

# Beurteilung der Anpassungsgüte multivariater Regressionsmodelle in Überlebenszeitmodellen unter besonderer Berücksichtigung des C-Index

Katrin Kupas  
Statistical Consultant  
Im Rosenträger 22  
60388 Frankfurt am Main  
katrin.kupas@web.de

Hendrik Schmidt  
Boehringer Ingelheim Pharma GmbH & Co KG  
Birkendorfer Straße 65  
88397 Biberach an der Riß  
hendrik.schmidt@boehringer-ingelheim.com

## Zusammenfassung

In der Arzneimittelentwicklung haben Subgruppenanalysen und die Identifizierung bestimmter prognostischer Faktoren und Prädiktoren für Krankheiten und besonders Krankheitsverläufe in den letzten Jahren an Bedeutung gewonnen. Verstärkt wird diese Bedeutung zudem durch die zunehmend frühe Nutzenbewertung von Arzneimitteln sowie durch den Trend zu personalisierter Medizin, bei der anhand von bestimmten Charakteristika der Erfolg einer Therapie bei individuellen Patienten eingeschätzt wird.

In vielen Therapiegebieten ist die Zeit bis zum Eintritt eines Ereignisses (Tod, Verschlechterung des Krankheitszustandes, etc.) ein wichtiger und patientenrelevanter Endpunkt zur Bestimmung des Krankheitsverlaufes. Häufig wird zu Beginn einer Studie von jedem Patienten eine Vielzahl an Charakteristika erhoben, die als mögliche prognostische Faktoren und Prädiktoren für diese Endpunkte in Frage kommen. Multivariate Regressionsmodelle sind eine Möglichkeit, aus diesen Kovariaten signifikante prognostische Faktoren und Prädiktoren und damit entsprechende Patientensubgruppen zu identifizieren und auszuwerten.

SAS<sup>®</sup> bietet mit PROC PHREG für Überlebenszeitdaten verschiedene Algorithmen an, aus einem Pool an Kovariaten automatisch das optimale Modell auszuwählen. Beispiele sind die vorwärts- oder rückwärtsgerichtete oder auch die bidirektionale Selektion. Dabei werden anhand der einzelnen Chi-Quadrat-Statistiken und einer vom Benutzer zu setzenden Signifikanzgrenze nacheinander die signifikantesten Kovariaten ausgewählt oder die am wenigsten signifikanten Kovariaten eliminiert.

Insbesondere beim Vorliegen vieler Kovariaten und/oder kleiner Patientenzahlen stellt sich die Frage nach der Anpassungsgüte der erhaltenen Cox-Modelle. Der C-Index als absolute Maßzahl bietet eine Möglichkeit, die Anpassungsgüte des Modells an die Daten zu bewerten und verschiedene Modelle zu vergleichen. Wir stellen den mathematischen Hintergrund zum C-Index dar und präsentieren einen Algorithmus zu dessen Berechnung, der mit SAS<sup>®</sup> implementiert wurde. Zudem präsentieren wir ein Beispiel, wie die Variablenselektion mit Hilfe des C-Index bewertet werden kann.

**Schlüsselwörter:** Modellanpassung, multivariate Cox-Regression, optimales Modell, Anpassungsgüte, C-Index

## **1 Einleitung und Überblick**

Subgruppenanalysen und die Identifizierung bestimmter prognostischer Faktoren und Prädiktoren für Krankheitsereignisse und insbesondere Krankheitsverläufe haben in den letzten Jahren zunehmend an Bedeutung gewonnen. Diese gestiegene Bedeutung wird zudem durch die zunehmend häufigere und in vielen Ländern institutionalisierte frühe Nutzenbewertung von Arzneimitteln sowie durch den Trend zu personalisierter Medizin unterstrichen.

Die Überlebenszeit beziehungsweise die Zeit bis zum Eintritt eines bestimmten Ereignisses ist in vielen Therapiegebieten ein wichtiger und patientenrelevanter Endpunkt zur Bestimmung des Krankheitsverlaufes. Beispiele finden sich in vielen Therapiegebieten. In der Onkologie wird häufig die Zeit bis zum Eintritt einer Progression untersucht, in den Atemwegserkrankungen ist es die Zeit bis zur ersten Exazerbation (Verschlechterung des Krankheitszustandes). Exazerbationen sind durch eine Kombination aus Symptomen (zum Beispiel Sputum und Husten) und Ereignissen (zum Beispiel Behandlung mit Steroiden und/oder Antibiotika) definiert und werden üblicherweise in moderate und schwere (zusätzlich mit Hospitalisierung) Exazerbationen eingeteilt.

Zu Beginn einer Studie werden üblicherweise von jedem Patienten eine Vielzahl an Charakteristika erhoben, die als mögliche prognostische Faktoren und Prädiktoren für das Risiko des Patienten, ein solches Ereignis zu haben, in Frage kommen. Multivariate Regressionsmodelle, angewendet auf Überlebenszeitdaten, sind eine Möglichkeit, aus diesen Charakteristika signifikante prognostische Faktoren und Prädiktoren und damit entsprechende Patientensubgruppen zu identifizieren und auszuwerten.

Die klassische Analyse von Überlebenszeitdaten erfolgt mit Hilfe der Cox-Regression, mit Studienbehandlung als alleinige Kovariate in der Analyse, um den Effekt der zu testenden Behandlung im Vergleich zu Placebo oder aktiver Kontrolle zu berechnen. Um jedoch Vorhersagen über Therapieerfolge bzw. Krankheitsverläufe aufgrund bestimmter Patientencharakteristika machen zu können, können als Sensitivitätsanalyse noch weitere Kovariaten in das Cox'sche Regressionsmodell mit aufgenommen werden. Dann spricht man von einem multivariaten Regressionsmodell. Die statistische Aufgabe besteht darin, ein „optimales“ Modell zu finden, d.h. ein Modell, bei dem alle enthaltenen Kovariaten einen signifikanten Einfluss auf die Zeit bis zum Eintritt des interessierenden Ereignisses haben.

In diesem Beitrag beschreiben wir den C-Index als absolutes Maß zur Bewertung der Anpassungsgüte eines multivariaten Cox'schen Regressionsmodell. Zunächst wird jedoch die multivariate Cox-Regression eingeführt und es werden ganz allgemein die in SAS<sup>®</sup> vorhandenen automatischen Algorithmen zur Variablenselektion vorgestellt. Da-

nach wird der C-Index theoretisch eingeführt und ein Algorithmus zu seiner Berechnung angegeben. Schließlich wird dieser Algorithmus auf ein konkretes Beispiel mit Daten einer Atemwegsstudie angewendet.

## 2 Überlebenszeit und multivariate Cox-Regression

### 2.1 Überlebenszeitanalysen und -daten

Überlebenszeiten, also Zeiten bis zum Eintritt eines bestimmten klinisch relevanten Ereignisses, sind häufig wichtige Endpunkte in klinischen Studien. Sie berücksichtigen die Zeit, in der ein Patient unter Risiko ist, das Ereignis von Interesse zu erfahren. Eine Besonderheit solcher Daten ist, dass es sich um zensierte Daten handelt, da der Beobachtungszeitraum einer klinischen Studie nur eine bestimmte Zeitspanne umfasst und somit das Ereignis von Interesse bei einer bestimmten Anzahl an Patienten gar nicht beobachtet werden kann.

Die Cox-Regression, beziehungsweise das proportionale Hazardmodell (daher PROC PHREG), ist in der Praxis eine der populärsten Regressionsmethoden zur Analyse von Überlebenszeitdaten. Sie erlaubt in ihrer multivariaten Form Analysen der Hazardrate und des Hazard-Ratios zwischen verschiedenen Behandlungen, die für diverse Kovariaten adjustiert werden können [1]. Solche Kovariaten sind meist Charakteristika der Patienten, wie zum Beispiel Demographie, Begleitmedikationen und –erkrankungen oder auch Daten aus Patienten-Fragebögen, und werden in klinischen Studien vor Einnahme der ersten Studienmedikation (Baseline) erhoben. Diese Kovariaten können sowohl prognostische als auch prädiktive Faktoren für das Risiko sein, ein bestimmtes Ereignis zu erfahren.

### 2.2 Multivariate Cox-Regression

Im Folgenden wird die Überlebenszeit als Zufallsvariable mit Dichtefunktion  $f(t)$ , kumulativer Verteilungsfunktion  $F(t)$  und Überlebensfunktion  $S(t)=1-F(t)$  für  $t \geq 0$  modelliert [2]. Die Hazardfunktion  $h(t)$  gibt das Risiko eines Patienten für das Eintreten des Ereignisses zu einer bestimmten Zeit  $t$  an, unter der Annahme, dass das Ereignis noch nicht eingetreten ist. Zwischen der Hazardfunktion, der Dichtefunktion und der Verteilungsfunktion besteht der Zusammenhang  $h(t) = f(t)/(1-F(t))$ .

In einem multivariaten Modell nehmen wir an, dass  $h(t)$  von  $m$  Kovariaten  $X_1, \dots, X_m$  mit Werten  $x_1, \dots, x_m$  abhängt. Die Zufallsvariablen  $X_1, \dots, X_m$  sind üblicherweise Charakteristika von Patienten, die vor Einnahme der Studienmedikation an Tag 0 ( $t=0$ , also zu Baseline) erhoben werden, sowie die Behandlungsgruppe der Patienten. Die Hazardfunktion von Patient  $i$  zum Zeitpunkt  $t$  ist dann gegeben durch

$$h_i(t) = h_0(t) \exp\{\eta_i(x_1, \dots, x_m)\},$$

wobei  $\eta_i(x_1, \dots, x_m)$  eine lineare Kombination der Werte  $x_1, \dots, x_m$  der Kovariaten  $X_1, \dots, X_m$  und den unbekannt, zu berechnenden Regressionskoeffizienten  $\beta_1, \dots, \beta_m$  des Modells ist. Die Funktion  $h_0(t)$  ist die Basis-Hazardfunktion und gibt die Wahrscheinlichkeit des Eintritts des Ereignisses in Abhängigkeit von der Zeit an, wenn alle Kovariaten null sind.

Ein optimales multivariates Modell für ein bestimmtes Ereignis enthält dann genau die (und nicht mehr) Kovariaten, die einen bedeutenden Einfluss auf die Zeit bis zum Eintritt des Ereignisses haben.

Innerhalb des SAS<sup>®</sup>/STAT Systems kann die multivariate Cox-Regression mit PROC PHREG modelliert werden [3].

### 2.3 Kovariatenselektion mit SAS<sup>®</sup> PROC PHREG

Um solche Kovariaten zu selektieren, die einen signifikanten Einfluss auf die Zeit bis zum Eintritt des interessierenden Ereignisses haben, existiert eine Reihe automatischer Methoden innerhalb der SAS<sup>®</sup>-Prozedur PROC PHREG. Dazu gehören die Backward-, Forward-, und Stepwise-Selektion [4] (auszuwählen über die Option SELECTION), siehe auch Tabelle 1. Außerdem existiert noch eine Best-Subset-Selektion in SAS<sup>®</sup>, die allerdings nicht für kategorielle Kovariaten implementiert ist.

**Tabelle 1:** Automatische Kovariatenselektion in SAS<sup>®</sup> PROC PHREG

Backward	Forward	Stepwise
<ul style="list-style-type: none"> <li>• Alle Kovariaten im Modell</li> <li>• Löschen von nicht-signifikanten Kovariaten</li> <li>• Stop, wenn alle Signifikanzen &lt; gewählter Signifikanz-Schranke</li> </ul>	<ul style="list-style-type: none"> <li>• Keine Kovariaten im Modell</li> <li>• Hinzufügen von Kovariaten nach Signifikanz</li> <li>• Stop, wenn alle übrigen Signifikanzen &gt; gewählter Signifikanz-Schranke</li> </ul>	<ul style="list-style-type: none"> <li>• Kombination von forward und backward</li> <li>• Start ohne Kovariaten</li> <li>• Stop, wenn weder hinzufügen noch löschen nach Schranke möglich</li> </ul>

Alle diese Selektionsmethoden liefern ein „optimales“ multivariates Cox-Modell, definiert über den p-Wert des Chi<sup>2</sup>-Tests der einzelnen Kovariaten. Nur solche Kovariaten verbleiben im Modell oder werden in das Modell aufgenommen, deren p-Wert kleiner als eine vom Benutzer vordefinierte Schranke ist.

Das resultierende Cox-Modell ist jedoch nicht immer notwendigerweise wirklich das optimale Modell [5]. Zum einen ist die Auswahl der Kovariaten stark abhängig von der gewählten Schranke für den p-Wert des Chi<sup>2</sup>-Tests. Zum anderen kann die Reihenfolge, mit der die Kovariaten für das Cox-Modell selektiert werden, entscheidend dafür sein, ob eine weitere Kovariate in Zusammenhang mit den schon im Modell befindlichen Kovariaten einen signifikanten p-Wert hat oder nicht, und dementsprechend in das Modell gewählt wird oder nicht. Es ist daher notwendig, die Auswahl der Kovariaten kri-

tisch zu hinterfragen und die Anpassungsgüte (goodness-of-fit) des erhaltenen Cox-Modells zu bewerten. Dabei können Maßzahlen, wie beispielsweise der C-Index, hilfreich sein [6].

### 3 Der C-Index

#### 3.1 Beschreibung

Der C-Index ist eine absolute Maßzahl um die Diskriminanz eines Cox-Modells zu bewerten. In dieser Arbeit wurde Harrells C, formalisiert von Pencina und D'Agostino, als Maßzahl benutzt [7]. Er gibt die Wahrscheinlichkeit an, dass ein Patient mit Ereignis auch eine kürzere berechnete Überlebenszeit hat als ein Patient ohne Ereignis oder ein Patient mit einem späteren Ereignis. Der C-Index nimmt Werte zwischen Null und Eins an, und je näher der berechnete C-Index bei der Eins liegt, desto besser ist die Diskriminanz des Modells.

Besonders in Studien mit Fokus auf das Langzeit-Risiko, in denen nicht nur das Eintreten des Ereignisses an sich, sondern auch die Zeit bis zum Eintreten des Ereignisses von Interesse ist, liefert Harrells C und die zugrundeliegende Definition von Diskriminanz eine geeignete Maßzahl, um das erhaltene Cox-Modell zu bewerten [8].

Dieser C-Index ist abhängig von der Zensierungsverteilung, da nur sogenannte „nutzbare“ Paare verwendet werden. Hierbei bedeutet „nutzbar“, dass einer der Patienten des Paares, die miteinander verglichen werden, ein Ereignis hatte. Es gibt zwar auch Methoden zur Berechnung des C-Index, die unabhängig von der Zensierungsverteilung arbeiten [9], jedoch sind die Unterschiede zu Harrells C bei Studien mit großer Fallzahl und wenigen Abbrechern ohne Ereignis marginal [8].

#### 3.2 Definition

Seien  $i, j$  zwei Patienten mit beobachteter Überlebenszeit  $T_i, T_j$  und berechneter Überlebenszeit  $Z_i, Z_j$ .

Gilt nun  $(T_i > T_j \wedge Z_i > Z_j) \vee (T_i < T_j \wedge Z_i < Z_j)$ , wobei „ $\vee$ “ eine logische „ODER“-Verknüpfung und „ $\wedge$ “ eine logische „UND“-Verknüpfung bezeichnet, so ist das Paar der beiden Patienten  $i$  und  $j$  konkordant.

Gilt dagegen  $(T_i > T_j \wedge Z_i < Z_j) \vee (T_i < T_j \wedge Z_i > Z_j)$ , so ist das Patientenpaar diskordant.

Der C-Index ist dann die Wahrscheinlichkeit, dass ein Patientenpaar aus der Menge aller nutzbaren Patientenpaare konkordant ist, das heißt der C-Index  $C$  ist gegeben durch

$$C = P(\text{Paare sind konkordant} | \text{Paare sind nutzbar}) \\ = P(T_i > T_j \wedge Z_i > Z_j | \text{Paare sind nutzbar}) + P(T_i < T_j \wedge Z_i < Z_j | \text{Paare sind nutzbar})$$

Anstelle der geschätzten Überlebenszeiten  $Z_i$  und  $Z_j$  können in der Formel zur Berechnung von  $C$  auch die Werte der geschätzten Überlebensfunktion an einem fixen Zeit-

punkt  $t_{fix}$  verwendet werden. Diese Überlebensfunktion lässt sich in SAS<sup>®</sup> mit Hilfe der PROC PHREG leicht berechnen.

### 3.3 Algorithmus

Für die Berechnung des C-Index wird PROC PHREG zusammen mit dem BASELINE-Befehl ausgeführt. Dieser liefert die Werte der Überlebensfunktion zu jedem Zeitpunkt  $t$  im Datensatz für die entsprechenden Kombinationen der Ausprägungen der Kovariaten. Im Gegensatz dazu liefert der OUTPUT-Befehl nur die Werte der Überlebensfunktion zu der Beobachtungszeit eines jeden Patienten, also für jeden Patienten zu unterschiedlichen Zeiten. Der OUTPUT-Befehl ist daher nicht geeignet, um die nutzbaren Paare zu bilden [10], und wird hier nur zum Zusammenfügen der Patienteninformation und der Ergebnisse aus dem BASELINE-Befehl benutzt.

Sei  $t_{fix}$  ein beliebiger fixer Zeitpunkt nach Baseline im Datensatz, „Kovariaten“ die entsprechenden Kovariaten,  $t_{event}$  die Zeit bis zum Eintritt des Ereignisses, und „Zensierung“ die Indikatorvariable für die Zensierung. Dann ist der SAS<sup>®</sup>-Code zur Berechnung des C-Index folgendermaßen gegeben.

```
DATA covals(KEEP=Kovariaten);
SET indata;
RUN;

PROC SORT DATA=covals NODUPKEY;
BY Kovariaten;
RUN;

PROC PHREG DATA=indata;
CLASS Kovariaten;
MODEL tevent * Zensierung(0)= Kovariaten / RISKLIMITS;
OUTPUT OUT=results SURVIVAL=surv;
BASELINE COVARIATES=Kovariaten OUT=outdata SURVIVAL=preds surv;
RUN;
```

Danach werden alle nutzbaren Patientenpaare gebildet und die Anzahlen der nicht-konkordanten und der konkordanten Paare bestimmt:

```

DATA predsurv;
MERGE results(IN=a) outdata(WHERE=(tevent=tfix);
BY Kovariaten;
IF a;
RUN;

DATA evtset;
SET predsurv;
IF exaccen=1;
RENAME ptno=ptno_j predsurv=y_j exac=x_j;
KEEP ptno predsurv exac;
RUN;

PROC SQL;
CREATE TABLE allset AS
SELECT ptno_j, y_j, x_j, ptno AS ptno_i, predsurv AS y_i, exac AS
x_i
FROM evtset, predsurv
WHERE ptno_j ^= ptno;
QUIT;

DATA concord;
SET allset;
IF (x_i<x_j AND y_i < y_j) OR (x_i>x_j AND y_i > y_j) THEN
concord=1;
IF (x_i < x_j AND y_i > y_j) OR (x_i > x_j AND y_i < y_j) THEN
concord=0;
RUN;

DATA _null_;
SET concord END=eof;
RETAIN nch ndh;
IF _N_=1 THEN DO;
nch=0;
ndh=0;
END;
IF concord=1 THEN nch+1;
IF concord=0 THEN ndh+1;
IF eof=1 THEN DO;
CALL SYMPUT('ch',trim(left(nch)));
CALL SYMPUT('dh',trim(left(ndh)));
CALL SYMPUT('uspairs',trim(left(_n_)));
END;
RUN;

DATA _null_;
SET predsurv END=eof;
IF eof=1 THEN CALL SYMPUT('totobs',trim(left(_n_)));
RUN;

```

Damit wird dann der C-Index letztendlich berechnet:

```
DATA cindex;
ch=INPUT("&ch",12.0);
dh=INPUT("&dh",12.0);
uspairs=INPUT("&uspairs",12.0);
totobs=INPUT("&totobs",10.0);
pc=ch/(totobs*(totobs-1));
pd=dh/(totobs*(totobs-1));
c_hat=pc/(pc+pd);
w=(2*1.96**2)/(totobs*(pc+pd));
low_ci_w=((w+2*c_hat)/(2*(1+w)))-(sqrt((w**2+4*w*c_hat*(1-
c_hat))/(2*(1+w))));
up_ci_w=((w+2*c_hat)/(2*(1+w)))+(sqrt((w**2+4*w*c_hat*(1-
c_hat))/(2*(1+w))));
RUN;
```

### 3.4 Der C-Index im Kontext multivariater Modellanpassung

Der C-Index ist eine absolute Maßzahl für die Diskriminanz des Cox-Modells, d.h. für die Wahrscheinlichkeit, dass ein Patient mit einem Ereignis eine kürzere geschätzte Überlebenszeit hat als ein Patient mit keinem oder einem späteren Ereignis. Damit ist es mit dem C-Index möglich, das durch automatische Algorithmen gefundene multivariate Cox-Modell in Bezug auf seine Diskriminanz zu bewerten, und damit auch mit anderen Modellen zu vergleichen. Der C-Index ist ein wichtiger Baustein bei der Suche nach dem optimalen Cox-Modell, am Besten in Verbindung mit weiteren Such- und Evaluierungsmethoden wie dem Vergleich der AIC-Werte oder der Analyse der Robustheit der Modellselektion durch Bootstrapping [11].

Der C-Index kann als absolute Maßzahl auch dazu verwendet werden, ein Modell an unterschiedlichen Datensätzen, z.B. durch Bootstrapping generiert, zu bewerten, um die Auswahl der Kovariaten für dieses Modell zu validieren. Da der C-Index jedoch von der Verteilung der Zensierung der Überlebenszeitdaten abhängig ist, sollten die verwendeten Datensätze ähnlich gestaltet sein.

## 4 Beispiele

### 4.1 Die POET-COPD<sup>®</sup> Studie - Prevention Of Exacerbations with Tiotropium

Die POET-COPD<sup>®</sup> Studie [12] war eine randomisierte Doppelblind-Studie in der Patienten mit chronisch-obstruktiver Lungenkrankheit (COPD), in zwei aktiven Behandlungsarmen (eine Gruppe mit Tiotropium und die andere Gruppe mit Salmeterol) für ein Jahr beobachtet wurden. Der primäre Endpunkt war die Zeit bis zur ersten Exazerbation. Für dieses Beispiel wurden 11 kategorielle Kovariaten von insgesamt 3989 Patienten betrachtet (siehe Tabelle 2):



**Tabelle 2:** Untersuchte Patientencharakteristika als mögliche Kovariaten in der POET-COPD<sup>®</sup>-Studie

Patientencharakteristikum	Ausprägungen
Anzahl Steroid-Behandlungen im Jahr vor Einschluß in die Studie	<2; ≥2
Anzahl Antibiotika-Behandlungen im Jahr vor Einschluß in die Studie	<2; ≥2
Medikation mit inhalativen Steroiden (ICS) zu Randomisierung	ja; nein
COPD Status nach GOLD*	I; II; III; IV
Dauer der Krankheit	< 6 Jahre; ≥ 6 Jahre
Geschlecht	männlich; weiblich
Alter	<55; ≥55 bis < 65; ≥65 bis < 75; ≥75
BMI	<20; ≥20 bis <25; ≥25 bis < 30; ≥30
Raucherstatus	Nichtraucher (Ex-Raucher und Nichtraucher); Raucher
Packungsjahre	<36; ≥36
Beschäftigungsstatus	Vollzeit; Teilzeit; Zuhause arbeitend; In Rente; Arbeitslos; Anderes

\*Stage I = Mild: FEV1 ≥80% pred.; Stage II = Moderate: FEV1 <80% to ≥50% pred.; Stage III = Severe: FEV1 <50% to ≥30% pred.; Stage IV = Very severe: FEV1 <30% pred.

## 4.2 Multivariate Cox-Regression mit Behandlung und Baseline-Messungen als Kovariaten

Die SAS<sup>®</sup> Prozedur PROC PHREG mit der Option SELECTION=STEPWISE wurde für eine erste Kovariatenauswahl benutzt. Als Schranken für die p-Werte zum Einschluß beziehungsweise zum Verbleib im Modell wurden folgende Werte gesetzt:

- Einschluß ins Modell: p-Wert=0.25 (in SAS<sup>®</sup>: SLENTY=0.25)
- Verbleib im Modell: p-Wert=0.15 (in SAS<sup>®</sup>: SLSTAY=0.15)

Das so erhaltene Modell enthält zusätzlich zur Kovariate „Behandlung“ neun weitere Kovariaten, die nach dem Chi<sup>2</sup>-Test und den gewählten p-Wert-Schranken signifikant für die Zeit bis zur ersten Exazerbation sind (siehe Abbildung 1). Nur die Kovariaten „Beschäftigungsstatus“ und „Raucherstatus“ sind nicht im finalen Modell enthalten.

Zusätzlich wurde die Robustheit dieses von SAS<sup>®</sup> generierten Modells überprüft [13]. Dafür wurden 1000 Replikate des Datensatzes durch Bootstrapping erstellt, und für alle erzeugten Datensätze wurde die oben genannte automatische Selektion durchgeführt. Dieselben neun Kovariaten wurden in ≥ 80% der Datensätze in das finale Modell aufgenommen, während die Kovariate Beschäftigungsstatus nur in 67% der Datensätze und der Raucherstatus nur in 26% der Datensätze im finalen Modell enthalten war (siehe Tabelle 3).

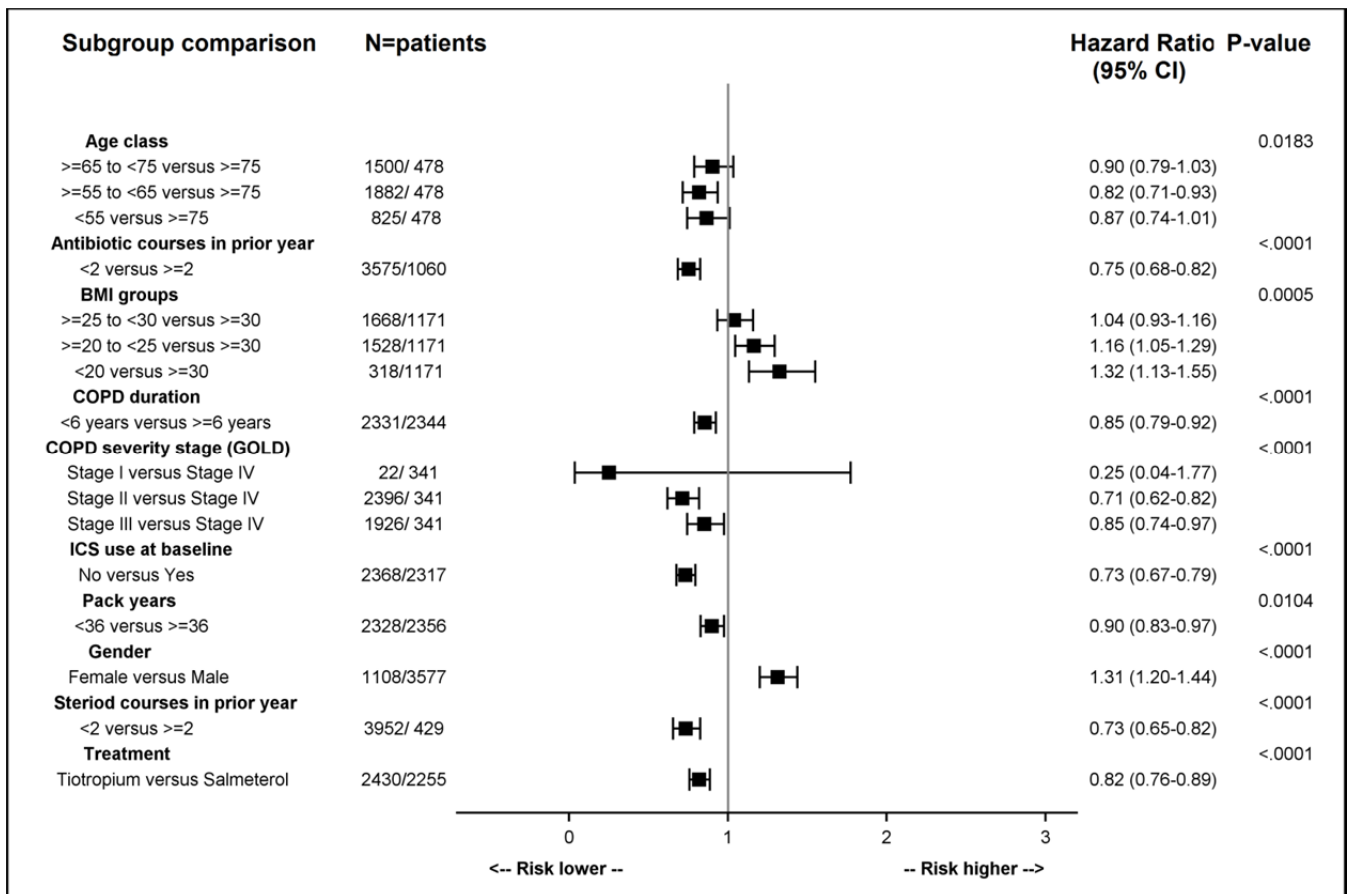


Abbildung 1: Forest plot der Ergebnisse der automatischen Kovariatenselektion

Tabelle 3: Häufigkeit der Kovariatenselektion in 1000 Bootstrapping-Datensätzen

Kovariate	Prozent [%] der Datensätze mit Kovariate im finalen Modell
Anzahl Antibiotika-Behandlungen im Jahr vor Einschluß in die Studie	100
Medikation mit inhalativen Steroiden (ICS) zu Randomisierung	100
Geschlecht	100
Anzahl Steroid-Behandlungen im Jahr vor Einschluß in die Studie	100
COPD Status nach GOLD	99.9
Dauer der Krankheit	99.4
BMI	99.2
Packungsjahre	87
Alter	79.9
Beschäftigungsstatus	67.3
Raucherstatus	26.3

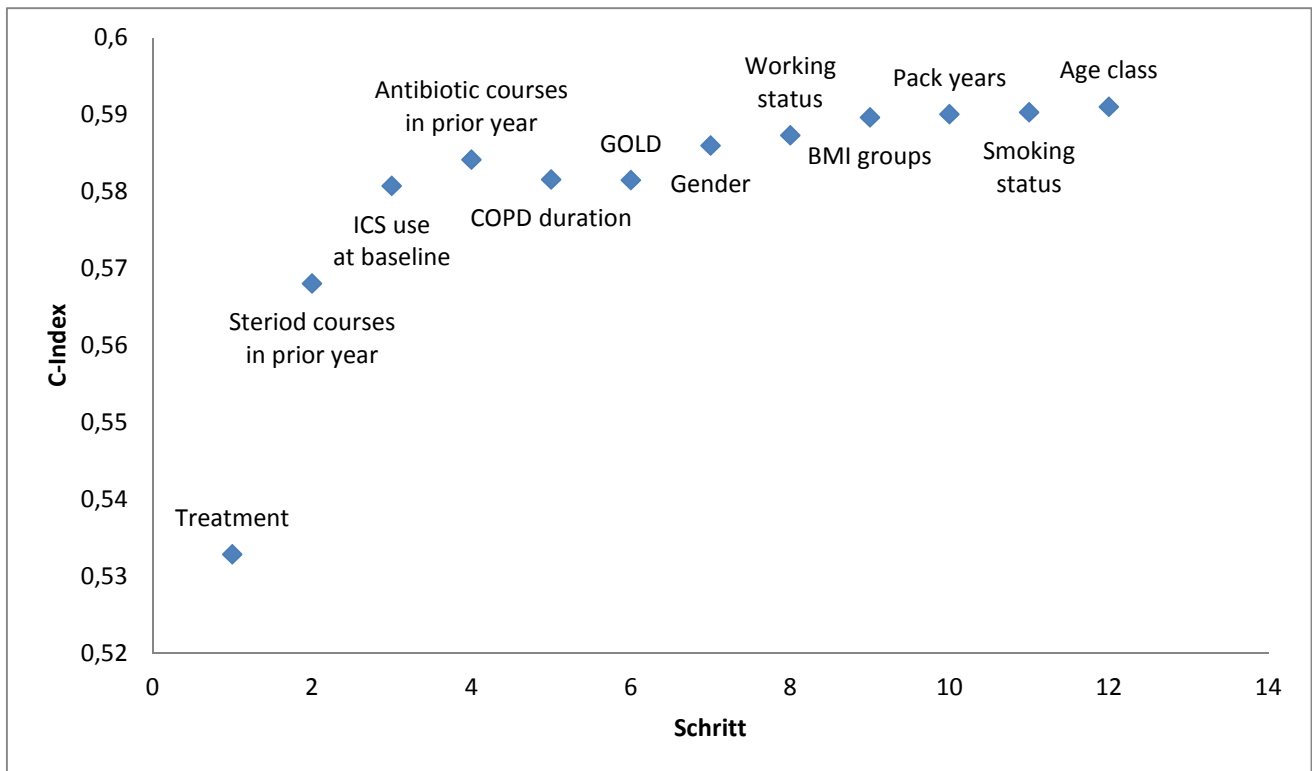
### 4.3 C-Index

Um die Anpassungsgüte des mit Hilfe der automatischen Selektionsalgorithmen in SAS<sup>®</sup> generierten multivariaten Cox-Modells zu überprüfen, wurde der C-Index für verschiedene Modelle mit unterschiedlichen Kovariaten berechnet. Zuerst wurde die Cox-Regression nur mit Behandlung und jeweils einer der Kovariaten durchgeführt und anschließend der C-Index für dieses Modell berechnet, um die Faktoren zu identifizieren, die die beste Diskriminanz liefern, siehe Tabelle 4.

Diese Analyse hat ergeben, dass die Kovariate „Anzahl Steroid-Behandlungen im Jahr vor Einschluß in die Studie“, die als Maßzahl für die Anzahl der Exazerbationen vor Studienbeginn angesehen werden kann, der am besten diskriminierende Faktor ist. Weiterhin scheinen Antibiotikabehandlungen im Jahr vor Einschluß in die Studie und Medikation mit inhalativen Steroiden (ICS) wichtig zu sein.

**Tabelle 4:** C-Index und 95% Konfidenzintervall für Modelle mit Behandlung und jeweils einer Kovariate

Kovariate	C-Index	95%CI untere Schranke	95%CI obere Schranke
Anzahl Steroid-Behandlungen im Jahr vor Einschluß in die Studie	0.57	0.52	0.62
Medikation mit inhalativen Steroiden (ICS) zu Randomisierung	0.56	0.52	0.61
Anzahl Antibiotika-Behandlungen im Jahr vor Einschluß in die Studie	0.56	0.52	0.61
Dauer der Krankheit	0.54	0.50	0.59
COPD Status nach GOLD	0.54	0.50	0.59
Geschlecht	0.54	0.50	0.59
Beschäftigungsstatus	0.54	0.50	0.58
BMI	0.53	0.49	0.58
Packungsjahre	0.53	0.49	0.57
Raucherstatus	0.53	0.49	0.57
Alter	0.53	0.49	0.57



**Abbildung 2:** C-Indices für Cox-Modelle mit sequentiell hinzugefügten Kovariaten in der Reihenfolge der C-Indices (absteigend) der Einzelanalyse

In einem zweiten Schritt wurden danach die Kovariaten in absteigender Reihenfolge der C-Indices der Modelle in Tabelle 4 nacheinander dem Cox-Modell zusätzlich zur Behandlung als Effekt hinzugefügt. Das erste Cox-Modell enthielt demnach nur Behandlung und Steroide im Jahr vor der Studie als Kovariaten, das zweite Cox-Modell zusätzlich noch ICS-Gaben, usw. Dann wurden die C-Indices der multivariaten Modelle mit den jeweiligen Kovariaten sowie des univariaten Cox-Modells nur mit Behandlung als Kovariate berechnet.

Abbildung 2 zeigt die C-Indices der so generierten Modelle. Die Punkte bilden eine virtuelle Kurve, die deutlich zeigt, dass der C-Index bei Hinzunahme weiterer Kovariaten ins Modell erst ansteigt und dann ein Plateau erreicht. Die Hinzunahme weiterer Kovariaten verbessert demnach die Anpassungsgüte des Modells in Bezug auf seine Diskriminanz nicht weiter.

Die Modellanpassung mit Hilfe des C-Index schlägt somit ein Modell mit Behandlung plus drei weiteren Kovariaten, nämlich „Anzahl Steroid-Behandlungen im Jahr vor Einschluß in die Studie“, „Anzahl Antibiotika-Behandlungen im Jahr vor Einschluß in die Studie“ und „Medikation mit inhalativen Steroiden (ICS) zu Randomisierung“ vor, im Gegensatz zu neun Kovariaten aus der automatischen Stepwise-Selektion mit SAS®.

Alle diese drei Kovariaten können als stark korreliert mit der Anzahl an Exazerbationen vor der Behandlung und der Schwere der COPD-Erkrankung angesehen werden. Somit

liefert die auf statistischen Modellen beruhende Modellwahl auch ein klinisch sinnvolles Modell für die signifikanten Faktoren, eine (weitere) Exazerbation zu bekommen.

## 5 Zusammenfassung und Ausblick

Multivariate Analysen und die dazugehörige Selektion wichtiger prognostischer oder prädiktiver Faktoren als Kovariaten in Regressionsmodellen werden angesichts des Trends zu personalisierter Medizin in der klinischen Entwicklung und der zunehmend frühen Nutzenbewertung von Arzneimitteln immer wichtiger.

Für Überlebenszeitdaten bietet SAS<sup>®</sup>/STAT mit PROC PHREG eine leicht anwendbare Möglichkeit an, eine multivariate Cox-Regression zu implementieren. In SAS<sup>®</sup>/STAT eingebaute automatische Algorithmen wie die „Stepwise-“, „Backward-“ oder „Forward-“ Selektion sind eine einfache Möglichkeit, wichtige Kovariaten aus der großen Anzahl möglicher Patientencharakteristika herauszufiltern.

Diese Methoden basieren die Selektion der Kovariaten auf einer vom Benutzer gesetzten Schranke für den p-Wert des Chi<sup>2</sup>-Tests. Einfügen beziehungsweise Löschen von Kovariaten geschieht in der Reihenfolge der Signifikanzen. Mit diesen automatischen Methoden kann jedoch nicht immer ein optimales Modell für das Risiko eines interessierenden Ereignisses gefunden werden, also ein Modell, welches so wenige und so viele Kovariaten wie nötig enthält.

Um das Ergebnis der automatischen Selektion mit SAS<sup>®</sup> zu bewerten, braucht man daher Maßzahlen für die Anpassungsgüte und Validierungsmethoden für die Robustheit der Anpassung. Eine mögliche Maßzahl für die Diskriminanz des erhaltenen Cox-Modells ist der C-Index.

Der in dieser Arbeit vorgestellte C-Index ist leicht mit SAS<sup>®</sup> zu berechnen. Der Algorithmus zu seiner Berechnung ist nicht unabhängig von der Zensierungsverteilung, jedoch gibt es Untersuchungen, die zeigen, dass bei großen Studien mit wenigen Studienabbruchern nur ein geringer Unterschied zu verteilungsfreien Berechnungsmethoden besteht.

Am Beispiel einer Atemwegsstudie mit COPD-Patienten wurde die automatische Selektion in SAS<sup>®</sup> für die Zeit bis zur ersten Exazerbation dargestellt und die Berechnung des C-Index zur Bewertung des erhaltenen Cox-Modells demonstriert. Die Stepwise-Selektion in SAS<sup>®</sup> führte zu einem Modell mit neun Kovariaten, zusätzlich zu der Behandlung. Der C-Index jedoch weist darauf hin, dass bereits ein Modell mit deutlich weniger (nämlich drei) Kovariaten (und Behandlung) ausreicht, dieselbe Anpassungsgüte zu erreichen. Die drei mit Hilfe des C-Index ausgewählten Kovariaten stehen alle im Zusammenhang mit der Anzahl an Exazerbationen vor der Behandlung und der Schwere der Erkrankung, und sind somit auch klinisch plausible Einflussfaktoren für die Zeit bis zur Exazerbation.

Das Beispiel zeigt, dass die Ergebnisse einer automatischen Kovariatenselektion nicht in jedem Fall unkritisch übernommen werden sollten, sondern dass jede Modellselektion immer auch mit zusätzlichen Maßzahlen und Methoden hinterfragt und auf Plausibilität sowie Robustheit geprüft werden sollte. Der C-Index ist hierzu eine gute Möglichkeit.

Zusätzlich zum C-Index stehen noch weitere Methoden zur Verfügung, wie die bekannte relative Maßzahl Akaike's Information Criterion (AIC) für die Anpassungsgüte, die von SAS<sup>®</sup> bei Anwendung von PROC PHREG standardmäßig mit ausgegeben wird. Auch kann die Robustheit der Modellselektion überprüft werden, indem man Replikat des Datensatzes durch Bootstrapping erstellt und dann die automatische Modellselektion mit diesen replizierten Datensätzen durchführt. Die Häufigkeit, mit der eine Patientencharakteristik als Kovariate in das finale Cox'sche Regressionsmodell gewählt wird, ist eine Maßzahl für die Robustheit der Faktoren.

Eine weitere Möglichkeit ein robustes Cox-Modell zu wählen ist die Anwendung von Selektionsverfahren, die nicht standardmäßig von PROC PHREG angeboten werden. Beispiele hierzu sind genetische Algorithmen deren Gütekriterium zur Auswahl eines „optimalen“ Modells frei gewählt werden kann, wie zum Beispiel AIC oder der C-Index [14].

## Literatur

- [1] Breslow, NE. Covariance analysis of censored survival data. *Biometrics* 1974; 30:89-100
- [2] Collet D. *Modelling Survival Data in Medical Research*. 2nd edition. Chapman&Hall, 2003
- [3] SAS<sup>®</sup> Institute Inc. *SAS/STAT<sup>®</sup> 9.2 User's Guide*. 2nd edition. SAS<sup>®</sup> Institute, Cary, NC, 2009
- [4] Steyerberg EW. *Clinical Prediction Models: A practical Approach to Development, Validation and Updating*. Springer, 2010
- [5] Wiegand RE. Performance of using multiple stepwise algorithms for variable selection. *Statist Med* 2010; 29:1647-1659
- [6] Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, Pencina MJ and Kattan MW. Assessing the Performance of Prediction Models – A Framework for Traditional and Novel Measures. *Epidemiology* 2010; 21:128-138
- [7] Pencina MJ, D'Agostino RB. Overall C as a measure of discrimination in survival analysis: model specific population value and confidence interval estimation. *Statist Med* 2004; 23:2109-2123
- [8] Pencina MJ, D'Agostino RB, Song L. Quantifying discrimination of Framingham risk functions with different survival C statistics. *Statist Med* 2012; 31:1543-1553

- [9] Uno H, Cai T, Pencina MJ, D'Agostino RB and Wei LJ. On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. *Statist Med* 2011; 30:1105-1117
- [10] Liu L, Forman S, Barton B. Fitting Cox Model Using PROC PHREG and Beyond in SAS<sup>®</sup>. *Proceedings of SAS<sup>®</sup> Global Forum 2009*, Paper 236
- [11] Efron B and Tibshirani, R J. *An Introduction to the Bootstrap*. Chapman&Hall, 2003
- [12] Vogelmeier C, Hederer B, Glaab T, Schmidt H, Rutten-van Moelken MPMH, Beeh KM, Rabe KF, Fabbri LM, POET-COPD Investigators. Tiotropium versus Salmeterol for the Prevention of Exacerbations of COPD. *N Engl J Med* 2011; 364:1093-1103
- [13] Cassell D. Don't Be Loopy: Re-Sampling and Simulation the SAS<sup>®</sup> Way. *Proceedings SAS Global Forum 2007*, Paper 183.
- [14] Tsai JS. Optimal Model Selection By a Genetic Algorithm Using SAS<sup>®</sup>. *Western Users of SAS<sup>®</sup> Software Conference, San Jose, CA, 2009*