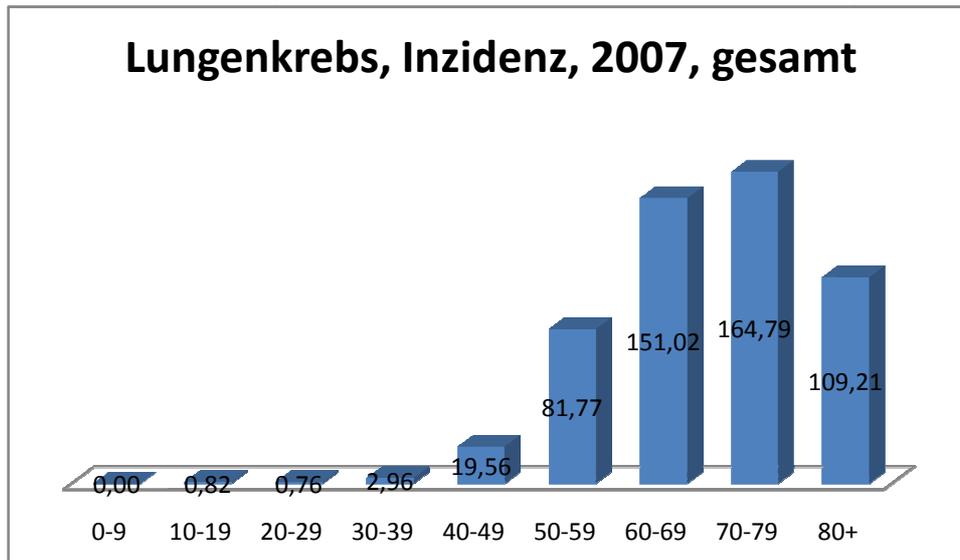
Aus diesen Graphiken erkennt man einen Anstieg der Neuerkrankungsfälle ab einem Alter von 40-49 Jahren und einen deutlichen Anstieg ab einem Alter von 50. Die Darstellung in Form eines Säulendiagramms legt eine normalverteilte Erkrankungslast mit Maximum in der Altersgruppe der 60-69 Jährigen nahe. Das Maximum der Absolutfallzahlen wird bei Männern und Frauen im gleichen Altersintervall erreicht. Eine reduzierte Fallzahl für die Gruppe der über 80 Jährigen bedeutet nicht gesichert eine geringere Inzidenz. Diese muss noch gesondert analysiert werden.

Zur Berechnung der Inzidenz werden die Bevölkerungsdaten der Statistik Austria für das Jahr 2007 zu Grunde gelegt. Man erhält die in der nachstehenden Tabelle angeführten Inzidenzen in Fällen pro 100.000 Einwohner und Jahr für 2007:

Altersgruppe	männlich	weiblich	gesamt
0-9	0,00	0,00	0,00
10-19	0,80	0,84	0,82
20-29	1,13	0,38	0,76
30-39	3,45	2,47	2,96
40-49	22,74	16,31	19,56
50-59	102,95	61,17	81,77
60-69	200,73	105,85	151,02
70-79	263,63	90,33	164,79
80+	194,94	73,43	109,21

Klar ersichtlich ist aus der tabellarischen Darstellung eine stark erhöhte Inzidenz bei Männern im Vergleich zu Frauen für die Altersgruppen ab 40 Jahren. Für die Altersgruppe der 70-79 Jährigen erreicht die Neuerkrankungsanzahl von 263,63 Personen pro 100.000 Männern dieser Altersgruppe und Jahr ihr Maximum. Das bedeutet, dass die Absolutanzahl an Neuerkrankungen und die höchste Inzidenz in unterschiedlichen Altersgruppen der Männer auftreten. Die Erkrankungswahrscheinlichkeit von Frauen liegt in den stark betroffenen Altersgruppen bei nur etwa der Hälfte der Rate von Männern.

Die folgende Darstellung zeigt die Ergebnisse bezüglich der geschlechterneutralen Inzidenzen (in Fällen pro 100.000 Einwohner und Jahr für 2007) in Form eines Säulenplots.



Abschließend kann aus den berechneten Daten ersehen werden, dass die Anzahl an Neuerkrankungen bei den unter 40 Jährigen minimal ist. Darüber hinaus erkennt man, dass die aus der GAP-DRG Datenbank identifizierten Fälle (3877), mit den Daten der Statistik Austria (http://www.statistik.at/web_de/statistiken/gesundheit/krebserkrankungen/luftroehre_bronchien_lunge/021766.html), die 4.087 Neuerkrankungsfälle berichten, tendenziell übereinstimmt. Abweichungen können unter anderem aus nicht berücksichtigten weiteren Diagnosen oder auf Grund der Nichtbeachtung von Prikraf-Daten resultieren.

5 Brustkrebs

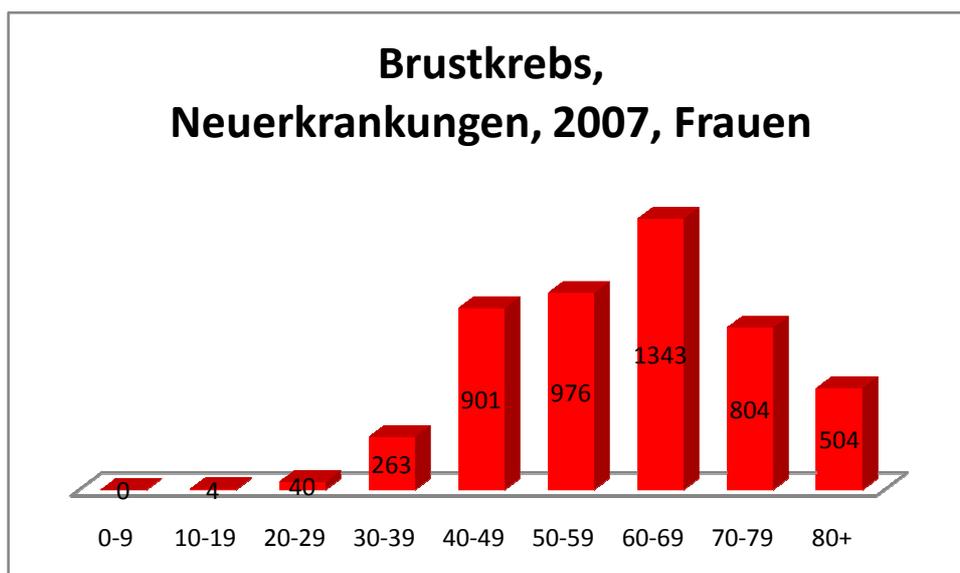
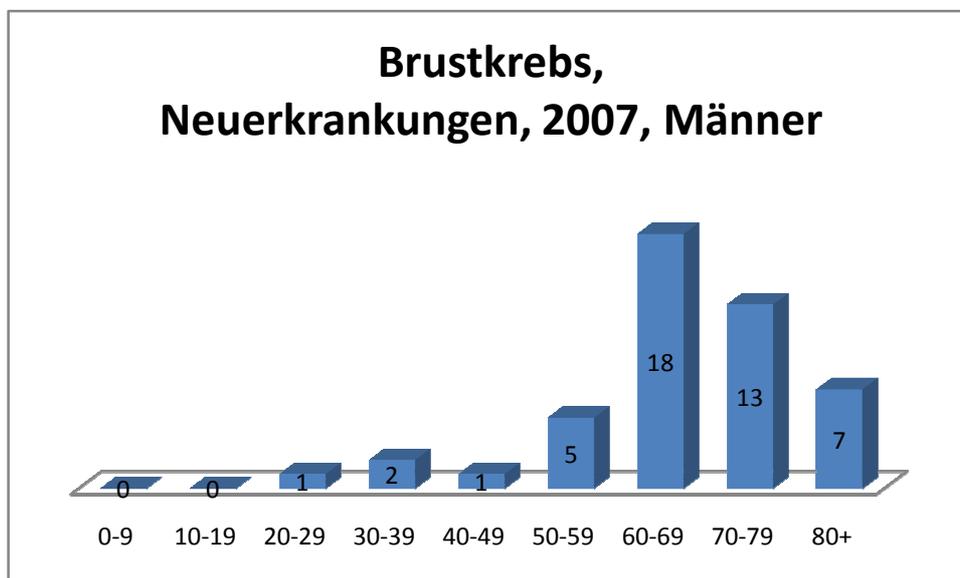
Basierend auf den in der Aufgabenstellung beschriebenen Fragestellungen werden Mammakarzinome im Detail analysiert. Dazu werden folgende Diagnosen herangezogen:

C50.X Bösartige Neubildung der Brustdrüse [Mamma]

D05.X Carcinoma in situ der Brustdrüse [Mamma]

D48.6 Brustdrüse [Mamma]

Neben der Gesamtanzahl an Neuerkrankungen von 4882 Fällen für das Jahr 2007 ist die Darstellung der Verteilung nach Absolutzahlen und Altersklassen von besonderem Interesse. Die folgenden beiden Graphiken zeigen die Fallzahlen für Männer und Frauen getrennt.



Aus diesen Graphiken erkennt man einen Anstieg der Neuerkrankungsfälle ab einem Alter von 30-39 Jahren bei Frauen und einen deutlichen Anstieg ab einem Alter von 40 bei Frauen. Da die Fallzahlen für Männer (in Summe 47 Fälle in allen Altersgruppen gesamt im Jahr 2007) im Vergleich zu denen der Frauen minimal sind, werden die folgenden Aussagen nur für Frauen getroffen.

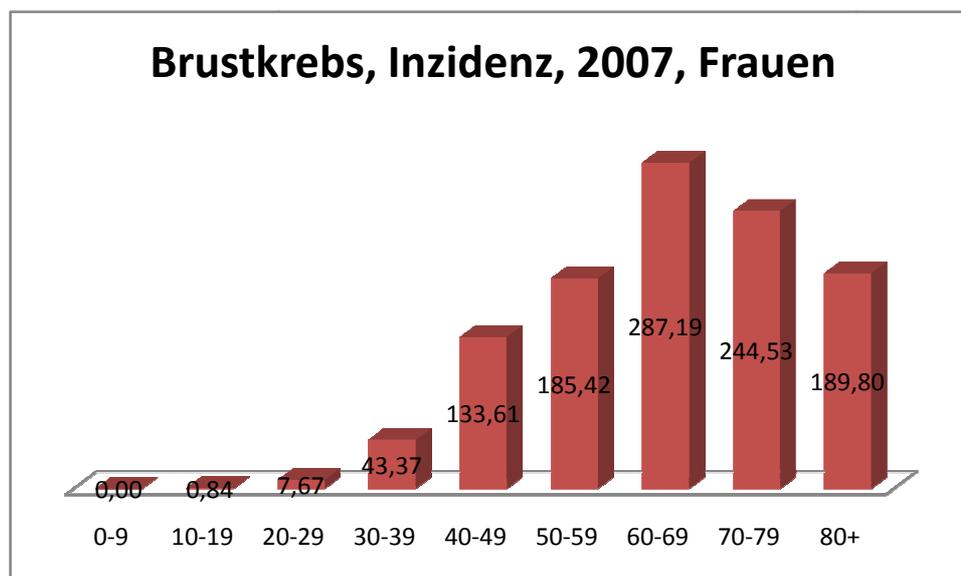
Eine reduzierte Fallzahl für die älteren Personengruppen bedeutet wieder nicht gesichert eine geringere Inzidenz. Diese wird gesondert betrachtet.

Zur Berechnung der Inzidenz werden die Bevölkerungsdaten der Statistik Austria für das Jahr 2007 zu Grunde gelegt. Man erhält die in der nachstehenden Tabelle angeführten Inzidenzen:

Altersgruppe	männlich	weiblich	gesamt
0-9	0,00	0,00	0,00
10-19	0,00	0,84	0,41
20-29	0,19	7,67	3,89
30-39	0,33	43,37	21,82
40-49	0,14	133,61	66,09
50-59	0,98	185,42	94,48
60-69	4,24	287,19	152,48
70-79	5,25	244,53	141,72
80+	6,32	189,80	135,78

Klar ersichtlich ist aus der tabellarischen Darstellung genau wie in den Absolutzahlen eine stark erhöhte Inzidenz bei Frauen im Vergleich zu Männern für alle Altersgruppen. Für die Altersgruppe der 60-69 Jährigen erreicht die Neuerkrankungsanzahl von 287,19 Personen pro 100.000 Frauen dieser Altersgruppe und Jahr ihr Maximum. Aber bereits ab der Altersgruppe der 40-49 Jährigen liegt die Neuerkrankungsrate über 100 Frauen pro 100.000 und Kalenderjahr und damit über einem Promille der gesamten weiblichen Bevölkerung dieser Altersgruppe.

Die folgende Darstellung zeigt die Ergebnisse der Inzidenzen für Frauen in Form eines Säulenplots:





Abschließend kann aus den berechneten Daten ersehen werden, dass die Anzahl an Neuerkrankungen bei den unter 30 Jährigen minimal ist. Darüber hinaus erkennt man, dass die aus der GAP-DRG Datenbank identifizierten Fälle (4882), sehr genau mit den Daten der Statistik Austria

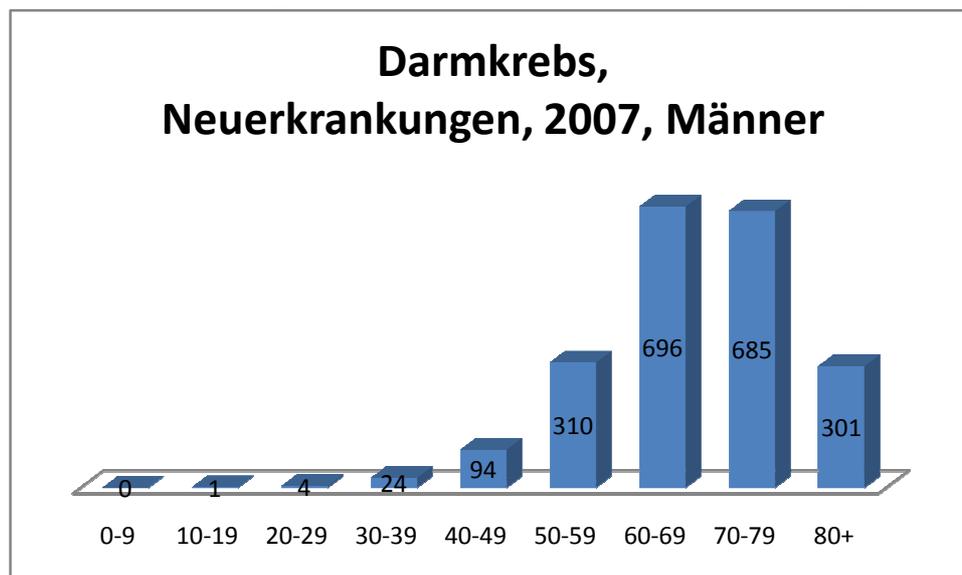
(http://www.statistik.at/web_de/statistiken/gesundheit/krebserkrankungen/brust/021696.html), die 4.833 Neuerkrankungsfälle berichten, übereinstimmt.

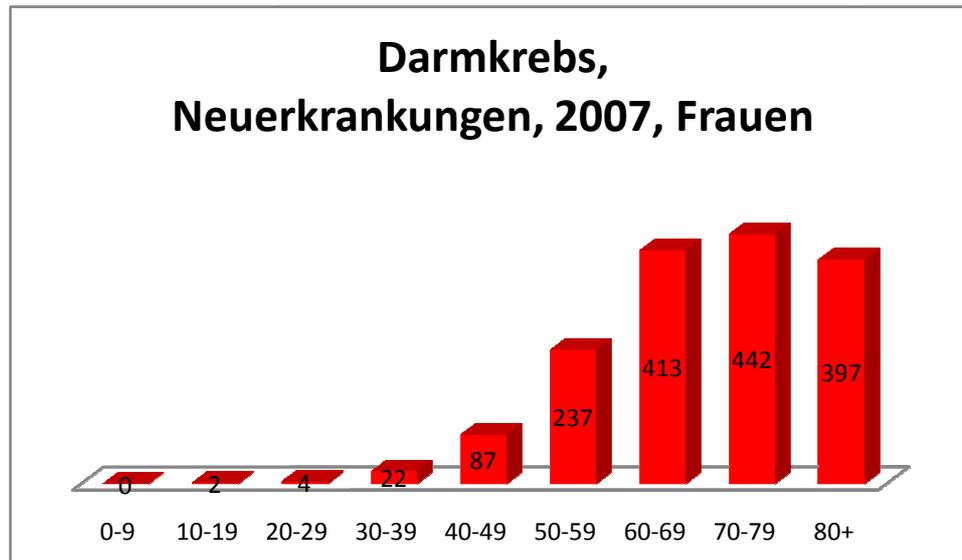
6 Darmkrebs

Basierend auf den in der Aufgabenstellung beschriebenen Fragestellungen werden Darmkarzinome im Detail analysiert. Dazu werden folgende Diagnosen herangezogen:

- C18.X Bösartige Neubildung des Kolons
- C19.X Bösartige Neubildung am Rektosigmoid, Übergang
- C20.X Bösartige Neubildung des Rektums
- C21.X Bösartige Neubildung des Anus und des Analkanals
- D01.0 Kolon
- D01.1 Rektosigmoid, Übergang
- D01.2 Rektum
- D01.3 Analkanal und Anus
- D37.4 Kolon

Neben der Gesamtanzahl an Neuerkrankungen von 3719 Fällen für das Jahr 2007 ist die Darstellung der Verteilung nach Absolutzahlen und Altersklassen von besonderem Interesse. Die folgenden beiden Graphiken zeigen die Fallzahlen für Männer und Frauen getrennt.





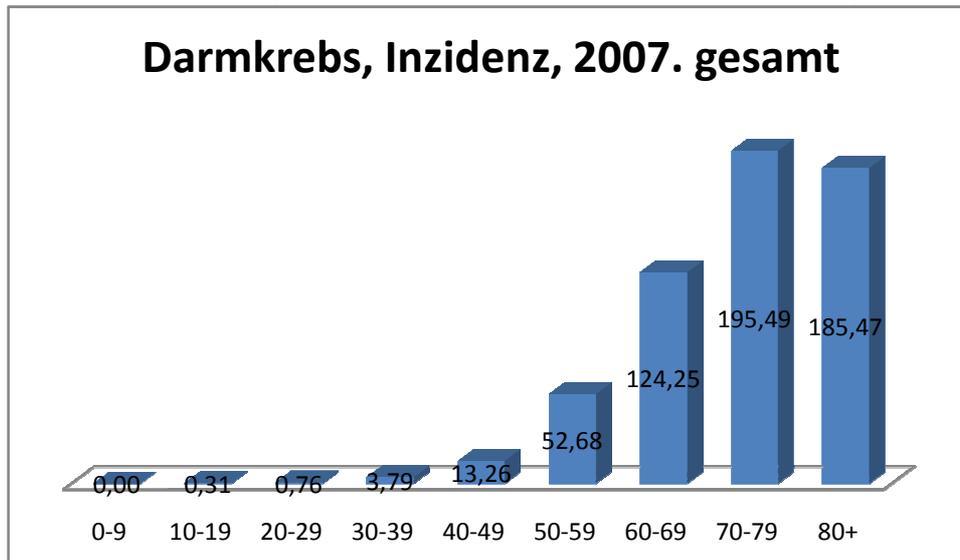
Aus diesen Graphiken erkennt man einen deutlichen Anstieg der Neuerkrankungsfälle ab einem Alter von 40-49 Jahren. Interessant ist, dass die Fallzahlen für Männer in den Altersgruppen der 50-79 Jährigen die der Frauen signifikant übertrifft. Die Altersgruppe der 80+ Jährigen wird hingegen bei den Absolutfallzahlen von den Frauen dominiert. Dennoch erhält man daraus keine Aussage zur Inzidenz, da die Lebenserwartung von Frauen in Österreich um ca. 5 Jahre höher ist und somit die Altersgruppe der 80+ Jährigen zu über 2/3 aus Frauen besteht.

Zur Berechnung der Inzidenz werden die Bevölkerungsdaten der Statistik Austria für das Jahr 2007 zu Grunde gelegt. Man erhält die in der nachstehenden Tabelle angeführten Inzidenzen:

Altersgruppe	männlich	weiblich	gesamt
0-9	0,00	0,00	0,00
10-19	0,20	0,42	0,31
20-29	0,75	0,77	0,76
30-39	3,95	3,63	3,79
40-49	13,61	12,90	13,26
50-59	60,56	45,03	52,68
60-69	163,79	88,32	124,25
70-79	276,54	134,43	195,49
80+	271,65	149,50	185,47

Klar ersichtlich ist aus der tabellarischen Darstellung eine stark erhöhte Inzidenz bei Männern im Vergleich zu Frauen für die Altersgruppen ab 60 Jahren. Die Erkrankungswahrscheinlichkeit liegt hier bei ca. dem doppelten Wert der Frauen. Für die Altersgruppe der 70-79 Jährigen erreicht die Neuerkrankungsanzahl von 276,54 Personen pro 100.000 Männern dieser Altersgruppe und Jahr ihr Maximum. Auch ohne Einbezug des Geschlechts sind Personen des Alters 70+ die Hauptrisikogruppe.

Die folgende Darstellung zeigt die Ergebnisse bezüglich der geschlechterneutralen Inzidenzen in Form eines Säulenplots:



Abschließend kann aus den berechneten Daten ersehen werden, dass die Wahrscheinlichkeit an Darmkrebs zu erkranken bei unter 40 Jährigen minimal ist. Darüber hinaus erkennt man, dass die aus der GAP-DRG Datenbank identifizierten Fälle (3719), nur zu ca.75% der von der Statistik Austria

(http://www.statistik.at/web_de/statistiken/gesundheit/krebserkrankungen/dickdarm_enddarm/021701.html), die 4610 Neuerkrankungsfälle berichten, erfasst. Diese Abweichung kann mit der Einschränkung der in diese Betrachtung und Analyse aufgenommenen Diagnosen erklärt werden. Eine weitere Begründung der reduzierten Fallzahlen erhält man aus den vorgegebenen Regeln zu vorhergehenden Aufenthalten. Dadurch wird verhindert, dass erneute Aufnahmen einer Person (auch in anderen Krankenanstalten, versichert bei einem anderen Träger, ...) als neue Fälle erkannt werden.

7 Bauchspeicheldrüsenkrebs

Basierend auf den in der Aufgabenstellung beschriebenen Fragestellungen werden Bauchspeicheldrüsenkarzinome im Detail analysiert. Dazu werden folgende Diagnosen herangezogen:

C25.X Bösartige Neubildung des Pankreas

D01.7 Sonstige näher bezeichnete Verdauungsorgane

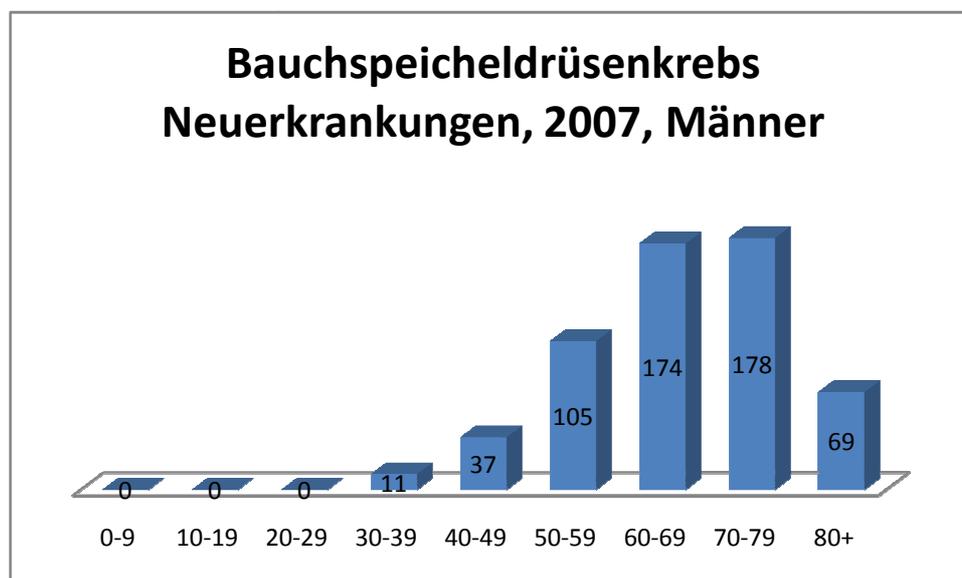
Inkl.:Pankreas

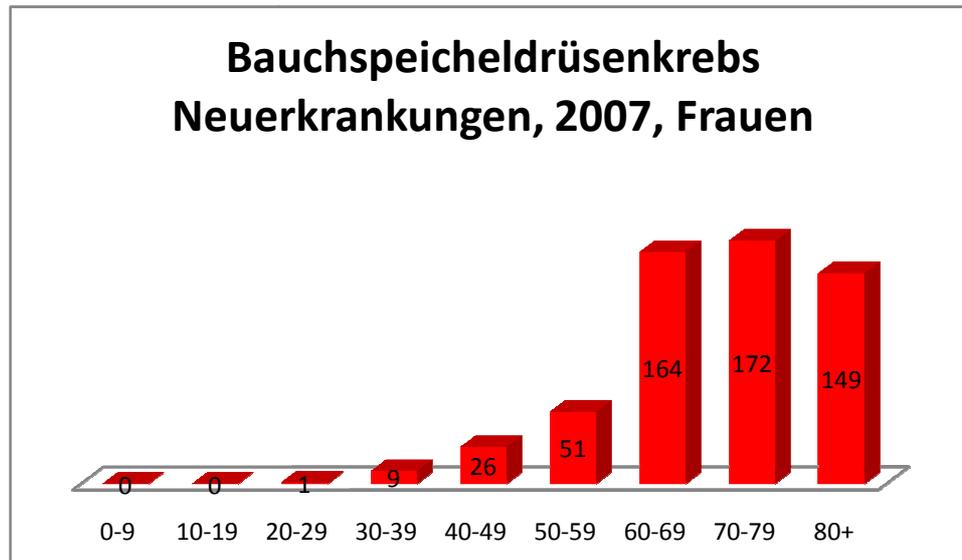
D37.7 Sonstige Verdauungsorgane

Inkl.: Anus o.n.A., Canalis analis, Darm o.n.A., Ösophagus, Pankreas, Sphincter ani

Zu beachten ist bei den hier angegebenen Diagnosen, dass speziell D37.7 eine nicht sehr detaillierte Beschränkung vorgibt. Es kann daher zu falsch aufgenommenen Fällen kommen. Für neue Erkrankungen die vor November 2007 auftreten wird die Fehlerwahrscheinlichkeit auf Grund des geforderten Folgeaufenthalts minimiert. Da D37-D48 Neubildungen unsicheren oder unbekanntes Verhaltens darstellen, ist bei einem Folgeaufenthalt von einer bereits vorhandenen exakteren Diagnose auszugehen.

Neben der Gesamtanzahl an Neuerkrankungen von 1146 Fällen für das Jahr 2007 ist die Darstellung der Verteilung nach Absolutzahlen und Altersklassen von besonderem Interesse. Die folgenden beiden Graphiken zeigen die Fallzahlen für Männer und Frauen getrennt.





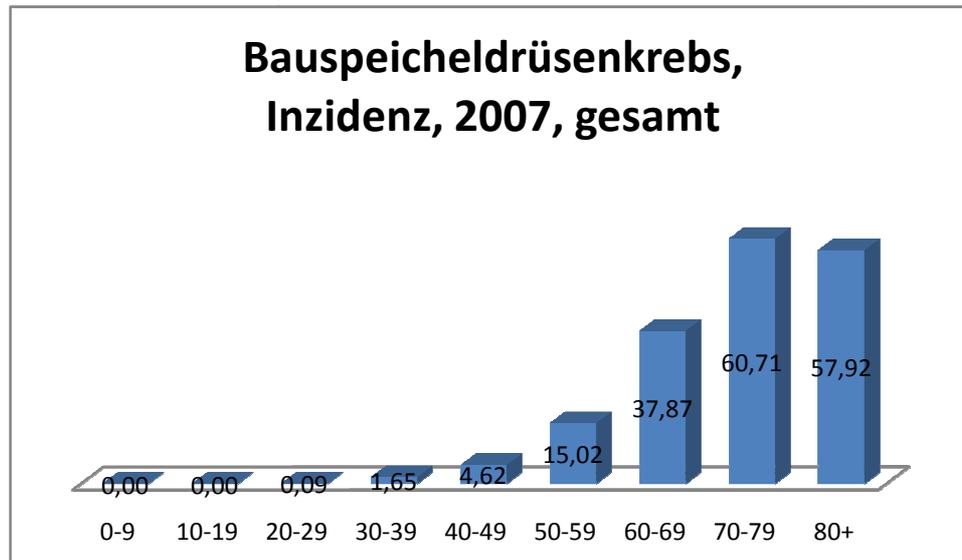
Aus diesen Graphiken erkennt man einen deutlichen Anstieg der Neuerkrankungsfälle ab einem Alter von 60-69 Jahren bei Frauen, bei Männern jedoch bereits in der Altersgruppe der ab 50 Jährigen. Eine reduzierte Fallzahl für die Gruppe der über 80 Jährigen bedeutet nicht gesichert eine geringere Inzidenz, jedoch speziell bei Männern lässt sich diese vermuten.

Zur Berechnung der Inzidenz werden die Bevölkerungsdaten der Statistik Austria für das Jahr 2007 zu Grunde gelegt. Man erhält die in der nachstehenden Tabelle angeführten Inzidenzen:

Altersgruppe	männlich	weiblich	gesamt
0-9	0,00	0,00	0,00
10-19	0,00	0,00	0,00
20-29	0,00	0,19	0,09
30-39	1,81	1,48	1,65
40-49	5,36	3,86	4,62
50-59	20,51	9,69	15,02
60-69	40,95	35,07	37,87
70-79	71,86	52,31	60,71
80+	62,27	56,11	57,92

Klar ersichtlich ist aus der tabellarischen Darstellung eine leicht erhöhte Inzidenz bei Männern im Vergleich zu Frauen für die Altersgruppen ab 40 Jahren, speziell für die über 70 Jährigen ist dieser Unterschied signifikant. Für die Altersgruppe der 70-79 Jährigen erreicht die Neuerkrankungsanzahl von 71,86 Personen pro 100.000 Männern dieser Altersgruppe und Jahr ihr Maximum.

Die folgende Darstellung zeigt die Ergebnisse bezüglich der geschlechterneutralen Inzidenzen in Form eines Säulenplots:



Abschließend kann aus den berechneten Daten ersehen werden, dass die Anzahl an Neuerkrankungen bei den unter 50 Jährigen minimal ist, es erkrankten in Summe nur 21 Personen neu im Jahr 2007. Darüber hinaus erkennt man, dass die aus der GAP-DRG Datenbank identifizierten Fälle (1146), gut mit den Daten der Statistik Austria (http://www.statistik.at/web_de/statistiken/gesundheit/krebserkrankungen/bauchspeicheldruese/021691.html), die 1.406 Neuerkrankungsfälle berichten, übereinstimmt. Da in der GAPDRG mehr neue Fälle identifiziert werden als von der Statistik Austria werden mögliche Ursachen für die Differenz gesucht. Einerseits ist es möglich, dass Fälle nicht an die Statistik Austria gemeldet werden. Andererseits ist es möglich, dass speziell für Bauspeicheldrüsenkrebs die Auswahlkriterien nicht exakt genug gewählt sind. Speziell der Diagnosecode D37.7 ist nicht sehr spezifisch.

8 Speiseröhre

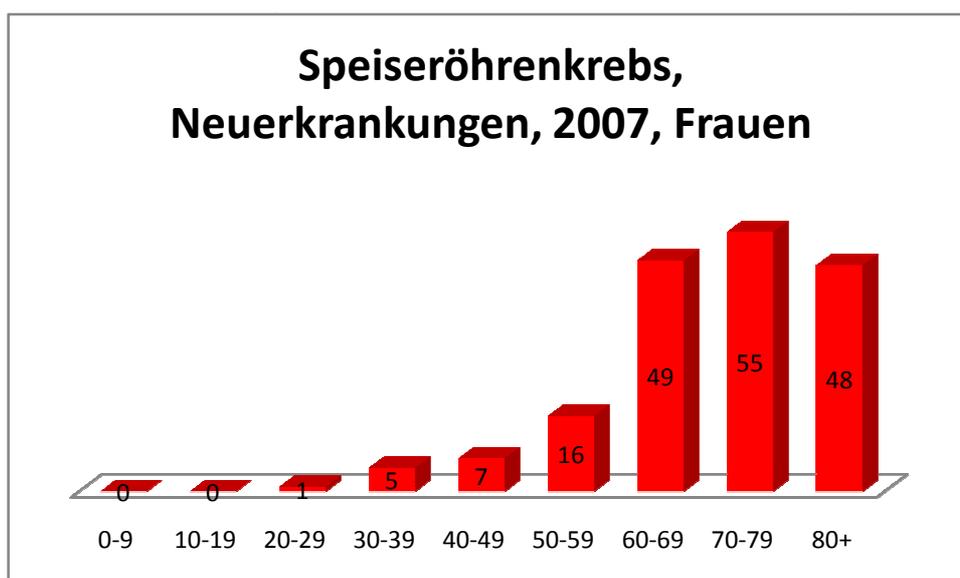
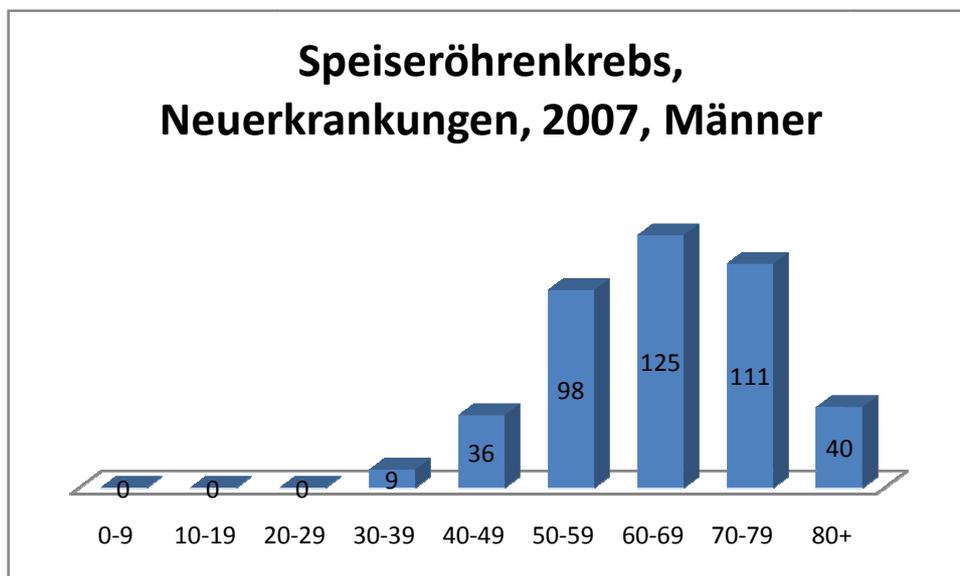
Basierend auf den in der Aufgabenstellung beschriebenen Fragestellungen werden Speiseröhrenkarzinome im Detail analysiert. Dazu werden folgende Diagnosen herangezogen:

C15.X Bösartige Neubildung des Ösophagus

D00.1 Ösophagus

D37.7 Sonstige Verdauungsorgane inkl.:Anus o.n.A., Canalis analis, Darm o.n.A., Ösophagus, Pankreas, Sphincter ani

Neben der Gesamtanzahl an Neuerkrankungen von 600 Fällen für das Jahr 2007 ist die Darstellung der Verteilung nach Absolutzahlen und Altersklassen von besonderem Interesse. Die folgenden beiden Graphiken zeigen die Fallzahlen für Männer und Frauen getrennt.



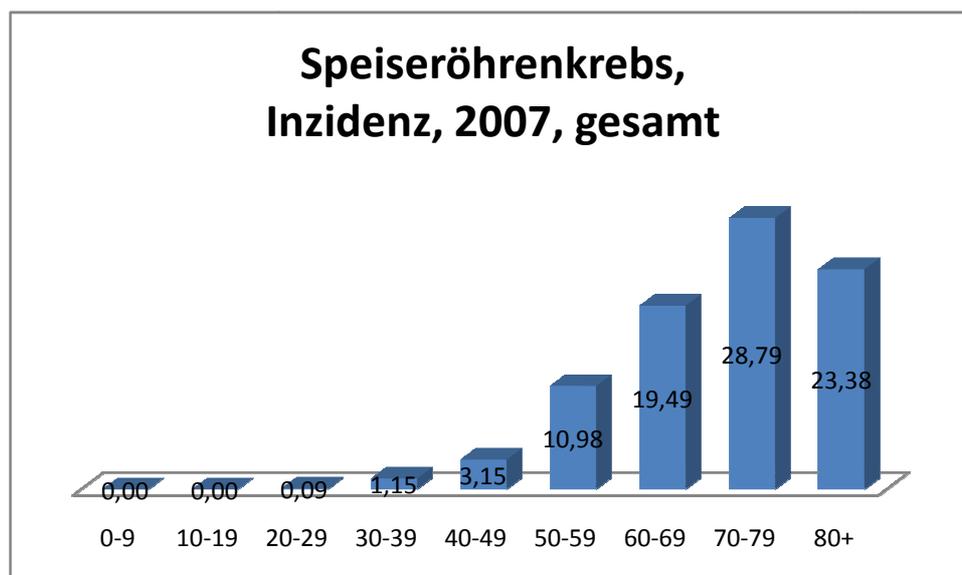
Aus diesen Graphiken erkennt man einen deutlichen Anstieg der Neuerkrankungsfälle ab einem Alter von 60-69 Jahren bei Frauen. Männer erkranken bereits ab einem Alter von 50 Jahren verstärkt. Generell ist die Krankheitslast in Österreich aber bedeutend geringer als bei anderen Krebsarten wie etwa Lungenkrebs oder Darmkrebs.

Zur Berechnung der Inzidenz werden die Bevölkerungsdaten der Statistik Austria für das Jahr 2007 zu Grunde gelegt. Man erhält die in der nachstehenden Tabelle angeführten Inzidenzen:

Altersgruppe	männlich	weiblich	gesamt
0-9	0,00	0,00	0,00
10-19	0,00	0,00	0,00
20-29	0,00	0,19	0,09
30-39	1,48	0,82	1,15
40-49	5,21	1,04	3,15
50-59	19,14	3,04	10,98
60-69	29,42	10,48	19,49
70-79	44,81	16,73	28,79
80+	36,10	18,08	23,38

Klar ersichtlich ist aus der tabellarischen Darstellung eine stark erhöhte Inzidenz bei Männern im Vergleich zu Frauen für die Altersgruppen ab 40 Jahren. Für die Altersgruppe der 70-79 Jährigen erreicht die Neuerkrankungsanzahl von 44,81 Personen pro 100.000 Männern dieser Altersgruppe und Jahr ihr Maximum.

Die folgende Darstellung zeigt die Ergebnisse bezüglich der geschlechterneutralen Inzidenzen in Form eines Säulenplots:



Abschließend kann aus den berechneten Daten ersehen werden, dass die Anzahl an Neuerkrankungen bei den unter 50 Jährigen minimal ist. Darüber hinaus erkennt man, dass die aus der GAP-DRG Datenbank identifizierten Fälle (600), ein deutliches Overreporting im



Vergleich zu den Daten der Statistik Austria (http://www.statistik.at/web_de/statistiken/gesundheit/krebserkrankungen/speiseroehre/021801.html), die 384 Neuerkrankungsfälle berichten, liefern. Diese zu hohe Fallzahl resultiert primär aus der Berücksichtigung der Diagnose D37.7 als Indikator. Für weiterführende Arbeiten ist daher speziell auf diese Diagnosen zu achten und bei Bedarf weitere Ausschlusskriterien (zum Beispiel auf Basis der Medizinischen Einzelleistungen) zu treffen.

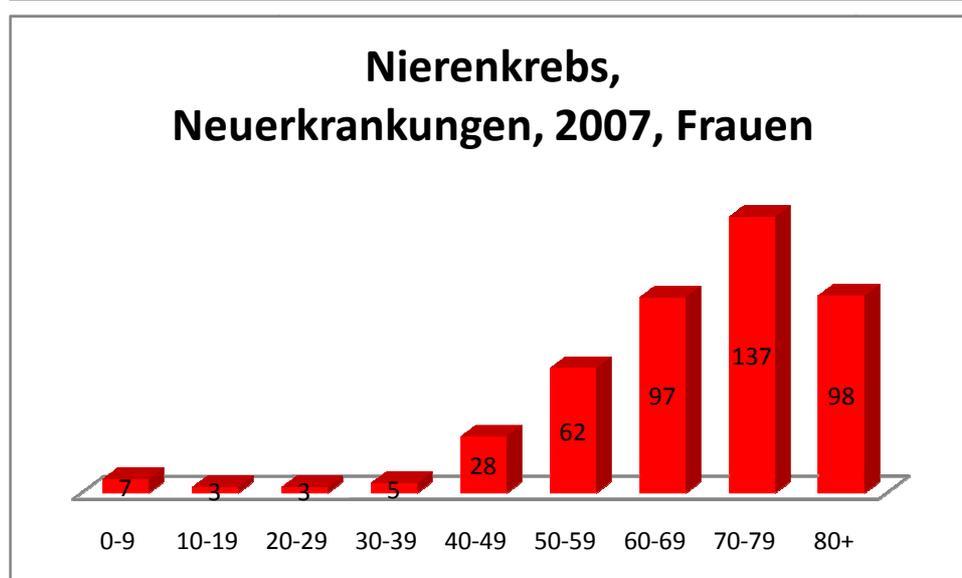
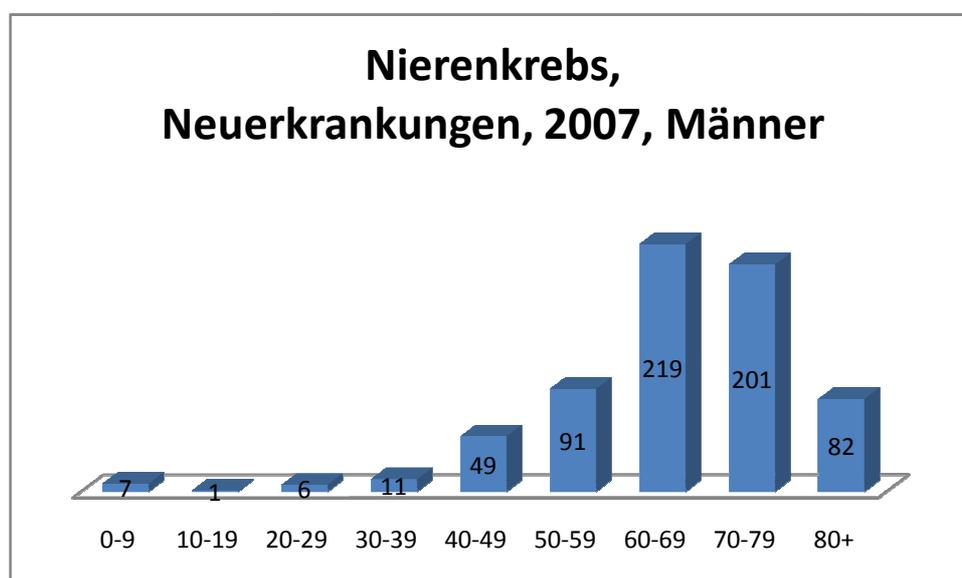
9 Niere

Basierend auf den in der Aufgabenstellung beschriebenen Fragestellungen werden Nierenkarzinome im Detail analysiert. Dazu werden folgende Diagnosen herangezogen:

C64.X Bösartige Neubildung der Niere, ausgenommen Nierenbecken

D41.0 Niere Exkl.: Nierenbecken (D41.1)

Neben der Gesamtanzahl an Neuerkrankungen von 1107 Fällen für das Jahr 2007 ist die Darstellung der Verteilung nach Absolutzahlen und Altersklassen von besonderem Interesse. Die folgenden beiden Graphiken zeigen die Fallzahlen für Männer und Frauen getrennt.



Aus diesen Graphiken erkennt man einen deutlichen Anstieg der Neuerkrankungsfälle ab einem Alter von 50-59 Jahren. Speziell die Darstellung der Neuerkrankungen bei Männern lässt ein

Abklingen mit höherem Alter vermuten. Die Überprüfung dieser Annahme geschieht abermals durch Berechnung der Inzidenzen.

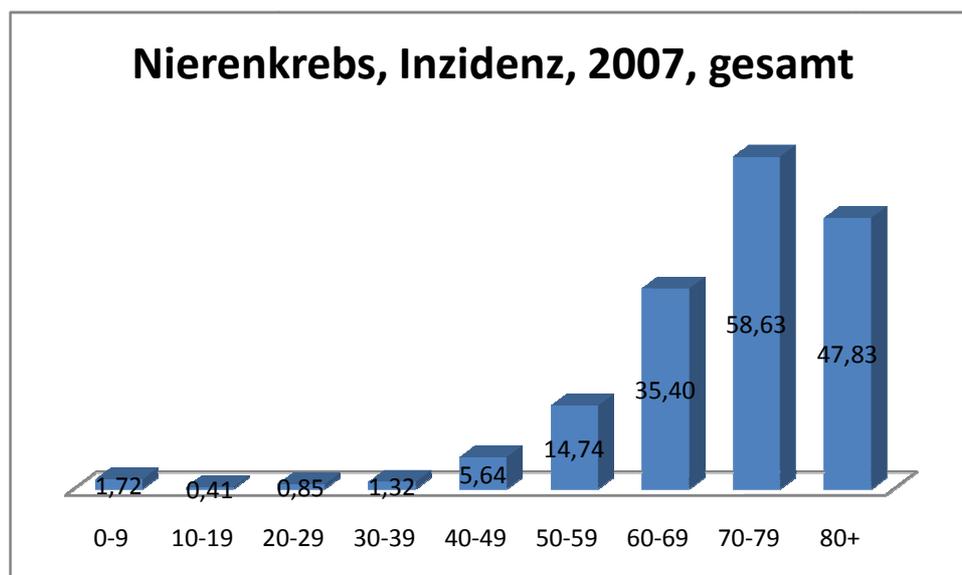
Zur Berechnung der Inzidenz werden die Bevölkerungsdaten der Statistik Austria für das Jahr 2007 zu Grunde gelegt. Man erhält die in der nachstehenden Tabelle angeführten Inzidenzen:

Altersgruppe	männlich	weiblich	gesamt
0-9	1,68	1,77	1,72
10-19	0,20	0,63	0,41
20-29	1,13	0,58	0,85
30-39	1,81	0,82	1,32
40-49	7,10	4,15	5,64
50-59	17,78	11,78	14,74
60-69	51,54	20,74	35,40
70-79	81,15	41,67	58,63
80+	74,01	36,90	47,83

Klar ersichtlich ist aus der tabellarischen Darstellung eine stark erhöhte Inzidenz bei Männern im Vergleich zu Frauen für die Altersgruppen ab 40 Jahren. Für die Altersgruppe der 70-79 Jährigen erreicht die Neuerkrankungszahl von 81,15 Personen pro 100.000 Männern dieser Altersgruppe und Jahr ihr Maximum. Das bedeutet, dass annähernd ein Promille der über 70 jährigen Männer jährlich neu an Nierenkrebs erkrankt.

Die Erkrankungswahrscheinlichkeit für Männer über 40 liegt etwa doppelt so hoch als die von gleichaltrigen Frauen.

Die folgende Darstellung zeigt die Ergebnisse bezüglich der geschlechterneutralen Inzidenzen in Form eines Säulenplots:





Abschließend kann aus den berechneten Daten ersehen werden, dass die Anzahl an Neuerkrankungen bei den unter 40 Jährigen minimal ist. Darüber hinaus erkennt man, dass die aus der GAP-DRG Datenbank identifizierten Fälle (1107), sehr genau mit den Daten der Statistik Austria

(http://www.statistik.at/web_de/statistiken/gesundheit/krebserkrankungen/niere/021781.html), die 1209 Neuerkrankungsfälle berichten, übereinstimmt.

Dieses sehr gute Ergebnis lässt sich nicht zuletzt auf die klar definierten Diagnosecodes in ICD10 für Nierenkrebs zurückführen.

10 Prostata

Basierend auf den in der Aufgabenstellung beschriebenen Fragestellungen werden Prostatakarzinome im Detail analysiert. Dazu werden folgende Diagnosen herangezogen:

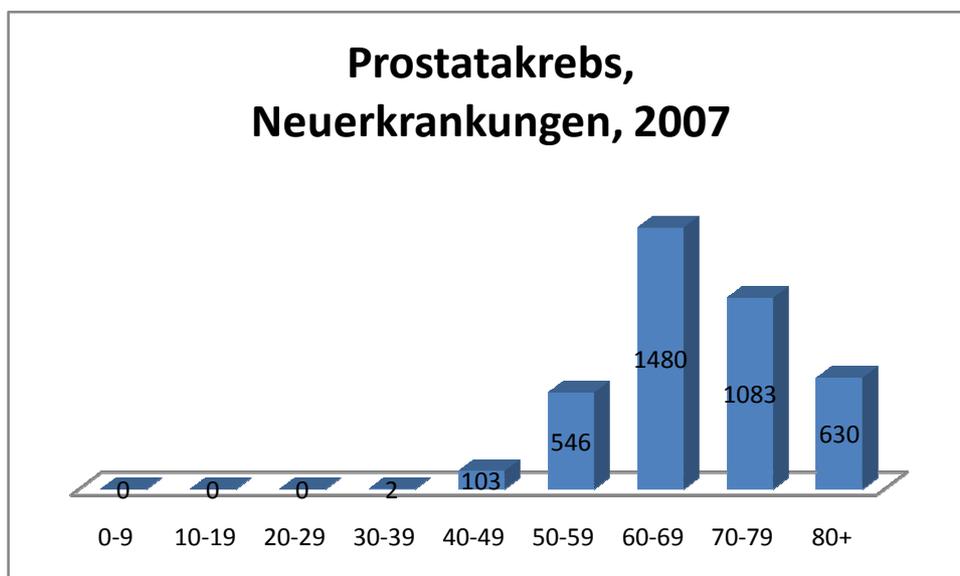
C61.X Bösartige Neubildung der Prostata

D07.5 Prostata; inkl.: Hochgradige intraepitheliale Neoplasie der Prostata [high-grade PIN]

Intraepitheliale Neoplasie der Prostata [PIN] II. und III. Grades

D40.0 Prostata

Da Prostataerkrankungen nur Männer betreffen können, werden im Folgenden alle Aussagen nur für Männer getroffen. Neben der Gesamtanzahl an Neuerkrankungen von 3844 Fällen für das Jahr 2007 ist die Darstellung der Verteilung nach Absolutzahlen und Altersklassen von besonderem Interesse. Die folgende Graphik zeigt die Fallzahlen für die unterschiedlichen Altersklassen.



Aus diesen Graphiken erkennt man einen Anstieg der Neuerkrankungsfälle ab einem Alter von 40-49 Jahren mit einer Steigerung um den Faktor 51 im Vergleich zur Altersgruppe der 30-39 Jährigen.

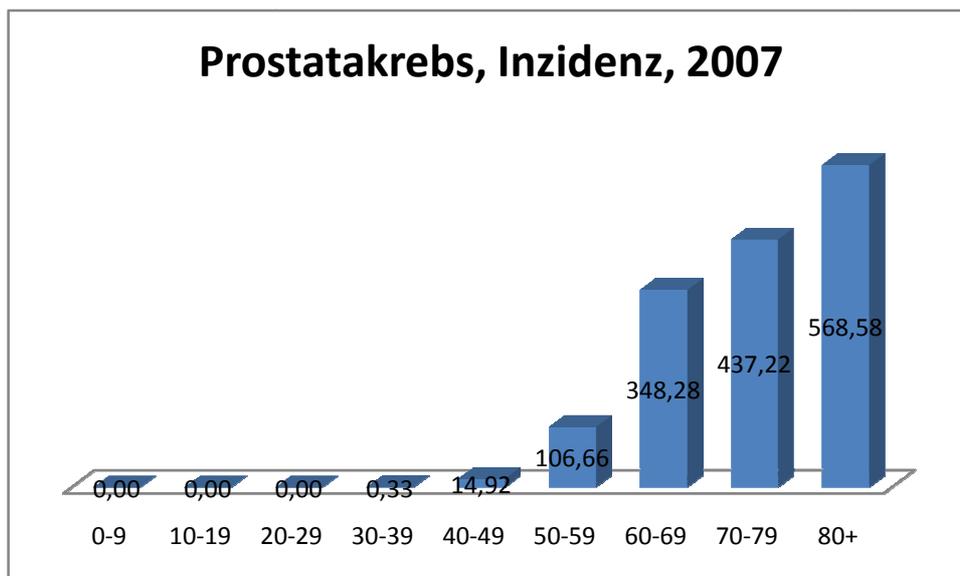
Ein absolutes Maximum wird für Männer im Alter zwischen 60 und 69 Jahren erreicht. Die Altersgruppe der über 80 jährigen Männer hat eine geringere Fallzahl. Speziell bei Prostatakarzinomen ist jedoch zu beachten, dass nicht alle Neuerkrankungen zwingend zu einer stationären Behandlung führen. Außerdem liefert eine geringere Absolutzahl noch keine gesicherte Aussage zur Inzidenz.

Zur Berechnung der Inzidenz werden die Bevölkerungsdaten der Statistik Austria für das Jahr 2007 zu Grunde gelegt. Man erhält die in der nachstehenden Tabelle angeführten Inzidenzen:

Altersgruppe	männlich
0-9	0,00
10-19	0,00
20-29	0,00
30-39	0,33
40-49	14,92
50-59	106,66
60-69	348,28
70-79	437,22
80+	568,58

Das Maximum ist bei den über 80 Jährigen mit 568,58 Personen pro 100.000 Männern dieser Altersgruppe und Jahr erreicht.

Die folgende Darstellung zeigt die Ergebnisse bezüglich der Inzidenzen in Form eines Säulenplots:



Abschließend kann aus den berechneten Daten ersehen werden, dass die Anzahl an Neuerkrankungen bei den unter 50 Jährigen gering ist. Darüber hinaus erkennt man, dass die aus der GAP-DRG Datenbank identifizierten Fälle (3844), weniger Erkrankte erfasst, als aus den Daten der Statistik Austria (http://www.statistik.at/web_de/statistiken/gesundheit/krebserkrankungen/prostata/021791.html), die 5108 Neuerkrankungsfälle berichten, ersichtlich sind.

11 Hoden

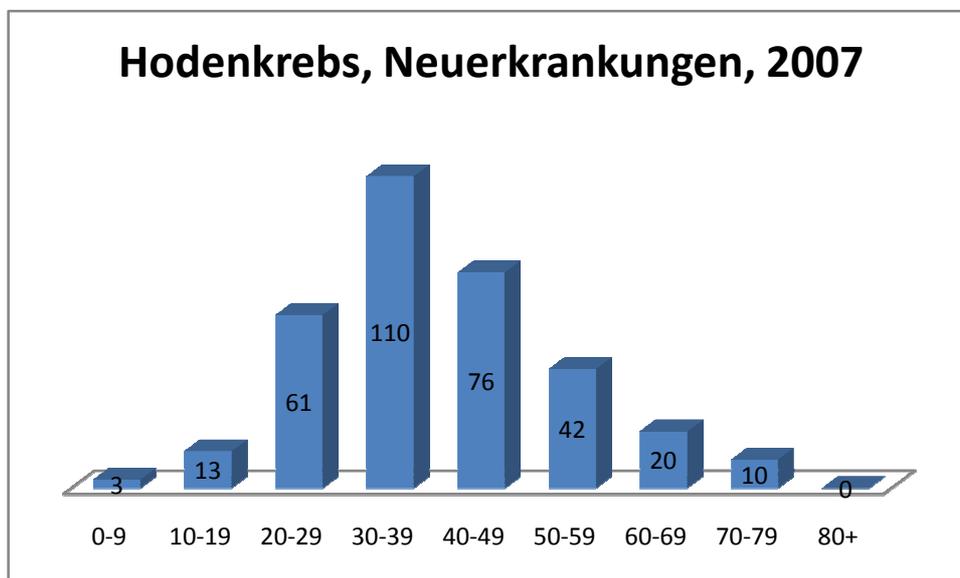
Basierend auf den in der Aufgabenstellung beschriebenen Fragestellungen werden Hodenkarzinome im Detail analysiert. Dazu werden folgende Diagnosen herangezogen:

C62.X Bösartige Neubildung des Hodens

D07.6 Sonstige und nicht näher bezeichnete männliche Genitalorgane

D40.1 Hoden

Neben der Gesamtanzahl an Neuerkrankungen von 335 Fällen für das Jahr 2007 ist die Darstellung der Verteilung nach Absolutzahlen und Altersklassen von besonderem Interesse. Die folgende Graphik zeigt die Fallzahlen für die unterschiedlichen Altersklassen.



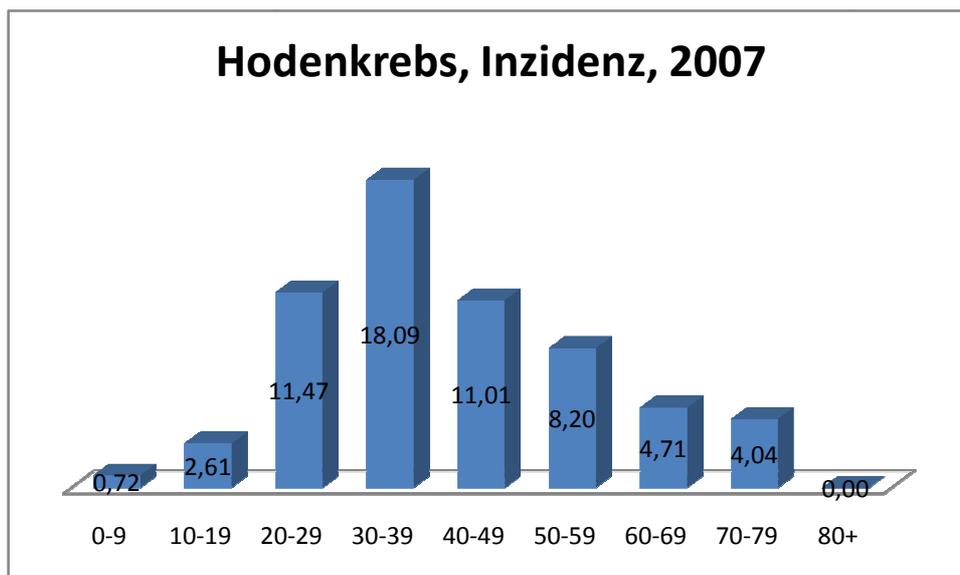
Aus diesen Graphiken erkennt man einen deutlichen Anstieg der Neuerkrankungsfälle ab einem Alter von 20-29 Jahren. Eine reduzierte Fallzahl ist ab der Gruppe der über 50 Jährigen erkennbar. Eine Besonderheit des Hodenkrebses ist somit das erhöhte Krankheitsrisiko in jungen Bevölkerungsschichten. Die Betrachtung der Daten der Statistik Austria liefert für das Jahr 2007 eine Gesamtzahl von 26 Todesfällen.

Zur Berechnung der Inzidenz werden die Bevölkerungsdaten der Statistik Austria für das Jahr 2007 zu Grunde gelegt. Man erhält die in der nachstehenden Tabelle angeführten Inzidenzen:

Altersgruppe	männlich
0-9	0,72
10-19	2,61
20-29	11,47
30-39	18,09
40-49	11,01
50-59	8,20
60-69	4,71
70-79	4,04
80+	0,00

Das Maximum ist bei den 30-39 Jährigen mit 18,09 Personen pro 100.000 Männern dieser Altersgruppe und Jahr erreicht, danach erkennt man eine absteigende Inzidenz. Die Inzidenz scheint aus statistischen Gesichtspunkten bestmöglich durch eine nicht symmetrische Dreiecksverteilung dargestellt werden zu können.

Die folgende Darstellung zeigt die Ergebnisse bezüglich der Inzidenzen in Form eines Säulenplots:



Signifikant für den Hodenkrebs ist unter anderem, dass die Inzidenzen für über 60 Jährige sehr gering sind und keine Fälle bei Personen, älter als 80 Jahre, diagnostiziert wurden.

Darüber hinaus erkennt man, dass die aus der GAP-DRG Datenbank identifizierten Fälle (335), sehr genau mit den Daten der Statistik Austria (http://www.statistik.at/web_de/statistiken/gesundheit/krebserkrankungen/hoden/021741.html), die 345 Neuerkrankungsfälle berichten, übereinstimmt.

12 Gebärmutterhalskrebs (Zervix)

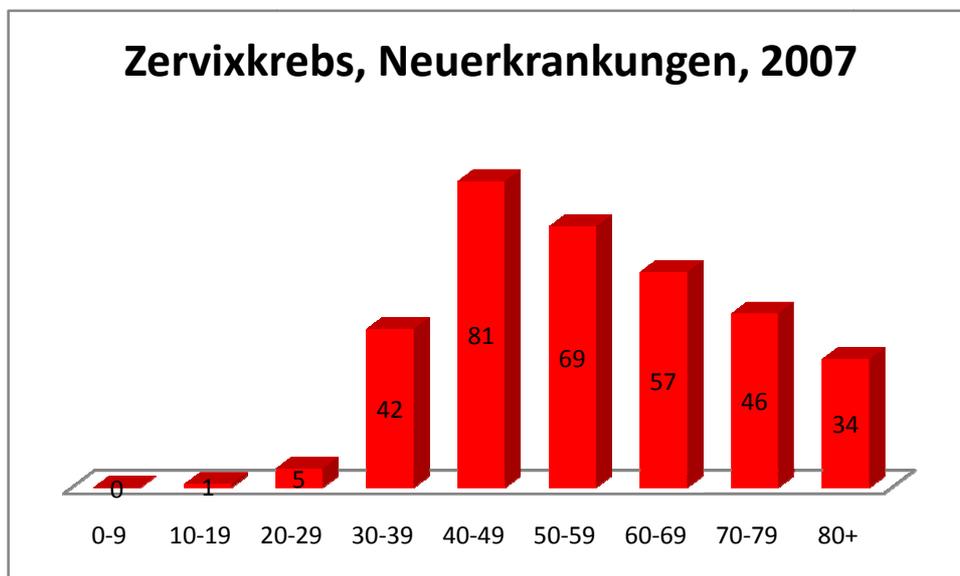
Basierend auf den in der Aufgabenstellung beschriebenen Fragestellungen werden Zervixkarzinome im Detail analysiert. Dazu werden folgende Diagnosen herangezogen:

C53.X Bösartige Neubildung der Cervix uteri

D06.X Carcinoma in situ der Cervix uteri Inkl.: Zervikale intraepitheliale Neoplasie [CIN] III. Grades, mit oder ohne Angabe einer hochgradigen Dysplasie

D39.9 Weibliches Genitalorgan, nicht näher bezeichnet

Neben der Gesamtanzahl an Neuerkrankungen von 335 Fällen für das Jahr 2007 ist die Darstellung der Verteilung nach Absolutzahlen und Altersklassen von besonderem Interesse. Die folgende Graphik zeigt die Fallzahlen für die unterschiedlichen Altersklassen.



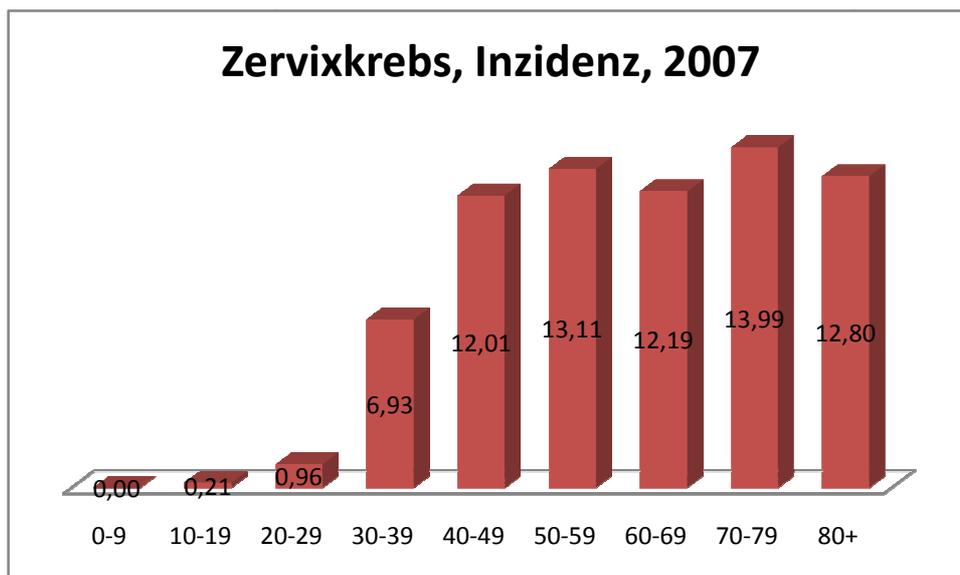
Aus diesen Graphiken erkennt man einen deutlichen Anstieg der Neuerkrankungsfälle ab einem Alter von 30-39 Jahren. Ein gleichmäßiger Abfall der Fallzahlen pro Altersgruppe ist ab der Gruppe der über 50 Jährigen erkennbar. Interessant ist hierbei, dass der Abfall annähernd linear um 11,5 Fälle auf die jeweils folgende Altersgruppe ist.

Zur Berechnung der Inzidenz werden die Bevölkerungsdaten der Statistik Austria für das Jahr 2007 zu Grunde gelegt. Man erhält die in der nachstehenden Tabelle angeführten Inzidenzen:

Altersgruppe	weiblich
0-9	0,00
10-19	0,21
20-29	0,96
30-39	6,93
40-49	12,01
50-59	13,11
60-69	12,19
70-79	13,99
80+	12,80

Das Maximum wird bei den 70-79 Jährigen mit 13,99 neu erkrankten Personen pro 100.000 Frauen dieser Altersgruppe und Jahr erreicht. Anders als bei den Absolutzahlen ist die Inzidenz für Personen im Alter ab 40 Jahren annähernd stabil.

Die folgende Darstellung zeigt die Ergebnisse bezüglich der Inzidenzen in Form eines Säulenplots:



Abschließend kann aus den berechneten Daten ersehen werden, dass die Anzahl an Neuerkrankungen bei den unter 30 Jährigen minimal ist. Darüber hinaus erkennt man, dass die aus der GAP-DRG Datenbank identifizierten Fälle (335), sehr genau mit den Daten der Statistik Austria

(http://www.statistik.at/web_de/statistiken/gesundheit/krebserkrankungen/gebaermutterhals/021716.html), die 362 Neuerkrankungsfälle berichten, übereinstimmt.

13 Eierstock

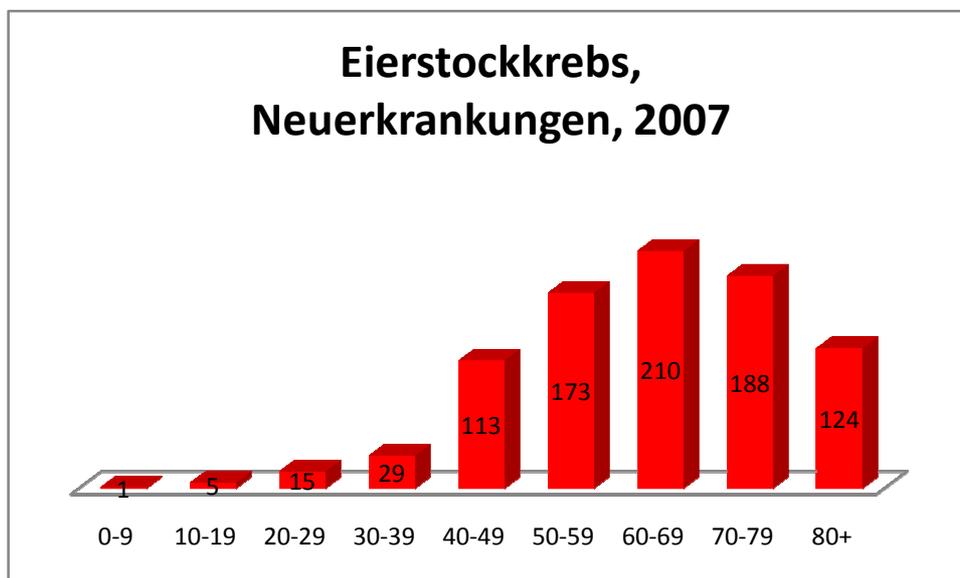
Basierend auf den in der Aufgabenstellung beschriebenen Fragestellungen werden Eierstockkarzinome im Detail analysiert. Dazu werden folgende Diagnosen herangezogen:

C56.X Bösartige Neubildung des Ovars

D07.3 Sonstige und nicht näher bezeichnete weibliche Genitalorgane

D39.1 Ovar

Neben der Gesamtanzahl an Neuerkrankungen von 858 Fällen für das Jahr 2007 ist die Darstellung der Verteilung nach Absolutzahlen und Altersklassen von besonderem Interesse. Die folgende Graphik zeigt die Fallzahlen für die unterschiedlichen Altersklassen.



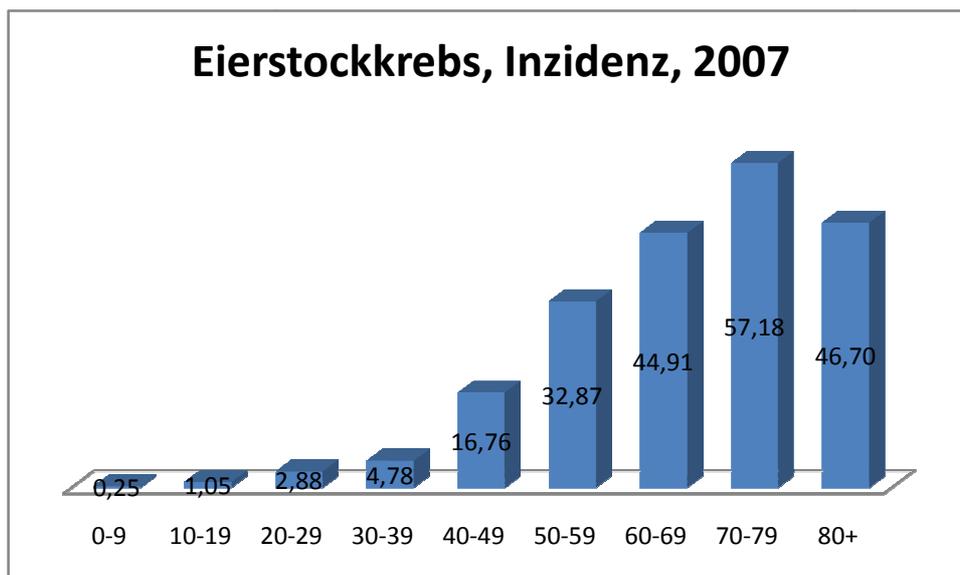
Aus diesen Graphiken erkennt man einen deutlichen Anstieg der Neuerkrankungsfälle ab einem Alter von 40-49 Jahren, dieser beträgt den Faktor 4 im Vergleich zur vorhergehenden Altersgruppe. Eine leicht reduzierte Fallzahl ist ab der Gruppe der über 70 Jährigen erkennbar, was nicht gesichert eine geringere Inzidenz bedeutet. Diese muss noch gesondert analysiert werden.

Zur Berechnung der Inzidenz werden die Bevölkerungsdaten der Statistik Austria für das Jahr 2007 zu Grunde gelegt. Man erhält die in der nachstehenden Tabelle angeführten Inzidenzen:

Altersgruppe	weiblich	gesamt
0-9	0,25	0,12
10-19	1,05	0,51
20-29	2,88	1,42
30-39	4,78	2,39
40-49	16,76	8,28
50-59	32,87	16,66
60-69	44,91	23,53
70-79	57,18	32,61
80+	46,70	32,95

Das Maximum ist bei den 70-79 Jährigen mit 57,18 Personen pro 100.000 Frauen dieser Altersgruppe und Jahr erreicht. Das bedeutet, dass ca. jede 2000ste Frau im Alter zwischen 70 und 79 Jahren jährlich neu an Eierstockkrebs erkrankt.

Die folgende Darstellung zeigt die Ergebnisse bezüglich der Inzidenzen in Form eines Säulenplots:



Man sieht deutlich, dass die Inzidenz bis zur Altersgruppe der unter 80 Jährigen steigt, wohingegen die Absolutzahl an Neuerkrankungen bereits früher überschritten wird.

Abschließend kann aus den berechneten Daten ersehen werden, dass die Anzahl an Neuerkrankungen bei den unter 40 Jährigen minimal ist. Darüber hinaus erkennt man, dass die aus der GAP-DRG Datenbank identifizierten Fälle (858) mit den Daten der Statistik Austria (http://www.statistik.at/web_de/statistiken/gesundheit/krebserkrankungen/eierstock/021706.html), die 673 Neuerkrankungsfälle berichten, nur bedingt übereinstimmt. Dies ist auf die nicht gesicherten Fälle, die durch D39.1 kodiert werden, zurückzuführen.

14 Resümee

Die hier beschriebene Vorgangsweise in Kombination mit der Überprüfung und Analyse der Datensätze im Vergleich zu Statistik Austria Ergebnissen liefert eine breite Anwendungsbasis. Denn neben dem reinen Erhalt von Fallzahlen kann durch die Verknüpfung von weiteren Datenquellen auf Personenebene eine Vielzahl von Besonderheiten der Behandlungswege, Leistungen im niedergelassenen Bereich als auch Medikamentenabgabe detaillierter untersucht werden.

Die folgende Tabelle fasst für das Kalenderjahr 2007 die Gesamtfallzahlen der ausgewählten Krebserkrankungen analysiert aus den GAP-DRG Daten zusammen und stellt sie den Daten der Statistik Austria gegenüber:

Tabelle 14-1: Tabelle der Absolutfallzahlen für Krebsneuerkrankungen analysiert aus der GAP-DRG Datenbank im Vergleich mit Daten der Statistik Austria; die 4.Spalte stellt die Abweichung von den Daten der Statistik Austria dar.

	Anzahl GAP-DRG	Anzahl Statistik Austria	prozentuelle Abweichung
Leberkrebs	906	892	1,57
Lungenkrebs	3877	4087	-5,14
Brustkrebs	4882	4833	1,01
Darmkrebs	3719	4610	-19,33
Bauchspeicheldrüsenkrebs	1146	1406	-18,49
Speiseröhrenkrebs	600	384	56,25
Nierenkrebs	1107	1209	-8,44
Prostatakrebs	3844	5108	-24,75
Hodenkrebs	335	345	-2,90
Gebärmutterhalskrebs	335	362	-7,46
Eierstockkrebs	858	673	27,49

Klar erkennbar ist in dieser Darstellung die hohe Übereinstimmung der Fallzahlen für Krebsarten mit klar gegebenen Diagnosen (ohne nicht spezifischen Diagnosen aus der Gruppe D, wie zum Beispiel D37.1).

Dieses Potential in Verbindung mit dem Detailwissen der Statistik Austria zu Tumorstadien und deren Verteilung kann als Basis für die Analyse von Szenarien dienen. Besonders hervorgehoben muss in diesem Zusammenhang die Fragestellung der Wirksamkeit von Screeningmaßnahmen werden. Durch die Nutzung von GAP-DRG Daten können der gesamte Behandlungsweg und die Überlebenswahrscheinlichkeiten in höherem Detailgrad abgeschätzt und dadurch eine sozio-ökonomische Gesamtbewertung auf Basis von Realdaten durchgeführt werden.

Des Weiteren ist auf die vorhandene hohe Rationalitätsauflösung der Datensätze in GAP-DRG hinzuweisen. Diese in Verbindung mit der Möglichkeit der Betrachtung eines gesamten Behandlungsweges (intramuraler und extramuraler Bereich) und den abgegebenen Medikamenten stellt einen weiteren Schritt in der Behandlungsstrategiebewertung und der Ressourcenplanung dar. Denn neben klassischen Überlebenswahrscheinlichkeiten können so durch Anwendung von dynamischen Modellen Behandlungs- und Vorsorgestrategien gegenüber gestellt werden und für jeweils vorgegebene regionalen Strukturen getestet werden.

15 Qualitätssicherung – Vergleich mit Krebsregister

Basierend auf den nun ermittelten Neuerkrankungen basierend auf den MBDS – Daten in Kombination mit Filtern aus den Heilmitteldaten und den Daten des niedergelassenen Bereichs kann eine Qualitätssicherung durch Vergleich mit Daten aus dem Krebsregister der Statistik Austria durchgeführt werden. Um eine Aussage über die Datenqualität und die Vergleichbarkeit mit den Daten der Statistik Austria zu erhalten wird im folgenden Unterkapitel eine Zusammenfassung und Bewertung durchgeführt.

15.1 Kurzzusammenfassung der relevanten Passagen aus „Standard-Dokumentation Metainformationen (Definitionen, Erläuterungen, Methoden, Qualität) zur Krebsstatistik (Krebsregister)“

Aus Datenschutz- und Qualitätsgründen können z.B. für einzelne Gemeinden keine Auswertungen angeboten werden.

Zur Qualität der Datenquellen:

Grundlage für die Krebsmeldeblätter bilden die Befunde der Krankengeschichte. Die Qualität dieser Aufzeichnung ist sehr gut, da sie im Rahmen der Behandlung eines Patienten gemacht werden. Daraus ergibt sich allerdings auch, dass gewisse Merkmale, die für die Behandlung nicht von unmittelbarem Interesse sind, häufig vernachlässigt werden.

Zur Abdeckung:

Personen mit Wohnsitz in Österreich, die sich jedoch ausschließlich einer Behandlung im Ausland unterziehen, können nicht erfasst werden, da die Dokumentationspflicht nur für Krankenanstalten in Österreich rechtsverbindlich ist.

Alle Krebserkrankungen von Personen mit Wohnsitz in Österreich, die in österreichischen Krankenanstalten diagnostiziert werden, sind per Gesetz meldepflichtig. Bezüglich der Vollzähligkeit ist die Einschränkung der Meldepflicht auf Krankenanstalten etwas problematisch. Da Ambulanzen außerhalb der Krankenanstalten und der niedergelassene Bereich nicht meldepflichtig sind, kommt es möglicherweise bei bestimmten Lokalisationen zu einer Untererfassung. Bei diesen Fällen handelt es sich laut Expertenauskunft in erster Linie um Melanome, Prostatakrebs sowie Gebärmutterhalskrebs in einem frühen Stadium und um ältere Lymphompatienten.

Zur Kontrolle:

Zur Kontrolle der Vollzähligkeit wird die DCO-Rate (Death-Certificate-Only-Rate) herangezogen. Dabei handelt es sich um den Prozentsatz der Personen, die laut Totenschein an Krebs verstorben sind, jedoch nicht im Rahmen der gesetzlich vorgeschriebenen Inzidenzmeldung an das Krebsregister der Statistik Austria gemeldet wurden, an allen Krebsmeldungen des jeweiligen Berichtsjahres.

Diese muss um die Datenqualitätsanforderungen der IARC (International Agency for Research on Cancer) Länder zu erreichen unter 10% liegen. Dies gilt für Österreich erst ab dem Jahr 1997.

15.2 Resumee

Die folgenden Aussagen stellen eine diskutierbare Einschätzung der Datensammlung des Krebsregisters dar:

- Der aktuelle Zeitversatz zwischen Meldungen und vollständiger Bearbeitung beträgt in Österreich aktuell ca. zwei Jahre. International erlaubt sind 5 Jahre. Daher kann davon ausgegangen werden, dass die Datensätze für 2006 und 2007 bereits vollständig



bearbeitet sind. Daten zu einem internationalen Vergleich sind allerdings im derzeitigen Stand nicht gesichert verfügbar.

- Ein Problem, das ein weiteres Underreporting durch das Krebsregister der Statistik Austria nahe legt, ist die Tatsache dass nur Krankenhausdaten verwendet werden. Dies gilt sowohl bei den Diagnosen als auch bei den betrachteten Sterbefällen! Vor allem bei diesen ist mit einer hohen Anzahl an nicht erfassten Fällen, speziell im ländlichen Raum, zu rechnen.

Eine weitere Alternative zur Verbesserung der Qualität könnte eine Kombination des Wissens aus dem nationalen Melderegister mit auf Bundeslandebene vorhandenen Registern darstellen. In den Bundesländern Vorarlberg, Tirol, Salzburg und Kärnten gibt es landesweite Krebsregister.

B. Detailbetrachtung der ATC-ICD Ergebnisse speziell für Krebsdiagnosen

1 Aufgabenstellung

Um zu erhaltenen Krankenanstaltenaufenthalte medikamentöse Vorbehandlungen zu erkennen und dadurch Fälle, die keine Neuerkrankung darstellen, auszuschließen, ist es notwendig, detaillierter die Ergebnisse des ATC-ICD Projektes in der vorliegenden Version aus GAP-DRG zu untersuchen.

Dazu sind folgende ICD9 – Gruppen von besonderem Interesse:

140-149	Bösartige Neubildungen der Lippe, der Mundhöhle und des
150-159	Bösartige Neubildungen der Verdauungsorgane und des Bauchfells
160-165	Neubildungen der Atmungs- und intrathorakalen Organe
170-175	Bösartige Neubildungen der Knochen, des Bindegewebes, der Haut und der Brustdrüse
179-189	Bösartige Neubildungen der Harn- und Geschlechtsorgane
190-199	Bösartige Neubildungen sonstiger und n.n.bez. Sitze
200-208	Bösartige Neubildungen des lymphatischen und hämatopoetischen Gewebes
230-234	Carcinoma in situ
239	Neubildungen unbekanntes Charakters

Untersucht soll zunächst werden, welche Medikamente bzw. ATC-Codes abgegeben wurden, wenn aus dem ATC-ICD Projekt eine einschlägige Diagnose errechnet wurde.

2 Vorgangsweise bei den Analysen

Zunächst wird dieses Vorgehen auf folgende Gruppen eingeschränkt um exemplarische Ergebnisse zu erhalten:

150-159, 160-165, 170-175, 230-234

Aus den im ATC-ICD Projekt enthaltenen Daten werden mit den Rezeptdaten sowohl die Personen IDs als auch das enthaltene Datum für die jeweils im Fokus befindlichen ICD9 – Gruppen gematcht. Zu den so erhaltenen Medikamentenabgaben werden sowohl die Pharmanummern als auch der Name und der ATC-Code gespeichert. Anschließend wird nach den einzelnen ATC-Codes aufsummiert.

In SQL wurde diese Aufgabenstellung (am Beispiel ICD9 Gruppe 170-175) wie folgt gelöst:

```
/*  
  Ausgeben der Medikamente mit ATC-Code und Name, die von  
  Personen am gleichen Tag per Rezept eingeloest wurden,  
  an denen sie laut atcid ein Rezept eingeloest haben,  
  das auf die icd-9-gruppe 170-175 matcht.  
*/
```



```
create table p_einzinger.gr170175 as
select atc, pharnr, name, Anzahl
from
heilmittel_hvb natural join
  (select pharnr, count(*) as Anzahl
   from rezeptdaten join
     (select pers_id, datum
      from atcicd
       where icd9_gruppe = '170 - 175) as person_datum
   on id_person_p=pers_id and eindat=datum
   group by pharnr) as pharmanummern
order by atc asc;

/*
  Ausgeben der Tabelle, geordnet nach Anzahl.
*/
select *
from p_einzinger.gr170175
order by anzahl desc;
```

3 Analyse und Interpretation für die ICD9 Gruppe 170-175

Mit einer zugeordneten ICD9 Gruppe 170-175 wurden im Betrachtungszeitraum 2006-2007 im ATC-ICD Projekt 4.577.357 Rezept mit einschlägiger Diagnose erkannt. Durch die Verknüpfung mit der Rezepte – Datenbank erhält man zu 137.910 keinen ATC-Code. Dies entspricht ca. 3% der Gesamtdatensätze.

Bei den verschriebenen und eingelösten Rezepten sind folgende 10 Pharnummern am häufigsten enthalten:

ATC-Code	Pharnr	Name	Anzahl	Prozent
L02BG03	1302720	Arimidex Filmtabl.	134.474	2,94
	9999927	MAGISTRALE ZUBEREITUNGEN	91.469	2,00
L02BA01	967417	Nolvadex 20 mg Filmtabl.	78.891	1,72
A02BC02	1292631	Pantoloc 40 mg Filmtabl.	74.029	1,62
L03AX	967185	Broncho-Vaxom 7 mg Kaps. f. Erw.	64.374	1,41
L01BA01	2443061	Ebetrexat 10 mg Tabl.	55.792	1,22
L02BG04	1309225	Femara 2,5 mg Filmtabl.	53.795	1,18
L02BB03	1302714	Casodex 50 mg Filmtabl.	50.876	1,11
B03BB01	713987	Folsan 5 mg Tabl.	44.868	0,98
A02BC02	1328464	Pantoloc 20 mg Filmtabl.	44.385	0,97

Diese 10 häufigsten Pharnummern stellen in Summe 15,14% aller abgegebenen Rezepte denen die ICD9 – Gruppe 170-175 zugeordnet wurde. Insgesamt wurden 8691 unterschiedliche Pharnummern erfasst, von denen 3331 weniger als 10-mal abgegeben wurden.

- Es ergibt sich die Notwendigkeit für weitere Analysen nicht den vollständigen ATC-Code zu betrachten, sondern abzuschneiden. Daher wurden nur die ersten vier Stellen des ATC-Codes berücksichtigt und erneut analysiert.

Diese Analyse enthält keine Pharnummern und keine Medikamentennamen mehr, da diese nun in ATC-Gruppen auf vier Stellen zusammengefasst sind. Die 10 größten Gruppen ergeben sich wie folgt:

ATC-Code	Anzahl	Prozent
L02B	428059	9,35
A02B	337101	7,36
L03A	191104	4,17
M01A	160802	3,51
L01B	155110	3,39
N06A	148779	3,25
J01D	135686	2,96
C07A	122258	2,67
N02A	105899	2,31
C10A	100424	2,19

Die so erfassten Rezepte stellen 41,16% aller eingelösten Medikamente dar.

Die Klassifizierung ob ein auf diese Weise erstellter Filter angewandt werden darf/soll ist die nächste zu beantwortende Frage.

4 Analyse und Interpretation für die ICD9 Gruppe 150-159

Mit einer zugeordneten ICD9 Gruppe 150-159 wurden im Betrachtungszeitraum 2006-2007 im ATC-ICD Projekt 8.024.971 Rezept mit einschlägiger Diagnose erkannt. Durch die Verknüpfung mit der Rezepte – Datenbank erhält man zu 277.612 keinen ATC-Code (zum Beispiel, da es sich um MAGISTRALE ZUBEREITUNGEN oder MELOLIN WUNDKOMPR ST 10X10CM, ... handelt). Dies entspricht ca. 3,5% der Gesamtdatensätze.

Bei den verschriebenen und eingelösten Rezepten sind folgende 10 Pharnummern am häufigsten enthalten:

ATC-Code	Pharnr	Name	Anzahl	Prozent
A09AA02	774546	Kreon 10.000 E Kaps.	198966	2,48
A02BC02	1292631	Pantoloc 40 mg Filmtabl.	171254	2,13
A09AA02	41269	Pankreoflat Drag.	161460	2,01
	9999927	MAGISTRALE ZUBEREITUNGEN	132601	1,65
A05BA03	1166843	Legalon 140 mg Kaps.	103181	1,29
A02BC02	1328464	Pantoloc 20 mg Filmtabl.	87103	1,09
N06DX02	1199305	Tebofortan 40 mg Filmtabl.	63163	0,69
L01BA01	2443061	Ebetrexat 10 mg Tabl.	54024	0,67
B03BB01	713987	Folsan 5 mg Tabl.	53665	0,67
A02BC05	1345793	Nexium 40 mg magensaftres. Tabl.	51584	0,64

Diese 10 häufigsten Pharnummern stellen in Summe 13,42% aller abgegebenen Rezepte, denen die ICD9 – Gruppe 150-159 zugeordnet wurde, dar. Insgesamt wurden 9309 unterschiedliche Pharnummern erfasst, von denen 3314 weniger als 10-mal abgegeben wurden.

- Es ergibt sich die Notwendigkeit für weitere Analysen nicht den vollständigen ATC-Code zu betrachten, sondern abzuschneiden. Daher wurden nur die ersten vier Stellen des ATC-Codes berücksichtigt und erneut analysiert.

Diese Analyse enthält keine Pharnummern und keine Medikamentennamen mehr, da diese nun in ATC-Gruppen auf vier Stellen zusammengefasst sind. Die 10 größten Gruppen ergeben sich wie folgt:

ATC-Code	Anzahl	Prozent
N02A	794451	9,9
A02B	727385	9,06
A09A	443601	5,53
M01A	300492	3,74
N06A	246443	3,07
C07A	219917	2,74
C10A	187811	2,34
A05B	171839	2,14
C09A	170693	2,13
L01B	167626	2,09

Die so erfassten Rezepte stellen 42,74% aller eingelösten Medikamente dar.

5 Analyse und Interpretation für die ICD9 Gruppe 160-165

Mit einer zugeordneten ICD9 Gruppe 160-165 wurden im Betrachtungszeitraum 2006-2007 im ATC-ICD Projekt 8.831.905 Rezepte mit einschlägiger Diagnose erkannt. Durch die Verknüpfung mit der Rezepte – Datenbank erhält man zu 222.520 keinen ATC-Code (zum Beispiel, da es sich um MAGISTRALE ZUBEREITUNGEN oder MELOLIN WUNDKOMPR ST 10X10CM, ... handelt).
C09A 209926 2,38

Dies entspricht ca. 2,5% der Gesamtdatensätze.

Bei den verschriebenen und eingelösten Rezepten sind folgende 10 Pharmanummern am häufigsten enthalten:

ATC-Code	Pharnr	Name	Anzahl	Prozent
R03AK03	1350357	Berodual Dosieraerosol	192232	2,18
A02BC02	1292631	Pantoloc 40 mg Filmtabl.	161065	1,82
	9999927	MAGISTRALE ZUBEREITUNGEN	129931	1,47
R03AK06	1330515	Seretide Diskus forte 50/500 mcg Plv. z. Inh.	121998	1,38
R03AK06	1330509	Seretide Diskus standard 50/250 mcg Plv. z. Inh.	121892	1,38
R03AK07	1346924	Symbicort Turbohaler Dosier-Pulverinhalator	114125	1,29
A02BC02	1328464	Pantoloc 20 mg Filmtabl.	83449	0,94
R03AC13	1300483	Foradil Kaps. z. Trockeninh.	80686	0,91
G04CA02	1304274	Alna retard 0,4 mg Kaps.	65279	0,74
N02CC06	2427091	Relpax 40 mg Filmtabl.	59157	0,67

Diese 10 häufigsten Pharmanummern stellen in Summe 12,79% aller abgegebenen Rezepte, denen die ICD9 – Gruppe 160-165 zugeordnet wurde, dar. Insgesamt wurden 9070 unterschiedliche Pharmanummern erfasst, von denen 3265 weniger als 10-mal abgegeben wurden.

- Es ergibt sich die Notwendigkeit für weitere Analysen nicht den vollständigen ATC-Code zu betrachten, sondern abzuschneiden. Daher wurden nur die ersten vier Stellen des ATC-Codes berücksichtigt und erneut analysiert.

Diese Analyse enthält keine Pharmanummern und keine Medikamentennamen mehr, da diese nun in ATC-Gruppen auf vier Stellen zusammengefasst sind. Die 10 größten Gruppen ergeben sich wie folgt:



ATC-Code	Anzahl	Prozent
R03A	851179	9,64
N02A	742806	8,41
A02B	682735	7,73
R03B	565672	6,4
M01A	303583	3,44
N06A	237961	2,69
C07A	229820	2,6
C10A	227737	2,58
R03D	216489	2,45
C09A	209926	2,38

Die so erfassten Rezepte stellen 48,32% aller eingelösten Medikamente dar.

6 Analyse und Interpretation für die ICD9 Gruppe 230-234 (Carcinoma in situ)

Mit einer zugeordneten ICD9 Gruppe 230-234 wurden im Betrachtungszeitraum 2006-2007 im ATC-ICD Projekt 83.173 Rezepte mit einschlägiger Diagnose erkannt. Durch die Verknüpfung mit der Rezepte – Datenbank erhält man zu 2872 keinen ATC-Code (zum Beispiel, da es sich um MAGISTRALE ZUBEREITUNGEN oder MELOLIN WUNDKOMPR ST 10X10CM, ... handelt). Dies entspricht ca. 3,5% der Gesamtdatensätze.

Bei den verschriebenen und eingelösten Rezepten sind folgende 10 Pharmanummern am häufigsten enthalten:

ATC-Code	Pharnr	Name	Anzahl	Prozent
A04AA03	1279518	Navoban 5 mg Kaps.	1880	2,26
	9999927	MAGISTRALE ZUBEREITUNGEN	1685	2,03
A02BC02	1292631	Pantoloc 40 mg Filmtabl.	1630	1,96
L02BB03	1313528	Casodex 150 mg Filmtabl.	1603	1,93
L02BB03	1302714	Casodex 50 mg Filmtabl.	1034	1,24
L01AA02	31696	Leukeran 2 mg Filmtabl.	894	1,07
C02KX01	2436894	Tracleer 125 mg Filmtabl.	825	0,99
L03AA13	2442972	Neulasta 6 mg Inj.lsg.	757	0,91
D10BA01	2441872	Ciscutan 20 mg Kaps.	746	0,90
A11CC04	593773	Rocaltrol Roche 0,25 mcg Kaps.	704	0,85

Diese 10 häufigsten Pharmanummern stellen in Summe 14,14% aller abgegebenen Rezepte, denen die ICD9 – Gruppe 230-234 zugeordnet wurde, dar. Insgesamt wurden 3507 unterschiedliche Pharmanummern erfasst, von denen 2484 weniger als 10-mal abgegeben wurden.

- Es ergibt sich die Notwendigkeit für weitere Analysen nicht den vollständigen ATC-Code zu betrachten, sondern abzuschneiden. Daher wurden nur die ersten vier Stellen des ATC-Codes berücksichtigt und erneut analysiert.

Diese Analyse enthält keine Pharmanummern und keine Medikamentennamen mehr, da diese nun in ATC-Gruppen auf vier Stellen zusammengefasst sind. Die 10 größten Gruppen ergeben sich wie folgt:

ATC-Code	Anzahl	Prozent
A04A	6250	7,51
A02B	5827	7,01
L01A	5418	6,51
L02B	4130	4,97
A11C	3703	4,45
N06A	2326	2,8
H02A	2170	2,61
N02A	1950	2,34
M01A	1810	2,18
V03A	1734	2,08

Die so erfassten Rezepte stellen 42,46% aller eingelösten Medikamente mit zugeordneter ICD9 Gruppe 230-234 dar.



7 Resümee

Aus den Analysen der abgegebenen Medikamente auf Basis von Pharmanummern erhält man aus den Häufigkeiten neben allgemein sehr häufig verschriebenen Medikamenten (z.B.: Pantoloc 40mg Filmtablette) auch die wichtigsten Präparate zur medikamentösen Therapie von Karzinomen. Die so erhaltenen Medikamentenliste gekoppelt mit medizinischem Fachwissen kann für die Generierung einer detaillierten Liste zur medikamentösen Therapie spezieller Krebserkrankungen herangezogen werden.

Durch die Klassifizierung der Diagnosen in ICD9 – Gruppen ist diese Analyse allerdings in ihrem Detailgrad beschränkt. Dennoch kann für spezielle Medikamente ein Ausschlusskriterium definiert werden (falls eine Person in den GAP-DRG Daten bereits eine medikamentöse Therapie mit einem Chemotherapeutikum bekommen hat, bevor der erste Krankenhausaufenthalt mit einschlägiger Diagnose erkannt wird, muss diese Person aus den Neuerkrankungsfällen ausgeschlossen werden).



C. Kontrolle der Größenordnung von Todesfällen auf Grund von Krebs bei Kindern – Vergleich mit Statistik Austria Daten

Folgende SQL – Abfrage wurde auf der GAP-DRG Datenbank abgesetzt um die Anzahl an Kindern unter 15 Jahren zu erkennen die mit einer Diagnose C im Zeitraum 2006/07 verstorben sind:

```
select count(distinct pers_id)
from
(select *
from personen
where gebjahr >=1992
and extract(year from toddat) in (2006,2007)) as kinder
join mbds_kaperson on pers_id=id_person_p
natural join mbds_diagnosen
where icd_subkategorie like 'C%';
```

Man ordnet damit jeder Person, die ab 1992 geboren wurde (also in 2006 höchstens 14 wurde) und die außerdem in 2006 oder 2007 ein eingetragenes Todesdatum hat, ihre MBDS-Diagnosen zu. Daraus filtert man dann die C-Diagnosen und zählt die eindeutigen Personen-IDs. Diese Abfrage liefert 46 unterschiedliche Personen-IDs.

Laut Statistik Austria

(http://www.statistik.at/web_de/static/gestorbene_2009_nach_todesursachen_alter_und_geschlecht_021987.xls) gab es in 2009 26 Sterbefälle in der Gruppe der 0-15 jährigen (wenn man aufaddiert). Vergleicht man nun mit dem Mittelwert für die Jahre 2006/07 ermittelt aus Daten der GAP-DRG Datenbank so erhält man 23 Fälle (aus GAP-DRG) vs. 26 Todesfälle

Das bestätigt die Annahme, dass die Erfassung von Todesfällen in der richtigen Größenordnung durch die MBDS – Daten gewährleistet wird.

D. Krebserkrankungen – Literaturrecherche

Die folgende Literaturanalyse basiert auf einer Abstimmung zwischen dwh simulation services und HVB-EBM Mitarbeitern. Die Analyse wird nicht systematisch durchgeführt und beschränkt sich primär auf Fallzahlen, Screening und dessen Einfluss sowie gegebenenfalls auf klassische Behandlungswege.

1 Allgemeine Zielsetzung

Basierend auf den Auswertungsergebnissen der Krebsinzidenzen für 2007 von ausgewählten Krebserkrankungen an Hand der GAP-DRG Datenbank soll eine weiterführende Modellierbarkeit der Inzidenzentwicklung, optional auch des Screeningeinflusses, untersucht werden.

Dazu soll eine nicht systematische Literaturrecherche in PubMed durchgeführt werden. Die Anforderungen an die Suche beschränken sich primär auf Fallzahlen, Screening und dessen Einfluss sowie Basismodellierungsansätze.

1.1 Vorgangsweise und Beschränkungen

Um die Komplexität bzw. den Arbeitsaufwand in der nötigen Dimension zu halten, und da eine gewisse Aktualität der betrachteten Literatur gefordert wird, werden eine Reihe von Einschränkungen an die Suche getroffen:

Zeitlicher Horizont	Die letzten 5 Jahre (Stichtag 28.02.2011)
Sprache der Veröffentlichung	Englisch, Deutsch
Voraussetzungen an den Text	Der Abstract muss frei zugänglich sein
Modellierungsmethoden	Markov Modelle, agentenbasierte Simulation, Statistisches „Forecast Model“

Nicht analysiert werden können hierbei der Einfluss von neuen Behandlungsstrategien sowie klinische Studien zu unterschiedlicher medikamentöser Therapie.

Die Einschränkung der Literaturanalyse zu dynamischen Modellen auf primär Agenten basierte Modellbildungsansätze und Markov Ketten sowie Markov Modelle höherer Stufe wird vorgenommen, da diese Ansätze auf Grund des vorhandenen Datenwissens eine ausreichende Abbildungsmöglichkeit darstellen. Diese Einschränkung gegenüber weiterer dynamischer Methoden wie etwa System Dynamics und Differentialgleichungen ist möglich, da keine Ansteckung der betrachteten Bevölkerung untereinander angenommen wird und keine sozialen Einflüsse auf die Response Rate an einem Screening Programm modelliert werden soll.

2 Literaturrechercheergebnisse

2.1 Abfrageergebnisse für Markov Modelle

Zunächst soll nach dem Modellbildungsprinzip der Kohorten, und hier nach dem hauptsächlich verwendeten Prinzip der Markov Modelle gesucht werden. Unter den in 2.3.1 angegebenen Einschränkungen hinsichtlich Aktualität und Sprache wird folgender Suchbegriff abgesetzt:

cancer incidence model markov

Man erhält dadurch 208 Treffer die zunächst nach ihrem Erscheinungsjahr sortiert werden und anschließend auf Titelebene vorgefiltert werden.

Erscheinungsjahr:

2011	1
2010	63
2009	40
2008	44
2007	43
2006	17

Aus der folgenden Analyse auf Abstractebene werden folgende Hauptanwendungsgebiete identifiziert:

- HPV – Screening und Inzidenzen
- Darmkrebsscreening
- Prostatakarzinom (primär Screening)
- eine große Anzahl behandelt spezielle medikamentöse Therapie, diese Arbeiten scheiden für die gegebenen Fragestellung aus
- Leber (Krebs, Hepatitis, Transplantationen)

Man erkennt bei den zu berücksichtigenden Arbeiten eine klare Tendenz zur Untersuchung und Bewertung von Screeningmaßnahmen. Screening hat durch den direkten Einfluss auf das Erkennungsstadium von Tumoren einen hohen Einfluss auf Restlebenserwartung bei Erkrankung und die Todesfallwahrscheinlichkeit.

Aus den Abstracts kann im Allgemeinen nicht näher auf das umgesetzte Modell bzw. die verwendeten Datenquellen rückgeschlossen werden. Daher ist für eine detaillierte Bearbeitung einer Fragestellung im Bereich Krebsinzidenzentwicklung eine weitere Analyse auf Volltextebene unumgänglich.

Klar ersichtlich ist allerdings, dass die Methode der Markov Modelle in der Literatur zur Krebsinzidenzmodellierung Anwendung findet.

2.2 Abfrageergebnisse für Agenten basierte Modelle

Eine Bearbeitung der Behandlungsweg- bzw. Inzidenzmodellierung auf Einzelpersonenbasis kann mit Hilfe Agenten basierter Simulationsmethoden durchgeführt werden. Dieser Modellbildungsansatz ist konträr zu Kohorten basierten Methoden auf einzelne Individuen zugeschnitten und liefert Ergebnisse und Einschätzungen für die Gesamtbevölkerung auf Grund von Summation der Berechnungsergebnisse für Einzelpersonen. Das bedeutet, dass die Zusammenfassung und Verwendung von statistischen Gesamtgrößen erst im letzten Schritt erfolgt, bis dahin wird die Heterogenität der einzelnen betrachteten Personen falls erforderlich in einem höheren Detailgrad abgebildet.

Agenten basierte Modellierungsmethoden sind i.A. rechenaufwändig und nicht ohne mathematische Software bzw. der Anwendung einer Programmiersprache bewerkstelligbar. Daher benötigen sie interdisziplinäres Arbeiten und waren mit den bis vor Kurzem vorhandenen Berechnungskapazitäten nicht berechenbar. Es ist daher davon auszugehen, dass die Anzahl an Veröffentlichungen zu Krebsinzidenzmodellierung für diese Methode beschränkt ist, die vorhandenen Ansätze jedoch sehr wohl interessante Aspekte liefern können.

Unter den in 1.1 angegebenen Einschränkungen hinsichtlich Aktualität und Sprache wird folgender Suchbegriff in PubMed abgesetzt:



cancer incidence model agent based

Man erhält dadurch 34 Treffer die zunächst nach ihrem Erscheinungsjahr sortiert werden und anschließend auf Titel-/Abstractebene gefiltert werden.

Erscheinungsjahr:

2011	1
2010	7
2009	10
2008	4
2007	5
2006	7

Bei der Bearbeitung der Abfrageergebnisse kristallisiert sich eine Gruppe von falsch detektierten Arbeiten heraus, die den Begriff Agent aus molekularbiologischer Sicht verwenden. Diese Arbeiten legen den Fokus primär auf neue Behandlungsstrategien und den Vergleich unterschiedlicher medikamentöser Behandlungsmethoden.

Dennoch konnten Arbeiten, welche die Agenten basierte Simulationstechnik verwenden, identifiziert werden. Zwei von ihnen sind exemplarisch hier angeführt:

1. Modeling the cost-effectiveness of colorectal cancer screening: policy guidance based on patient preferences and compliance.
Subramanian S, Bobashev G, Morris RJ.
2. An "age"-structured model of hematopoietic stem cell organization with application to chronic myeloid leukemia.
Roeder I, Herberg M, Horn M.

Die primäre Begründung bzw. die Notwendigkeit zur Verwendung von auf Einzelpersonen bezogenen Methoden im Gegensatz zu Kohortenmodellen begründet sich in der mitunter gegebenen Notwendigkeit der Betrachtung von heterogenem Verhalten (zum Beispiel bei der Annahme von Screeningangeboten nach Alter oder sozialem Status) und fein gegliederter Altersstruktur (die im Aufzählungspunkt 2 angeführte Publikation ist hierzu ein gutes Beispiel).

2.3 Zusätzliche Internetliteratur, interessante Artikel und Ansätze

Um auch auf weitere Quellen bzw. Netzwerke aufmerksam zu werden wurde eine freie Internetsuche als Erweiterung der bisherigen Arbeiten angehängt. Dabei wurde nicht systematisch von einzelnen angeführten Quellen in den bereits gefundenen Arbeiten ausgegangen und von diesen aus weitergesucht.

Generell ist auf das Cisnet (<http://cisnet.cancer.gov/>) Netzwerk zu verweisen. Cisnet steht in diesem Zusammenhang für „Cancer Intervention and Surveillance Modelling Network“ und gehört zu den U.S. National Institutes of Health.

Unter anderem findet man in diesem Netzwerk <http://cisnet.cancer.gov/grants/lung/kimmel.html>. Es wird auf vier Krebsarten (Brust-, Darm-, Lungen- und Prostatakrebs) fokussiert. Die Erstellung von Modellen wird primär in Kooperationen mit Universitäten durchgeführt. Als Beispiel im Bereich Brustkrebs dient der folgende Link: <http://cisnet.cancer.gov/breast/profiles.html>

Zur Vorausschätzung von Krebsinzidenzen bzw. -sterblichkeit wurden folgende Arbeiten als möglicher Ausgangspunkt erkannt:



1. Eine Arbeit mit dem Titel „Mathematical models for human cancer incidence rates“ des Max-Planck Instituts für Demografische Forschung
<http://www.demogr.mpg.de/cgi-bin/publications/paper.plx?l=de&t=&pubid=2234&personalid=515&listtype=-98>
2. Die Arbeit mit dem Titel: „Estimates of the cancer incidence and mortality in Europe in 2006“, der Ende 2006 erschienen ist
<http://annonc.oxfordjournals.org/content/18/3/581.short>
Verwendet werden „short-term prediction methods“ und man erhält dadurch Schätzer für Europa.
3. Die Arbeit mit dem Titel: „Weitere Entwicklung der Krebssterblichkeit in Deutschland bis zum Jahr 2010“ aus dem Jahr 2003.
<http://www.ist-world.org/ResultPublicationDetails.aspx?ResultPublicationId=590e9faa33574e2ab8a9863e9bd5519f>

Diese Arbeit kann mit den nun vorhandenen Daten evaluiert werden und man erhält dadurch eine retrospektive Bewertung der Aussagequalität. Man ersieht dadurch ob dieser Ansatz auch für eine Anwendung in Österreich Potential birgt oder ob er strukturell erweitert werden soll.

2.4 Internetliteratursucheergebnisse mit Screeningbezug durch offene Suche in Google – auch domänenspezifische Modelle zugelassen

Um auch speziell Fragen des Screenings Rechnung zu tragen wurden die folgenden Arbeiten exemplarisch ausgewählt:

- „Overdiagnosis Due to Prostate-Specific Antigen Screening: Lessons From U.S. Prostate Cancer Incidence Trends“ aus dem Jahr 2002
- „The German Cervical Cancer Screening Model: development and validation of a decision-analytic model for cervical cancer screening in Germany“
<http://eurpub.oxfordjournals.org/content/16/2/185.full>



E. Technische Realisierung der Abfragen zu ausgewählten Krebserkrankungen am Beispiel Magenkrebs

Im Folgenden ist der umgesetzte Code für Abfragen zu Magenkrebs dargestellt. Kommentare sind in orange gehalten.

Die hier dargestellte Struktur wird nach eingehender Evaluierung für alle vorgesehenen Krebsarten abgearbeitet und die Resultate bilden den Basisdatensatz für die anschließende statistische Detailanalyse die nicht mehr mit SQL, sondern mit statistischer Standardsoftware erfolgt.

```
/*
*****
*****
Ermitteln der Personen fuer die Krebsinzidenz
*****
*****
*/

/*
  Selektieren der IDs der ICD-Codes
*/
create view p_einzinger.magen_diagnosen as
select id as diag_id, icd10_subkat
from diagnose
where icd10_subkat like 'C16%' or icd10_subkat like 'D00.2%'
or icd10_subkat like 'D37.1%'

/*
  Alle Aufenthalte (plus Aufnahme- und Entlassungsdatum,
  Jahr sowie Personen-ID) in 2006 und 2007 mit den Diagnosen
  in die Tabelle p_einzinger.magen_aufenthalte schreiben
*/
create table p_einzinger.magen_aufenthalte as
select aufenthalt_id, id_person_p, aufnahme_datum, entlassung_datum,
jahr
from
  (select distinct(aufenthalt_id)
   from mbds_diagnosen
   natural join p_einzinger.magen_diagnosen) as aufenthalte
natural join mbds_kaperson natural join mbds_aufenthalte;

/*
  Wie oben, aber nur Aufenthalte in 2006
*/
create table p_einzinger.magen_aufenthalte_2006 as
select aufenthalt_id, id_person_p, aufnahme_datum, entlassung_datum
from
  (select distinct(aufenthalt_id)
   from mbds_diagnosen
   natural join p_einzinger.magen_diagnosen) as aufenthalte
```



```
natural join mbds_kaperson natural join mbds_aufenthalte
where jahr=2006;

/*
*****
TESTEN, ob AUFENTHALTE EINDEUTIG SIND
*****
*/

/*
  Test, ob Aufenthalte in p_einzinger.magen_aufenthalte
  distinct sind.
  Es ergeben sich aber 8 Aufenthalte, die öfter als einmal
  vorkommen.
*/
select aufenthalt_id, count(*) as Anzahl
from p_einzinger.magen_aufenthalte
group by aufenthalt_id
having count(*)>1;

/*
  Test, ob Aufenthalte in p_einzinger.magen_aufenthalte_2006
  distinct sind.
  Es ergeben sich 5 Aufenthalte, die öfter als einmal vorkommen.
*/
select aufenthalt_id, count(*) as Anzahl
from p_einzinger.magen_aufenthalte_2006
group by aufenthalt_id
having count(*)>1;

/*
  Test, ob den Aufenthalts-IDs in mbds_kaperson jeweils
  nur eine Person zugeordnet ist.
  Ergibt allerdings 2496 Aufenthalt-IDs, die mehr als einer
  Person zugeordnet sind. Daher sind die Aufenthalte in
  magen_aufenthalte
  auch nicht distinct.
*/
select distinct(aufenthalt_id)
from mbds_kaperson as ka1
where exists(select *
from mbds_kaperson as ka2
where ka1.aufenthalt_id=ka2.aufenthalt_id and
ka1.id_person_p!=ka2.id_person_p)
order by aufenthalt_id;

/*
*****
PERSONEN MIT AUFENTHALT IN 2006 AUSSCHLIESSEN
*****
*/

/*
  Personen, die in obigen Tabellen einen Aufenthalt hatten,
```



```
    aber keinen in 2006, auslesen und in Tabelle
    p_einzinger.magen_personen schreiben
*/
create table p_einzinger.magen_personen as
((select distinct id_person_p
from mbds_kaperson natural join p_einzinger.magen_aufenthalte)
except
(select distinct id_person_p
from mbds_kaperson natural join
p_einzinger.magen_aufenthalte_2006));

/*
    Auslesen der Tabelle, geordnet nach Personen-IDs
*/
select *
from p_einzinger.magen_personen
order by id_person_p;

/*
    Aufenthalte nur von den Personen nehmen, die keinen Aufenthalt
    in 2006 hatten.
*/
create table p_einzinger.magen_aufenthalte_2007 as
select *
from p_einzinger.magen_personen natural join
p_einzinger.magen_aufenthalte;

/*
    Test, ob Aufenthalte in p_einzinger.magen_aufenthalte_2007
    distinct sind.
    Es ergeben sich 2 Aufenthalte, die öfter als einmal vorkommen.
*/
select aufenthalt_id, count(*) as Anzahl
from p_einzinger.magen_aufenthalte_2007
group by aufenthalt_id
having count(*)>1;

/*
*****
FOLGEAUFENTHALTE
*****
*/

/*
    Erstaufenthalte herausfinden
*/
create table p_einzinger.magen_erstaufenthalte as
select *
from p_einzinger.magen_aufenthalte_2007 as aufenthaltel
where aufnahme_datum <= all (select aufnahme_datum
    from p_einzinger.magen_aufenthalte_2007 as aufenthalte2
    where aufenthaltel.id_person_p=aufenthalte2.id_person_p);

select aufenthalt_id, count(*) as Anzahl
```



```
from p_einzinger.magen_erstaufenthalte
group by aufenthalt_id
having count(*) > 1;

select id_person_p, count(*) as Anzahl
from p_einzinger.magen_erstaufenthalte
group by id_person_p
having count(*) > 1;

select *
from p_einzinger.magen_erstaufenthalte
where aufenthalt_id in (2627664,2554009);

select *
from p_einzinger.magen_erstaufenthalte
where id_person_p=8801344;

/*
  Erstellen der Folgeaufenthalte
*/
create table p_einzinger.magen_folgeaufenthalte as
select *
from p_einzinger.magen_aufenthalte_2007
except select *
from p_einzinger.magen_erstaufenthalte;

/*
  Aufenthalte aus der Tabelle magen_aufenthalte_2007, die entweder
  noch einen Folgeaufenthalt hatten, oder deren Erstaufenthalt nach
  dem 30. November 2007 war. Die idente Abfrage wird abermals (hier
  nicht abgebildet) fuer den 31.Oktober durchgefuehrt)
*/
create table p_einzinger.magen_aufenthalte_final as
select *
from p_einzinger.magen_aufenthalte_2007 as ges
where exists
  (select *
   from p_einzinger.magen_folgeaufenthalte as folge
   where ges.id_person_p=folge.id_person_p)
union
select *
from p_einzinger.magen_erstaufenthalte as erst
where erst.entlassung_datum > '2007-11-30';

/*
  Diverse Tabellen ausgeben und dann exportieren.
*/
select *
from p_einzinger.magen_aufenthalte_final
order by id_person_p, aufenthalt_id;

select distinct id_person_p
from p_einzinger.magen_aufenthalte_final
```



```
order by id_person_p;

select *
from p_einzinger.magen_folgeaufenthalte
order by id_person_p, aufenthalt_id;

select *
from p_einzinger.magen_erstaufenthalte
order by id_person_p, aufenthalt_id;

select *
from p_einzinger.magen_aufenthalte_2007
order by id_person_p, aufenthalt_id;

select *
from p_einzinger.magen_aufenthalte_2006
order by id_person_p, aufenthalt_id;

select *
from p_einzinger.magen_aufenthalte
order by id_person_p, aufenthalt_id;

select *
from p_einzinger.magen_personen
order by id_person_p;

/*
  Ausgeben der Medikamente mit ATC-Code und Name, die von
  Personen am gleichen Tag per Rezept eingeloeset wurden,
  an denen sie laut atcid ein Rezept eingeloeset haben,
  das auf die icd-9-gruppe 150-159 matcht.
*/
create table p_einzinger.gr150159 as
select atc, pharnr, name, Anzahl
from
heilmittel_hvb natural join
  (select pharnr, count(*) as Anzahl
  from rezeptdaten join
    (select pers_id, datum
     from atcid
     where icd9_gruppe = '150 - 159') as person_datum
  on id_person_p=pers_id and eindat=datum
  group by pharnr) as pharmanummern
order by atc asc;

/*
  Ausgeben der Tabelle, geordnet nach Anzahl.
*/
select *
from p_einzinger.gr150159
order by anzahl desc;
```



Die durch diese Abfragen erzeugten Dateien haben folgende Inhalte:

magen_aufenthalte.csv ... sind alle Aufenthalte mit den Diagnosen zu Magenkrebs

magen_aufenthalte_2006.csv ... nur die aus 2006

magen_personen.csv ... sind dann die Personen, die Aufenthalte in 2007 aber nicht in 2006 hatten

magen_aufenthalte_2007.csv ... die Aufenthalte jener Personen

magen_erstaufenthalte.csv und magen_folgeaufenthalte.csv ... sind obige Aufenthalte, getrennt in Erst- und Folgeaufenthalte

magen_aufenthalte_final.csv ... sind dann die Aufenthalte von Personen, die einen Folgeaufenthalt hatten, oder bei denen der Erstaufenthalt nach dem 30.11.2007 war

magen_personen.csv ... sind die Personen zu diesen Aufenthalten

Außerdem zeigt Doppelte_Erstaufenthalte.csv die beiden Aufenthalte, die doppelt in den MBDS - Daten sind (weil verschiedenen Personen zugeordnet), und Person_zwei_Erstaufenthalte.csv zeigt die einzige Person, die zwei Aufenthalte hat, die potentiell als Erstaufenthalt identifiziert wurden. Das liegt daran, weil der erste Aufenthalt nur eintägig war und die Person noch am gleichen Tag in ein anderes Krankenhaus aufgenommen wurde.