

Diplomarbeit

**Validation of EuroSCORE II risk model for aortic  
valve replacement**

**Ist der neue EuroSCORE II im Vergleich zum bereits verwendeten  
EuroSCORE das probate Mittel zur Qualitätssicherung bei  
prothetischen Aortenklappenersatz in der Herzchirurgie?**

eingereicht von

**Yannik Snieder**

zur Erlangung des akademischen Grades

**Doktor der gesamten Heilkunde  
(Dr. med. univ.)**

an der

**Medizinischen Universität Graz**

ausgeführt an der

**Universitätsklinik für Chirurgie, klinische Abteilung für  
Herzchirurgie**

unter der Anleitung von

Univ.-Prof. Dr. Igor Knez

Graz, 13.02.2018 .....

## **Eidesstattliche Erklärung**

*Ich erkläre ehrenwörtlich, dass ich die vorliegende Arbeit selbstständig und ohne fremde Hilfe verfasst habe, andere als die angegebenen Quellen nicht verwendet habe und die den benutzten Quellen wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.*

*Graz, am 13.02.2018*

*Yannik Snieder eh*

## Danksagungen

An dieser Stelle möchte ich mich bei den Personen bedanken, die mich während der Anfertigung dieser Arbeit unterstützt und mir dabei geholfen haben, ohne sie wäre diese Arbeit nicht zustande gekommen.

Ganz besonders gilt mein Dank Herrn Univ.-Prof. Dr. Igor Knez für die Betreuung meiner Diplomarbeit. Seine Anleitung und Hilfe waren eine große Bereicherung. Bei jedem Gespräch, auch bei kurzen Telefonaten, ging er auf Probleme ein und half mit Ratschlägen und Lösungen.

Ich möchte mich auch bei Dr. Alissa Strugger für die Auseinandersetzung mit meiner Arbeit bedanken. Ihre Hilfestellung bei jeglichen Problemen war unbezahlbar, insbesondere bei der statistischen Auswertung.

Ebenfalls möchte ich mich bei meinem Bruder Laurin Snieder und meinem Vater Mag. Karl Snieder bedanken, welche zahlreiche Stunden Korrektur gelesen haben und durch deren Mühen meine Arbeit kontinuierlich verbessert wurde.

Außerdem gilt mein Dank meinen Freunden und Freundinnen, die mich in diesen 6 Jahren über begleitet haben und die Zeit mit zahlreichen schönen Erinnerungen gefüllt haben.

Abschließend möchte ich bei meiner Familie bedanken, die mich mein ganzes Studium lang unterstützt hat.

# Zusammenfassung

## Hintergrund:

Der EuroSCORE wurde entwickelt um das individuelle Risiko von Patienten und Patientinnen zu berechnen an einem herzchirurgischen Eingriff oder bis zu 30 Tage danach zu sterben. Jahre nach der Entwicklung erfolgte eine Überarbeitung des EuroSCORE um eine Anpassung an die veränderten Risikofaktoren und Techniken vorzunehmen. Diese Überarbeitung ist der EuroSCORE II, der in dieser Studie im Vergleich zu seinem Vorgänger validiert werden soll.

## Methoden:

Die Daten von 843 Patienten und Patientinnen, welche einen Aortenklappenersatz an der Abteilung für Herzchirurgie des Universitätsklinikum Graz zwischen 2012 und 2016 erhielten, wurden dazu analysiert. ROC-Analysen zur 30-Tages-Mortalität wurden erstellt und die AUC verglichen um die Diskriminationskraft zu beurteilen. VLADs wurden erstellt und performance ratios berechnet um die Kalibrierung der Scores zu vergleichen.

## Ergebnisse:

Für den Zeitraum von 2012 bis 2016 betrug die 30-Tages-Mortalität 2,1%. Der durchschnittliche logistische EuroSCORE und der EuroSCORE II betragen jeweils 9,07% und 3,64%, mit einer performance ratio von 0,23 und 0,58. Die AUC war für den logistischen EuroSCORE 0,847 (0,761-0,933 95% KI) und für den EuroSCORE II 0,829 (0,733-0,926 95% KI).

## Fazit:

Beide Scores zeigten eine ähnliche und außerordentlich gute Diskriminationskraft, die für die 30-Tages-Mortalität bei Aortenklappenoperationen signifikant war. Der EuroSCORE II zeigte jedoch eine bessere Kalibrierung als der logistische EuroSCORE, obwohl beide weit von einer performance ratio von 1 entfernt waren. Der EuroSCORE II, als Verbesserung des logistischen EuroSCORE, stellt einen guten Prädiktor der 30-Tages-Mortalität bei Aortenklappenoperation dar und kann in sämtlichen Situationen verwendet werden.

# **Abstract**

## **Background:**

The EuroSCORE was developed to calculate the individual risk of patients undergoing cardiac surgery to die within a period of 30 days after the surgery. Years after the development the EuroSCORE was revised to adapt to the changed risk factors and techniques. This revision is the EuroSCORE II which will be compared to its predecessor for validation in this study.

## **Methods:**

The data of 843 patients who received an aortic valve replacement at the Department of Cardiac Surgery of the University Hospital Graz between 2012 and 2016 were analyzed. In regard to 30-day mortality ROC analysis was performed and the AUC was compared to assess discrimination. VLADs were created and the performance ratios were calculated to compare the calibration of the scores

## **Results:**

For the period from 2012 to 2016 the observed 30-day mortality was 2.1%. The mean logistic EuroSCORE and EuroSCORE II were 9.07% and 3.64%, respectively, with a performance ratio of 0.23 and 0.58. The AUC for the logistic EuroSCORE and the EuroSCORE II was 0.847 (0.761-0.933 95% CI) and 0.829 (0.733-0.926 95% CI).

## **Conclusion:**

Both scores showed a similar and exceptionally good discrimination which was significant for the 30-day mortality in aortic valve surgery. The EuroSCORE II, however, showed a better calibration than the logistic EuroSCORE. Both were, nevertheless, far from a performance ratio of 1. The EuroSCORE II as an improvement of the logistic EuroSCORE is a good predictor of 30-day mortality in aortic valve surgery and can be used in all situations.

# Inhaltsverzeichnis

Eidesstattliche Erklärung .....	I
Danksagungen .....	II
Zusammenfassung .....	III
Abstract .....	IV
Inhaltsverzeichnis .....	V
Glossar und Abkürzungen .....	VII
Abbildungsverzeichnis .....	VIII
Tabellenverzeichnis .....	IX
1 Einführung .....	1
1.1 Qualitätsmanagement .....	1
1.1.1 Definition von Qualität und Qualitätsmanagement .....	1
1.1.2 Qualitätsmanagement in der Medizin .....	2
1.1.3 EuroSCORE .....	3
1.1.4 Die Entwicklung des EuroSCORE .....	4
1.1.5 Berechnung des EuroSCORE .....	5
1.2 Aortenklappen Chirurgie .....	13
1.2.1 Grundlagen .....	13
1.2.2 Aortenklappenstenose .....	13
1.2.3 Aortenklappeninsuffizienz .....	17
1.2.4 Mechanische oder biologische Klappe .....	21
1.2.5 Herzinsuffizienz .....	22
1.2.6 Chirurgisches Verfahren: .....	23
1.2.7 Interventionelles Verfahren .....	23
1.2.8 Risiken und Komplikationen: .....	24
1.2.9 Ziele und Prognose: .....	24
2 Methoden .....	25

2.1	Statistische Methoden.....	25
2.1.1	ROC- Kurven.....	25
2.1.1.1	Grundlagen.....	25
2.1.1.2	Vergleich von ROC-Kurven .....	30
2.1.1.3	Indices zur diagnostischen Genauigkeit von ROC-Kurven .....	33
2.1.1.4	Generalisierte ROC-Analyse .....	35
2.1.1.5	Kurvenanpassung.....	39
2.1.1.6	Statistische Signifikanz und Abweichungen.....	41
2.1.1.7	Idealer Grenzwert und Youden-Index.....	43
2.1.2	Variable life-adjusted display (VLAD) und Leistungsverhältnis .....	48
2.1.3	Datenreview, Zulassung und verwendete statistische Methoden.....	51
3	Ergebnisse .....	52
3.1	Demographische Daten.....	52
3.2	ROC-Analysen: logistischer EuroSCORE und EuroSCORE II für den Zeitraum 2012 bis 2016 .....	56
3.3	Variable life-adjusted display (VLAD) und performance ratio .....	65
3.3.1	Variable life-adjusted display (VLAD) und performance ratio für den logistischen EuroSCORE (2012-2016).....	65
3.3.2	Variable life-adjusted display (VLAD) und performance ratio für den EuroSCORE II (2012-2016).....	66
4	Diskussion .....	67
4.1	Allgemeines.....	67
4.2	Scoring.....	68
4.3	ROC-Analysen und performance ratio .....	73
4.4	Zusammenfassung.....	75
5	Literatur .....	77

## Glossar und Abkürzungen

<b>ACEF</b>	Age, Creatinin, Ejection Fraction	<b>TAVI</b>	transcatheter aortic valve replacement
<b>AUC</b>	Area Under the Curve	<b>TGFBR</b>	Transforming Growth Factor Beta Receptor
<b>CABG</b>	koronarerterieller Bypass Graft	<b>TGFBR</b>	Transforming Growth Factor Beta Receptor
<b>KÖF</b>	Klappenöffnungsfläche	<b>VLAD</b>	Variable Life-Adjusted Display
<b>LVEDD</b>	linksventrikulärer enddiastolischer Durchmesser	<b>V<sub>max</sub></b>	maximale transvalvuläre Flussgeschwindigkeit
<b>LVEF</b>	linksventrikuläre Ejektionsfraktion	<b>ΔP</b>	Druckgradient
<b>LVESD</b>	linksventrikulärer endsystolischer Durchmesser		
<b>MSCT</b>	Multislice Computertomographie		
<b>NYHA</b>	New York Heart Association		
<b>ROC</b>	Receiver Operating Characteristic		
<b>SAVR</b>	surgical aortic valve replacement		
<b>STS</b>	Society of Thoracic Surgeons		
<b>SV<sub>i</sub></b>	Schlag Volumen Index		

## Abbildungsverzeichnis

Abbildung 1: Entscheidungsachse eines Tests .....	28
Abbildung 2: ROC-Kurve .....	29
Abbildung 4: ROC-Kurve im Vergleich Test A und Test B .....	33
Abbildung 5: LROC-Kurve .....	38
Abbildung 6: Empirische ROC-Kurve .....	40
Abbildung 7: „cut-off“ Punkt $d^2$ .....	44
Abbildung 8: Youden-Index .....	48
Abbildung 9: VLAD .....	50
Abbildung 10: ROC-Kurve logistischer EuroSCORE 2012-2016 .....	58
Abbildung 11: ROC-Kurve EuroSCORE II 2012-2016 .....	58
Abbildung 12: ROC-Kurve logistischer EuroSCORE 2012 .....	60
Abbildung 13: ROC-Kurve EuroSCORE II 2012 .....	60
Abbildung 14: ROC-Kurve logistischer EuroSCORE 2013 .....	61
Abbildung 15: ROC-Kurve Euroscore II 2013 .....	61
Abbildung 16: ROC-Kurve logistischer EuroSCORE 2014 .....	62
Abbildung 17: ROC-Kurve EuroSCORE II 2014 .....	62
Abbildung 18: ROC-Kurve logistischer EuroSCORE 2015 .....	63
Abbildung 19: ROC-Kurve EuroSCORE II 2015 .....	63
Abbildung 20: ROC-Kurve logistischer EuroSCORE 2016 .....	64
Abbildung 21: ROC-Kurve EuroSCORE II 2016 .....	64
Abbildung 22: VLAD für den logistischen EuroSCORE 2012-2016 .....	66
Abbildung 23: VLAD für den EuroSCORE II 2012-2016 .....	67

## Tabellenverzeichnis

Tabelle 1: Die patientenbezogenen Risikofaktoren und deren spezifischer Koeffizient $\beta_i$ im logistischen EuroSCORE .....	8
Tabelle 2: Die herzbezogenen Risikofaktoren und deren spezifischer Koeffizient $\beta_i$ im logistischen EuroSCORE.....	9
Tabelle 3: Die operationsbedingten Risikofaktoren und deren spezifischer Koeffizient $\beta_i$ im logistischen EuroSCORE .....	9
Tabelle 4: Die patientenbezogenen Risikofaktoren und deren spezifischer Koeffizient $\beta_i$ im EuroSCORE II.....	11
Tabelle 5: Die herzbezogenen Risikofaktoren und deren spezifischer Koeffizient $\beta_i$ im EuroSCORE II .....	12
Tabelle 6: Die operationsbedingten Risikofaktoren und deren spezifischer Koeffizient $\beta_i$ im EuroSCORE II.....	12
Tabelle 7: Stadieneinteilung der Herzinsuffizienz nach NYHA .....	22
Tabelle 8: ABCD-Gruppen der Herzinsuffizienz nach AHA .....	23
Tabelle 9: Konfusionsmatrix .....	27
Tabelle 10: Konfusionsmatrix Test A .....	32
Tabelle 11: Konfusionsmatrix Test B .....	32
Tabelle 12: Risikofaktoren I .....	52
Tabelle 13: Risikofaktoren II .....	53
Tabelle 14: Status und 30-Tage-Status .....	54
Tabelle 15: Alter und Scores .....	54
Tabelle 16: 30-Tages-Status nach OP-Jahr .....	54
Tabelle 17: Alter und Scores nach OP-Jahr .....	55
Tabelle 18: AUC-Werte der Scores .....	56
Tabelle 19: Youden-Index der Scores .....	56
Tabelle 20: d-Wert der Scores.....	57
Tabelle 21: AUC-Werte nach OP-Jahr.....	59
Tabelle 22: Übersicht 30-Tages-Status .....	65
Tabelle 23: Performance ratio des logistischen EuroSCORE.....	65
Tabelle 24: Performance ratio für den EuroSCORE II .....	66

# **1 Einführung**

## **1.1 Qualitätsmanagement**

Durch die sich immer schneller ändernden Bedingungen und Möglichkeiten hat die Wichtigkeit des Qualitätsmanagements mit der Zeit zugenommen. Um die Qualität von Dienstleistungen und Produkten vergleichen zu können, werden einfache Parameter bis hin zu komplizierten Verfahren verwendet. Es werden außerdem laufend neue Parameter und Verfahren gesucht und weiterentwickelt. Nur so kann festgestellt werden, welche Produkte und Dienstleistungen den Ansprüchen entsprechen. Im medizinischen Bereich ist der wichtigste Parameter in erster Linie die Mortalität zu der aber noch weitere Parameter hinzukommen. Dazu gehören unter anderem die Dauer von Krankenhausaufenthalten, die Remobilisation, die Rehabilitation und die Fortsetzung des Berufslebens von Patienten.

Aufgrund des wachsenden Leistungs- und Kostendrucks herrscht zwischen den Krankenanstalten eine Konkurrenzsituation. Der Erfolg von Behandlungen und deren Qualität in den verschiedenen Anstalten kann verglichen werden. Daraus entsteht ein Optimierungsbedarf der Strukturierung und Leistungserbringung. Medizinisches Qualitätsmanagement als Teil der Klinikleitung ist darum unerlässlich. Die Zielsetzungen liegen in der Implementierung und dem Umsetzen von Qualitätssystemen zur Verbesserung der Sicherheit sowie der Steigerung von Effektivität und Effizienz.

### **1.1.1 Definition von Qualität und Qualitätsmanagement**

Qualität ist die Übereinstimmung von Leistung und Erwartungen von Kunden, die ein Produkt oder eine Dienstleistung in Anspruch nehmen. Dies ist abhängig von den Anforderungen des Kunden und kann teilweise gemessen werden, wie die Sachqualität. Andere Teile sind nicht so einfach messbar wie die Anmutsqualität. [1]

Qualitätsmanagement dient zur Sicherstellung der Qualität, also dem Übereinstimmen der Anforderungen von Kunden an Produkte und Dienstleistungen mit den erbrachten Leistungen und erzeugten Produkten. Es soll eine Vertrauensbasis für alle beteiligten Personen schaffen. Die Aufgaben des Qualitätsmanagements bestehen aus Qualitätspolitik, Qualitätslenkung, Qualitätsplanung, Qualitätsprüfung und Qualitätsverbesserung. Diese Aufgaben werden in den Normen DIN/ISO 9001:2000 geregelt. Gutes Qualitätsmanagement stellt somit lediglich fest ob die gewünschte Qualität erreicht wird, unabhängig der Höhe der Qualität. [2]

### **1.1.2 Qualitätsmanagement in der Medizin**

Qualitätsmanagement im Gesundheitswesen zur Sicherung und Weiterentwicklung der Qualität der erbrachten Leistungen ist gesetzlich vorgeschrieben. Die Qualität muss dem jeweiligen aktuellen Stand der wissenschaftlichen Erkenntnisse entsprechen. [2]

In der Medizin kann Qualität in drei Bereiche unterteilt werden. Die Strukturqualität beschreibt Strukturmerkmale wie die Qualifikationen von Ärzten, die Ausstattung usw. Die Prozessqualität gibt Bescheid über die Qualität der medizinischen Versorgung, wie Diagnostik, Behandlungsabläufe und die interpersonelle Kommunikation. Die Ergebnisqualität gibt Auskunft über die Qualität des Behandlungserfolges z.B. durch die Mortalitätsrate. Diese drei Bereiche sind im Gegensatz zur subjektiven Qualität, die ebenfalls eine wichtige Rolle einnimmt, objektivierbar. Die subjektive Qualität umfasst die individuellen Empfindungen der Patienten wie den Umfang des Services, die Freundlichkeit des Personals, die Länge der Wartezeiten und vieles mehr. Mit ihr wird maßgeblich bestimmt ob Patienten und Patientinnen mit einem Krankenhausaufenthalt zufrieden sind und das Krankenhaus anderen vorziehen oder dieses weiterempfehlen. [3]

Alle unterschiedlichen Bereiche müssen gute Qualität aufweisen, damit die Gesamtqualität als gut erachtet werden kann. Das soll Qualitätsmanagement sicherstellen, indem es in die jeweiligen Bereiche eingreift und Verbesserungen

vornimmt. Somit soll die Zufriedenheit aller beteiligten Personen gesteigert werden, was nicht nur Patienten und Patientinnen und deren Angehörige sondern auch die eigenen Mitarbeiter und Mitarbeiterinnen betrifft.

Ein wichtiger Bestandteil des Qualitätsmanagements ist die statistische Prozessregelung, sie dient zur Verbesserung von Produkten und Dienstleistungen. Im medizinischen Bereich ist sie dafür zuständig, dass eine Anpassung der Bewertung an das individuelle Risiko von Patienten und Patientinnen erfolgt. Denn ein Vergleich verschiedener Krankenanstalten, Abteilungen oder Ärzten und Ärztinnen könnte ansonsten ungerecht ausfallen. Die Qualität ist demnach nicht nur von den Ergebnissen wie der Mortalität und Morbidität, sondern auch von der Ausgangssituation abhängig. Unterschiedliche Risikofaktoren wie das Patientenalter spielen dabei eine entscheidende Rolle. Wenn dadurch ein höheres Risiko zu sterben besteht, muss dieses berücksichtigt werden. Eine Anpassung an das individuelle Risiko kann mittels variable life-adjusted display (VLAD) erfolgen. [4]

### **1.1.3 EuroSCORE**

EuroSCORE steht für European System for Cardiac Operative Risk Evaluation und wird zum Voraussagen der Mortalität bei Herzoperationen verwendet. Es werden sämtliche Risikofaktoren, die als wichtig erachtet werden, berücksichtigt um die 30-Tages-Mortalität vorauszusagen. Diese ist das wichtigste Qualitätsmerkmal in der Herzchirurgie und gibt an, ob ein Patient oder eine Patientin in den ersten 30 Tagen nach einem herzchirurgischen Eingriff stirbt oder überlebt. Diese Bewertung trifft auf alle herzchirurgischen Eingriffe, von einfachen bis zu äußerst komplexen Operationen, die anhand der Operationsart schon ein unterschiedliches Risiko aufweisen, zu. Dazu kommen die Risikofaktoren des Patienten oder der Patientin wie das Alter, das Geschlecht und der Gesamtzustand. Die herzspezifischen Risikofaktoren wie die Ventrikelfunktion oder das Vorhandensein von pulmonalem Hypertonus bilden eine eigene Gruppe. Ohne eine entsprechende Anpassung an diese Risikofaktoren kann die Qualität von Operationen nicht objektiv beurteilt werden. Wenn die 30-Tages-Mortalität allein zum Vergleichen genutzt wird und keine Anpassung erfolgt, werden somit Chirurgen oder Chirurginnen, Abteilungen oder

Krankenhäuser, welche mehr Patienten und Patientinnen mit höherem Risiko operieren, schlechter bewertet. Der EuroSCORE soll dies durch Voraussagen der Sterbewahrscheinlichkeit unter Einbeziehung der wichtigsten Risikofaktoren verhindern. Damit kann die risikobezogene Mortalität als gutes Qualitätsmerkmal genutzt werden. [5, 6]

#### **1.1.4 Die Entwicklung des EuroSCORE**

Um Patienten und Patientinnen auf Operationen vorzubereiten, um festzustellen ob eine Operation sinnvoll und nicht zu riskant ist und die Qualität der Operation bewerten zu können, mussten Systeme entwickelt werden. Die Mortalitätsrate, bei den sich ständig ändernden Risikoprofilen von Patienten und Patientinnen, reichte dafür nicht aus. Deswegen mussten Wege gefunden werden um eine objektive Bewertung zu ermöglichen. Dies erfolgte durch die risikobezogene Mortalität, die mit verschiedenen Methoden berechnet wird.

Es wurden viele Systeme entwickelt um diese Aufgaben zu erfüllen, es konnten sich jedoch nur wenige durchsetzen. Die am weitesten verbreiteten Systeme sind der STS (Society of Thoracic Surgeons) Score, der in Nordamerika und Kanada verwendet wird und der EuroSCORE, der in Europa verwendet wird. Zur Erstellung des EuroSCORE wurden die Daten aus 128 verschiedenen Zentren, aus acht europäischen Ländern gesammelt und analysiert. Bei fast zwanzigtausend Patienten und Patientinnen wurden aus 97 Risikofaktoren die Wichtigsten, Verlässlichsten und Objektivsten herausgenommen um ein Bewertungssystem zu erstellen. Dieses System wurde dann mit einem Teil der gesammelten Daten überprüft, validiert und 1998 präsentiert.

Des Weiteren wurde das System in verschiedenen Studien an unterschiedlichen Orten und Bevölkerungsgruppen getestet und konnte gute prädiktive Werte erzielen. Jedoch gab es Kritik, unter anderem dass das Mortalitätsrisiko überschätzt wird. Außerdem fanden sich schlechte prädiktive Werte bei Hochrisikopatienten und

Patientinnen für Aortenklappenoperationen. Im Laufe der Zeit und durch ständig neue medizinische Erkenntnisse veraltete der EuroSCORE. Es erfolgte eine Überarbeitung und schließlich wurde 2011 der EuroSCORE II veröffentlicht. Dieser soll eine Verbesserung gegenüber dem Vorgänger sein und keine Mängel mehr aufweisen, indem er neue Erkenntnisse und Fortschritte berücksichtigt. Er wurde auf dieselbe Weise wie sein Vorgänger konstruiert allerdings mit einem neuen Datensatz und teilweise neuen oder veränderten Risikofaktoren. Die Daten stammen von über zwanzigtausend Patienten und Patientinnen aus 154 Zentren in 43 Ländern aus einem Zeitraum von drei Monaten. [5–8]

### **1.1.5 Berechnung des EuroSCORE**

Der EuroScore wird aus drei verschiedenen Gruppen von Faktoren gebildet, den patientenbezogenen Faktoren, den herzbezogenen Faktoren und den operationsbedingten Faktoren. Es gibt die additive Berechnung die sehr einfach ist und die logistische Berechnung. Wie bereits erwähnt gibt es seit 2011 eine neue Version, den EuroSCORE II, der sich durch ein paar Faktoren vom Vorgänger unterscheidet und ebenso wie der logistische EuroSCORE mit einem logistischen Regressionsmodell errechnet wird.

Beim additiven EuroSCORE bekommen alle Faktoren je nach Vorhandensein einen Wert zugeordnet, welche dann zusammengezählt werden. Dies ist einfach anwendbar und kann am Patientenbett, ohne weitere technische Hilfsmittel, durchgeführt werden. Er basiert auf dem logistischen Modell und sollte möglichst viele Ärzte und Ärztinnen animieren ihn für das Qualitätsmanagement anzuwenden. Die Summe der zusammengezählten Werte ist dann der additive oder numerische EuroSCORE und gibt die Sterbewahrscheinlichkeit an. [9, 10]

Nachfolgend werden die patientenbezogenen Faktoren des additiven EuroSCORE aufgezählt. Für das Alter wird dabei ab 60 Jahren 1 Punkt vergeben und alle 5 Jahre kommt noch ein Punkt dazu. Das Geschlecht, weiblich bekommt 1 Punkt männlich 0.

Für COPD mit Langzeitanwendung von Bronchodilatoren oder Steroiden wird 1 Punkt vergeben. 2 Punkte gibt es für Extrakardiale Arterienerkrankung bei Vorliegen von Claudicatio, einem Arteria carotis interna Verschluss oder einer über 50% Stenose oder schon durchgeführten oder indizierten Operationen der Aorta, den Carotiden oder den Arterien der unteren Extremität. Neurologische Erkrankungen mit Einschränkungen der Fortbewegung oder des Alltages zählen auch 2 Punkte. Stattgefundene Herzoperationen bei denen das Perikard eröffnet werden musste geben 2 Punkte. Ein Serumkreatininwert über 200µmol/L präoperativ gibt 2 Punkte. Eine akute Endokarditis mit laufender Antibiotikatherapie stellt 3 Punkte beim additiven EuroSCORE dar. Ein kritischer präoperativer Zustand mit mindestens einem der folgenden Aspekte, mechanische Reanimation, Beatmung, intraaortale Ballonpumpe, Katecholaminpflicht, Nierenversagen mit Anurie oder Oligurie < 10 ml/h, Kammertachykardie, Kammerflimmern oder Asystolie, zählt 3 Punkte.

Zu den herzbezogenen Risikofaktoren gehört eine instabile Angina Pectoris mit i.v. Nitroverabreichung bis zur Narkoseeinleitung, welche 2 Punkte addiert. Eine eingeschränkte Ejektionsfraktion des linken Ventrikels gibt bei 50-30% 1 Punkt und bei <30% 3 Punkte. Ein rezenter Myokardinfarkt in den letzten 90 Tagen zählt 2 Punkte. Pulmonale Hypertonie mit einem Pulmonalarteriendruck von über 60 mmHg gibt 2 Punkte.

Die operationsbedingten Risikofaktoren beinhalten die Indikation, bei einem Notfall, mit Operation am Tag der Einweisung, gibt es 2 Punkte. Wenn die Operation eine andere ist als eine isolierte CABG kommen 2 Punkte dazu. Eine Operation an der thorakalen Aorta gibt 3 Punkte und eine Septumruptur nach einem Herzinfarkt 4 Punkte. [10]

Der logistische EuroSCORE beinhaltet dieselben Faktoren wie der Additive und wird mit einem logistischen Regressionsmodell berechnet. Das Risiko wird folgendermaßen berechnet:

$$\text{Vorausgesagte Mortalität} = \frac{e^{(\beta_0 + \sum \beta_i X_i)}}{1 + e^{(\beta_0 + \sum \beta_i X_i)}}$$

Dabei steht e für den natürlichen Logarithmus = 2,718281828...

$\beta_0$  ist die Konstante der logistischen Regression = -4,789594

$\beta_i$  ist der spezifische Koeffizient von  $X_i$  der einzelnen Risikofaktoren und der angeführten Tabelle zu entnehmen.

$X_i$  ist das Vorhandensein eines bestimmten Risikofaktors. Wenn der Risikofaktor präsent ist gilt  $X_i = 1$  sonst 0. Das Alter ist die einzige Ausnahme hier gilt Alter <60  $X_i = 1$ , ab dann steigt mit jedem Jahr  $X_i$  um 1 Punkt. [11, 9]

Patientenbezogene Faktoren		Beta
Alter		0,0666354
Geschlecht	<i>weiblich</i>	0,3304052
COPD	<i>Langzeitanwendung von Bronchodilatoren oder Steroiden</i>	0,4931341
Extrakardiale Arterienerkrankung	<i>Claudicatio, Arteria carotis interna Verschluss oder über &gt;50% Stenose oder schon durchgeführten oder indizierten Operationen der Aorta, den Carotiden oder den Arterien der unteren Extremität</i>	0,6558917
Neurologische Erkrankungen	<i>Einschränkungen der Fortbewegung oder des Alltages</i>	0,841626
stattgefundene Herzoperation	<i>Perikarderöffnung</i>	1,002625
Serumkreatinin	<i>&gt;200µmol/L</i>	0,6521653
akute Endokarditis	<i>laufender Antibiotikatherapie</i>	1,101265
kritischer präoperativer Zustand	<i>mechanische Reanimation, Beatmung, intraaortale Ballonpumpe, katecholaminpflicht, Nierenversagen mit Anurie oder Oligurie &lt; 10 ml/h, Kammertachykardie Kammerflimmern oder Asystolie</i>	0,9058132

Tabelle 1: Die patientenbezogenen Risikofaktoren und deren spezifischer Koeffizient  $\beta_i$  im logistischen EuroSCORE [11]

Herzbezogenen Risikofaktoren		Beta
instabile Angina Pectoris	<i>i.v. Nitroverabreichung bis zur Narkoseeinleitung</i>	0,5677075
eingeschränkte Ejektionsfraktion	<i>LVEF 30-50%</i>	0,4191643
	<i>LVEF &lt;30%</i>	1,094443
rezenter Myokardinfarkt	<i>in den letzten 90 Tagen</i>	0,5460218
Pulmonale Hypertonie	<i>Pulmonalarteriendruck &gt;60 mmHg</i>	0,7676924

Tabelle 2: Die herzbezogenen Risikofaktoren und deren spezifischer Koeffizient  $\beta_i$  im logistischen EuroSCORE [11]

Operationsbedingte Faktoren		Beta
Indikation	<i>Notfall am Tag der Einweisung</i>	0,7127953
größere Herzoperation	<i>andere als CABG</i>	0,5420364
Operation an der thorakalen Aorta		1,159787
Septumruptur nach Herzinfarkt		1,462009

Tabelle 3: Die operationsbedingten Risikofaktoren und deren spezifischer Koeffizient  $\beta_i$  im logistischen EuroSCORE [11]

Der EuroSCORE II wird ebenso wie der logistische EuroSCORE anhand eines Regressionsmodells berechnet. Die Formel die verwendet wird bleibt die gleiche.  $\beta_0$  die Konstante der logistischen Regression hat jedoch den Wert -5,324537. Einige Risikofaktoren sind verändert und entfernt worden, neue wurden hinzugefügt und die Gewichtung hat sich verändert. Bei den patientenbezogenen Risikofaktoren ist die Höhe des Serumkreatinins mit der Kreatininclearance ersetzt worden, die mehr Abstufungen bietet. Sie muss zuvor ausgerechnet werden und es fließen darin das Gewicht und nochmal das Alter und das Geschlecht ein. Die neurologische Erkrankung ist durch die allgemeine Beeinträchtigung der Mobilität ausgetauscht worden. Zusätzlich hinzugekommen ist der insulinpflichtige Diabetes.

Bei den herzbezogenen Faktoren ist das NYHA Stadium anzugeben und die linksventrikuläre Ejektionsfraktion wird genauer unterteilt. Die operationsbedingten Faktoren wurden ebenfalls verändert. Bei der Dringlichkeit sind jetzt mehr Optionen verfügbar und nicht nur ob es sich um einen Notfall handelt. Die Größe des Eingriffs oder die Komplexität bietet nun auch mehr Unterscheidungsmöglichkeiten. Der Septumdefekt nach Herzinfarkt wurde gestrichen.

Patientenbezogene Faktoren		Beta
Alter		0,0285181
Geschlecht	<i>weiblich</i>	0,2196434
COPD	<i>Langzeitanwendung von Bronchodilatoren oder Steroiden</i>	0,1886564
Extrakardiale Arterienerkrankung	<i>Claudicatio, Arteria carotis interna Verschluss oder über &gt;50% Stenose oder schon durchgeführten oder indizierten Operationen der Aorta, den Carotiden oder den Arterien der unteren Extremität</i>	0,5360268
allgemeine Beeinträchtigung der Mobilität		0,2407181
stattgefundene Herzoperation	<i>Perikarderöffnung</i>	1,118599
renale Beeinträchtigung Kreatinin Clearance	<i>CC 50-85 (ml/min)</i>	0,303553
	<i>CC &lt;50 (ml/min)</i>	0,8592256
	<i>dialysepflichtig</i>	0,6421508
akute Endokarditis	<i>laufender Antibiotikatherapie</i>	0,6194522
kritischer präoperativer Zustand	<i>mechanische Reanimation, Beatmung, intraaortale Ballonpumpe, katecholaminpflicht, Nierenversagen mit Anurie oder Oligurie &lt; 10 ml/h, Kammertachykardie Kammerflimmern oder Asystolie</i>	1,086517
Insulinpflichtiger Diabetes		0,3542749

Tabelle 4: Die patientenbezogenen Risikofaktoren und deren spezifischer Koeffizient  $\beta_i$  im EuroSCORE II [12]

Herzbezogenen Risikofaktoren		Beta
CCS Klasse 4 Angina	<i>Angina in Ruhe</i>	0,2226147
eingeschränkte Ejektionsfraktion	<i>LVEF 31-50%</i>	0,3150652
	<i>LVEF 21-30%</i>	0,8084096
	<i>LVEF &lt;21%</i>	0,9346919
rezenter Myokardinfarkt	<i>in den letzten 90 Tagen</i>	0,1528943
Pulmonale Hypertonie	<i>Pulmonalarteriendruck systolisch &gt;55 mmHg</i>	0,3491475
	<i>31-55 mmHg</i>	0,1788899
NYHA	<i>II</i>	0,1070545
	<i>III</i>	0,2958358
	<i>IV</i>	0,5597929

Tabelle 5: Die herzbezogenen Risikofaktoren und deren spezifischer Koeffizient  $\beta_i$  im EuroSCORE II [12]

Operationsbedingte Faktoren		Beta
Dringlichkeit	<i>dringend: während des Aufenthaltes ohne Termin</i>	0,3174673
	<i>Notfall am Tag der Einweisung lebensbedrohlich kommen unter Reanimation in den OP</i>	0,7039121
		1,362947
Art der Herzoperation CABG gilt als Basis	<i>eine nicht CABG Operation</i>	0,0062118
	<i>zwei schwere Operationen</i>	0,5521478
	<i>drei oder mehr Operationen</i>	0,9724533
Operation an der thorakalen Aorta		0,6527205

Tabelle 6: Die operationsbedingten Risikofaktoren und deren spezifischer Koeffizient  $\beta_i$  im EuroSCORE II [12]

Trotz der Verbesserungen in der Herzchirurgie sind die additive und logistische EuroSCORE immer noch Modelle mit guter Vorhersagekraft. Dies zeigen sie in ROC (Receiver Operating Characteristic) Analysen mit der Auswertung durch die AUC (Area under the curve). [12]

Unter der Homepage [www.euroscore.org](http://www.euroscore.org) werden sämtliche Rechner zum EuroSCORE gratis angeboten und können ohne großen Aufwand verwendet werden. Das Eingeben der Daten genügt und man bekommt sofort den EuroSCORE II, den Logistischen oder den Additiven angezeigt. Außerdem können einige der Rechner als Excel Tabellen heruntergeladen werden.

## **1.2 Aortenklappen Chirurgie**

### **1.2.1 Grundlagen**

Die Aortenklappe ist zuständig den linken Ventrikel vom großen Blutkreislauf abzuschließen um den Rückfluss des Blutes aus der Aorta in der Diastole zu verhindern. Wenn der linke Ventrikel Blut in der Systole auswirft, soll sie sich öffnen und das Blut passieren lassen. Es kann dabei zu zwei verschiedenen Problemen kommen, der Aortenklappenstenose und der Aortenklappeninsuffizienz, die alleine oder gemeinsam auftreten können. Die Klappen des linken Ventrikels, besonders die Aortenklappe sind bei Klappenfehlern häufiger betroffen als die des rechten Ventrikels. Dies ist auf die mechanische Beanspruchung zurückzuführen, welche wegen des höheren Druckes viel größer ist. [13, 14]

### **1.2.2 Aortenklappenstenose**

Hierbei wird die linksventrikuläre Ausflussbahn durch eine Stenose der Aortenklappe eingeeengt. Die Schwingungsfähigkeit der Klappe wird herabgesetzt und der Blutfluss wird behindert, die Klappenöffnungsfläche wird zu klein. Es kommt zu einer Druckbelastung des linken Ventrikels und zu einem niedrigeren Druck in der Aorta und im Körperkreislauf. Eine konzentrische Hypertrophie des linken Ventrikels entwickelt sich und der enddiastolische Ventrikeldruck wird erhöht. Symptome und Folgen davon können sein: [14, 15]

- Blässe
- Niedrige Blutdruckamplitude und niedriger Blutdruck oder Hypertonie
- Schwindel
- Synkopen, besonders bei Belastung
- Angina pectoris durch Koronarinsuffizienz
- Linksherzinsuffizienz mit Dyspnoe und Leistungsminderung
- Stauungs-Rasselgeräusche über den Lungen
- Palpitation durch Rhythmusstörungen
- Plötzlicher Herztod

Zu den Ursachen der Aortenklappenstenose gehören degenerativ-sklerotische Klappenveränderungen, die im Alter am häufigsten auftreten, rheumatisches Fieber mit rheumatischer Endokarditis, bakterielle Endokarditis und angeborene Kommissurenverschmelzungen. [15]

Die Diagnostik kann mittels Anamnese, körperlicher Untersuchung, EKG, Röntgen, Echokardiographie, MRT/CT und Linksherzkatheter erfolgen. [14]

Der natürliche Verlauf einer Aortenklappenstenose ist chronisch fortschreitend und beginnt mit einer langen asymptomatischen Phase, deren Länge sich je nach Person stark unterscheiden kann. Eine unbewusste körperliche Schonung der Patienten und Patientinnen verschleiert Symptome und verlängert somit den asymptomatischen Zeitraum. Die Prognose ist solange noch keine Symptome bestehen gut, sobald Symptome auftreten ist die Prognose schlecht, mit einer Lebenserwartung von 1-2 Jahren bei Herzinsuffizienz, nach Synkopen von 2-3 Jahren und bei Angina pectoris von 4-5 Jahren. [14, 16]

Für die Aortenklappenstenose gibt es unterschiedliche Einteilungen in Stufen oder Grade. Je nach Schwere sind unterschiedliche Therapieoptionen empfehlenswert. Nachstehend die Einteilung der Aortenklappenstenose und das Vorgehen der 2017 ESC/EACTS Guidelines for the management of valvular heart disease, die 2017 auf den neusten Stand gebracht wurden. [17]

Grundvoraussetzung für die Einteilung der Aortenklappenstenose ist die Echokardiographie. Erst wenn dabei die Klappenmorphologie suspekt erscheint erfolgt eine genau Evaluierung und Einteilung in vier verschiedene Kategorien.

- High-gradient Aortenstenose mit einer KÖF  $<1 \text{ cm}^2$  und einem  $\Delta P$  von  $\geq 40$  mmHg. Dies ist als schwere Aortenstenose anzusehen, unabhängig von der LVEF und der  $V_{\text{max}}$ .
- Low-flow, low-gradient Aortenstenose mit reduzierter Ejektionsfraktion, also KÖF  $<1 \text{ cm}^2$   $\Delta P$  von  $\leq 40$  mmHg, LVEF  $\leq 50\%$  und  $SV_i \leq 35 \text{ mL/m}^2$ . Eine Dobutamin Echokardiographie ermöglicht die weitere Unterscheidung zwischen einer echten schweren und einer pseudoschweren Aortenstenose. Bei einer pseudoschweren Aortenstenose steigt die KÖF auf  $\geq 1,0 \text{ cm}^2$  und das Schlagvolumen normalisiert sich, diese Konfiguration hat prognostisch bessere Aussichten.
- Low-flow, low-gradient Aortenstenose mit erhaltener Ejektionsfraktion also KÖF  $<1 \text{ cm}^2$   $\Delta P$  von  $\leq 40$  mmHg, LVEF  $\geq 50\%$  und  $SV_i \leq 35 \text{ mL/m}^2$ . Diese Art ist typisch bei älteren Patienten und Patientinnen mit kleiner Ventrikelgröße, linksventrikulärer Hypertrophie und bekanntem Hypertonus. Die Diagnose einer schweren Aortenstenose ist bei dieser Konfiguration schwierig und es muss sorgfältig auf Messfehler oder andere Ursachen für die Ergebnisse der Echokardiographie geachtet werden. Die Messung der Klappenverkalkung durch MSCT, um die Schwere der Aortenstenose zu bestimmen ist hilfreich und hat deswegen an Bedeutung gewonnen.
- Normal-flow, low-gradient Aortenstenose mit erhaltener Ejektionsfraktion also KÖF  $<1 \text{ cm}^2$   $\Delta P$  von  $\leq 40$  mmHg, LVEF  $\geq 50\%$  und  $SV_i \geq 35 \text{ mL/m}^2$ . Diese Patienten und Patientinnen haben meistens nur eine moderate Aortenstenose.

Eine frühe Therapie ist für alle symptomatischen Patienten und Patientinnen mit schwerer Aortenstenose aufgrund der schlechten Prognose sehr stark empfohlen. Eine Ausnahme sind Patienten und Patientinnen mit schwerwiegenden Komorbiditäten, die eine kurze Lebenserwartung von weniger als einem Jahr haben

oder bei schlechtem Allgemeinzustand und hohem Alter, wo ein Eingriff weder die Lebensqualität verbessern noch das Leben verlängern könnte. Solange der Druckgradient  $\geq 40$  mmHg ist, gibt es kein unteres Limit der LVEF bei dem nicht mehr operiert werden sollte. Das richtige Vorgehen bei den low-gradient Varianten ist schwieriger. [17]

- Bei low-flow, low-gradient Aortenstenose mit reduzierter Ejektionsfraktion, welche vor allem durch eine erhöhte Nachlast bedingt ist, verbessert ein Eingriff die LV Funktion und ist deswegen empfohlen. Im Gegensatz dazu ist unklar, ob sich die LV Funktion nach einem Eingriff verbessert, wenn die Ursache eine Vernarbung durch eine myokardiale Infektion oder eine Kardiomyopathie ist. Eine Intervention sollte ebenso durchgeführt werden, wenn eine echte schwere Aortenstenose bei erhöhtem Schlagvolumen weiterhin vorliegt. Patienten und Patientinnen mit pseudoschwerer Aortenstenose bei erhöhtem Schlagvolumen sollten eine konservative Therapie für Herzinsuffizienz erhalten.
- Die low-flow, low-gradient Aortenstenose mit erhaltener Ejektionsfraktion ist die herausforderndste Gruppe. Das Vorgehen und der Nutzen einer Intervention sind umstritten und sie sollte nur durchgeführt werden, wenn Symptome vorliegen und eine umfangreiche Evaluation eine signifikante Obstruktion der Klappe ergibt.
- Normal-flow, low-gradient Aortenstenosen mit erhaltener Ejektionsfraktion sollten neu evaluiert werden. Wenn die neue Evaluierung wieder eine normal-flow, low-gradient Aortenstenose bestätigt, ist keine schwere Aortenstenose vorliegend und eine Operation nicht sinnvoll.

Für einen Aortenklappenersatz gibt es zwei Möglichkeiten, die offen chirurgische SAVR (surgical aortic valve replacement) und die minimal invasive TAVI (transcatheter aortic valve replacement) Methode. Die Wahl der Interventionsmethode ist abhängig von unterschiedlichen Faktoren, wie dem Zustand des Patienten oder der Patientin, den Risikofaktoren, dem Risiko der Operation, der Einschätzung durch Spezialisten, der Bewertung durch Scores, der Durchführbarkeit

einer TAVI, der Erfahrung im Krankenhaus und den bisherigen Ergebnisse. In vielen Studien hat TAVI bei älteren Patienten und Patientinnen mit höherem Operationsrisiko bessere Ergebnisse als SAVR gezeigt und sollte in diesen Fällen bevorzugt werden. Die Entscheidung liegt jedoch beim Behandlungsteam und sollte nach sorgfältiger Evaluation individuell erfolgen. Eine Ballondilatation zur Aufdehnung der Aortenklappe kann als Überbrückung bis zum Ersatz in Erwägung gezogen werden. [17]

Die Behandlung von asymptomatischen schweren Aortenstenosen ist kontrovers. Es müssen fallspezifisch die Vorteile und die Risiken sorgfältig abgewogen werden. TAVI ist bei asymptomatischen Aortenstenosen nicht empfohlen. Eine frühe elektive Intervention ist nur bei erniedrigter LV Funktion, die nicht durch andere Ursachen hervorgerufen wurde, wenn sich Symptome während Belastungstests zeigen oder wenn Anzeichen bestehen, dass sich bald Symptome zeigen werden, sinnvoll. Außerdem sollte ein Eingriff nur dann erfolgen wenn das Operationsrisiko niedrig ist, ansonsten gilt watchful waiting. [17]

Eine medikamentöse Therapie für die Aortenstenose ist nicht vorgesehen, weil der Ausgang gegenüber dem natürlichen Verlauf nicht verbessert werden kann. Statine können den Fortschritt der Erkrankung nicht beeinflussen und sind deswegen nicht empfohlen. Patienten und Patientinnen die nicht für einen Eingriff in Frage kommen, oder auf einen Eingriff warten, sollten eine Therapie für Herzinsuffizienz bekommen. Ein vorhandener Hypertonus sollte vorsichtig behandelt werden um Hypotonus zu vermeiden und die Patienten und Patientinnen sollten regelmäßig neu evaluiert werden. Die Aufrechterhaltung des Sinusrhythmus ist wichtig. [17]

### **1.2.3 Aortenklappeninsuffizienz**

Die Schlussunfähigkeit der Aortenklappe ist die Aortenklappeninsuffizienz. Dabei fließt in der Diastole ein Teil des Blutes, abhängig von der Fläche der Regurgitationsöffnung, in den linken Ventrikel zurück. Es kommt zu einer

Volumenbelastung des linken Ventrikels mit exzentrischer Hypertrophie. Symptome und Folgen davon können sein: [15]

- Pulsus celer et altus, Wasserhammerpuls
- Hebender, nach unten außen verlagerter Herzspitzenstoß
- Pulssynchrones Kopfröhen, Musset Zeichen
- Linksherzinsuffizienz mit Dyspnoe und Leistungsminderung
- Stauungs-Rasselgeräusche über den Lungen
- Angina pectoris

Die häufigsten Ursachen für eine Aortenklappeninsuffizienz sind rheumatisches Fieber mit rheumatischer Endokarditis, bakterielle Endokarditis, bikuspide Aortenklappe, aortale Ektasie insbesondere bei älteren Hypertonie Patienten, Aortenaneurysma, Marfan-Syndrom, Lues, akute Aortenklappeninsuffizienz bei Dissektion der Aorta ascendens und posttraumatisch. [15]

Die Diagnostik kann mittels Anamnese, körperlicher Untersuchung, EKG, Röntgen, Echokardiographie, MRT/CT und Linksherzkatheter erfolgen. [14]

Der Verlauf der Aortenklappeninsuffizienz hat eine günstigere Prognose als der Verlauf der Aortenstenose. Dennoch haben Patienten und Patientinnen mit akuter schwerer Aorteninsuffizienz aufgrund der bedenklichen hämodynamischen Verhältnisse ohne Operation eine nachteilige Prognose. Die Langzeitprognose der chronischen Aorteninsuffizienz ist ebenfalls schlecht. Dies wirkt sich erst spät aus, weil der asymptomatische Zeitraum ähnlich wie bei der Aortenstenose sehr lang sein kann. Bei leicht- bis mittelgradiger Aorteninsuffizienz besteht eine 10-Jahresüberlebensrate von 90%, bei einer höher gradigen Aorteninsuffizienz nur mehr 50%. Sobald Symptome auftreten wird die Prognose sehr schlecht, mit einer mittleren Überlebensdauer von 5 Jahren bei Angina pectoris und 2 Jahren bei Herzinsuffizienz. [14, 16]

Für die Aortenklappeninsuffizienz gibt es ebenso unterschiedliche Einteilungen in Stufen oder Grade. Je nach Schwere sind unterschiedliche Therapieoptionen

empfehlenswert. Nachstehend das Vorgehen der 2017 ESC/EACTS Guidelines for the management of valvular heart disease, die 2017 auf den neusten Stand gebracht wurden. [17]

Um zu bestimmen ob es sich um eine schwere Aorteninsuffizienz handelt und welche Therapie sinnvoll ist, wird als wichtigste Untersuchung die Echokardiographie verwendet. Dabei können verschiedene Aspekte, die für die therapeutische Entscheidung bedeutend sind, bestimmt werden: [17]

- Die Klappenmorphologie
- Die Richtung des Insuffizienzjets und dessen Ursprung
- Die Bestimmung der Ätiologie
- Die Quantifizierung der Insuffizienz
- Die LV Funktion und Größe
- Die Maße der Aortenwurzel und der Aorta ascendens
- Der Zustand der einzelnen Taschen und die Abschätzung ob eine Reparatur möglich ist

Wenn die Ergebnisse nicht eindeutig sind kann eine CT oder MRT zur genaueren Abklärung eingesetzt werden. Eine MRT kann zur Bestimmung der Regurgitationsfraktion und eine CT zur Feststellung des exakten Durchmessers bei einer Dilatation durchgeführt werden. [17]

Eine Operation ist grundsätzlich nur indiziert wenn es sich um eine schwere Aorteninsuffizienz handelt oder eine Erweiterung der Aortenwurzel oder der Aorta ascendens vorliegt. Die Aorteninsuffizienz wird außerdem in akute und chronische Fälle unterteilt. Wenn es sich um einen akuten Fall handelt, muss eine Operation schleunigst erfolgen. Diese Fälle sind primär durch eine infektiöse Endokarditis oder eine Aortendissektion bedingt. Bei einer chronischen schweren Aorteninsuffizienz mit

Symptomen ist ein Eingriff unabhängig von der LVEF, außer wenn das Operationsrisiko zu hoch ist, immer indiziert. Bei asymptomatischen Patienten und Patientinnen ist eine Intervention indiziert, wenn die LVEF weniger als 50% beträgt oder der linke Ventrikel vergrößert ist. Dabei muss entweder der LVEDD mehr als 70 mm oder der LVESD mehr als 50 mm sein. Bei Patienten und Patientinnen mit geringer Körpergröße sollte der LVESD auf die Körperoberfläche bezogen werden. Wenn diese Kriterien nicht erreicht werden, sollten regelmäßige Kontrollen erfolgen und Belastungstests durchgeführt werden um grenzwertig symptomatische Patienten und Patientinnen zu finden. In Erwägung gezogen werden kann eine Operation bei asymptomatischen Patienten und Patientinnen mit einer raschen Vergrößerung des Ventrikels oder einer raschen Verschlechterung der Funktion. [17]

Die Operationsgründe der Aortenklappe bei einer Erweiterung der Aorta sind für das Marfan-Syndrom und für eine Dilatation der Aortenwurzel am besten erforscht, weswegen darüber die konkretesten Empfehlungen vorliegen. Ein Eingriff ist bei Marfan-Syndrom mit Erweiterung der Aorta ascendens, ab einem Durchmesser von 50 mm, indiziert. Bei Patienten und Patientinnen mit einer Erweiterung der Aorta ascendens ab 45 mm mit Marfan-Syndrom und zusätzlichen Risikofaktoren, wie eine positive Familienanamnese oder einer TGFBR1 oder TGFBR2 Mutation, sollte eine Operation in Erwägung gezogen werden. Ab einem Durchmesser von 50 mm, bei einer bikuspiden Klappe mit zusätzlichen Risikofaktoren oder einer Aortenisthmusstenose, sollte ein Eingriff in Erwägung gezogen werden. Ebenso sollte, bei allen anderen Patienten und Patientinnen ab einem Durchmesser der Aorta ascendens von 55 mm eine Intervention in Erwägung gezogen werden. Wenn eine Indikation zur Operation der Aortenklappe besteht sollte die Aorta ascendens, ab einem Durchmesser von 45 mm, gleichzeitig operiert werden. [17]

Eine Rekonstruktion der Aortenklappe oder eine klappenerhaltende Operation sollte bei nicht kalzifizierten trikuspiden Klappen oder bei bikuspiden Klappen mit einem Typ 1 oder Typ 2 Mechanismus in Erwägung gezogen werden. In erfahrenen Zentren können Klappenrekonstruktionen oder klappenerhaltende Verfahren bessere

Langzeitergebnisse als ein Klappenersatz erzielen und werden darum diesen vorgezogen. [17]

Eine medikamentöse Therapie kann bei Patienten und Patientinnen, bei denen eine Operation nicht in Frage kommt, helfen die Symptome zu verbessern. ACE-Hemmer, Sartane oder Beta-Blocker können bei der Behandlung hilfreich sein, wenn nach einer Operation noch immer eine Herzinsuffizienz oder ein Hypertonus besteht. Außerdem sollten, bei Patienten und Patientinnen mit Marfan-Syndrom, Beta-Blocker und/oder Losartan vor und nach der Operation in Erwägung gezogen werden. Obwohl keine Beweise existieren, wird dies auch bei allen anderen Arten von Erweiterungen der Aorta ascendens bei bikuspiden Aortenklappen empfohlen. [17]

#### **1.2.4 Mechanische oder biologische Klappe**

Wenn eine Rekonstruktion der Aortenklappe nicht möglich, ein Ersatz aber notwendig ist, gibt es zwei Optionen. Es kann entweder eine mechanische Klappe oder eine biologische verwendet werden. Biologische Klappen können tierischen oder menschlichen Ursprungs sein und können mit einem Stent als Gerüst versehen werden. Es ist ebenso möglich eine Klappe autolog zu transplantieren. Die transplantierte Klappe muss dabei durch eine Prothese ersetzt werden. Der große Unterschied der Klappen liegt in der Haltbarkeit. Biologische Klappen unterliegen einem Zerfallsprozess und verlieren über Jahre hinweg ihre Funktion, bis ein erneuter Ersatz erforderlich ist. Mechanische Klappen haben eine viel längere Haltbarkeit und müssen deswegen nicht ersetzt werden, jedoch ist eine lebenslange Antikoagulation notwendig. Um die richtige Klappe zu wählen muss abgewogen werden ob das Risiko von Blutungen durch Antikoagulation und Thromboembolien oder das Risiko eines Verschleißes eher annehmbar ist. Bei Frauen mit dem Wunsch schwanger zu werden sollte trotz des Zerfallsprozesses eine biologische Klappe verwendet werden, weil während der Schwangerschaft ein hohes Risiko besteht thromboembolische Komplikationen zu erleiden. Es gibt keine verpflichtenden

Altersvorgaben. Die Wahl der Klappe sollte zwischen Patient oder Patientin und Fachpersonal diskutiert werden und individuell erfolgen. [16, 17]

### 1.2.5 Herzinsuffizienz

Die durch Klappenfehler oder andere Erkrankungen entstehende Herzinsuffizienz wird in unterschiedliche Stadien nach AHA (American Heart Association) und NYHA (New York Heart Association) eingeteilt. Anfangs wird die Herzinsuffizienz noch physiologisch kompensiert oder kann therapeutisch kompensiert werden. Wenn die Kompensationsmöglichkeiten nicht mehr ausreichen besteht eine dekompensierte Herzinsuffizienz und die Pumpleistung des Herzens kann nicht mehr aufrechterhalten werden. [14]

#### Stadieneinteilung der Herzinsuffizienz nach NYHA

I	Beschwerdefreiheit, normale körperliche Belastbarkeit
II	Beschwerden bei starker körperlicher Belastung
III	Beschwerden bei leichter körperlicher Belastung
IV	Beschwerden in Ruhe

Tabelle 7: Stadieneinteilung der Herzinsuffizienz nach NYHA

## ABCD-Gruppen der Herzinsuffizienz nach AHA

---

A	Risikofaktoren für die Entwicklung einer Herzinsuffizienz sind vorhanden
B	Es besteht eine strukturelle Herzschädigung ohne Symptomatik
C	Es besteht eine strukturelle Herzschädigung mit Symptomatik
D	Terminale Herzinsuffizienz

Tabelle 8: ABCD-Gruppen der Herzinsuffizienz nach AHA

### 1.2.6 Chirurgisches Verfahren:

Als chirurgischer Zugang für einen Aortenklappenersatz wird in der Norm eine mediane Sternotomie durchgeführt. Es wird ein kardioplegischer Arrest herbeigeführt und die Herz-Lungen-Maschine kommt zur Anwendung. Die Aorta ascendens wird eröffnet und die nativen Taschen werden entfernt, der Klappenring wird jedoch belassen. Verkalkungen werden abgetragen und die exakte Größe des Ringes wird abgemessen. Dann wird die Klappenprothese, die anhand der Körperoberfläche bemessen ist, am Klappenring eingenäht. Die Aortotomie wird wieder verschlossen und die Herzhöhlen werden entlüftet. [18, 19]

### 1.2.7 Interventionelles Verfahren

Die Transkatheterverfahren werden am schlagenden Herzen ohne Herz-Lungen-Maschine durchgeführt. Dabei werden der Katheter und die Prothese über die Arteria femoralis oder nach einer Minithorakotomie transapikal direkt durch die Spitze des linken Ventrikels eingeführt. Die native erkrankte Klappe wird durch einen Hochdruckballon gesprengt und die Prothese wird zusammengefaltet auf einem Stentgerüst an die richtige Position geschoben und entfaltet. [18, 19]

### **1.2.8 Risiken und Komplikationen:**

Bei einem Aortenklappenersatz gibt es wie bei vielen Operationen die Risiken der Narkose und Intubation. Dazu kommen die Anwendung der Herz-Lungen-Maschine und die Kardioplegie. Die spezifischen Komplikationen beinhalten die Klappendysfunktion durch einen partiellen Klappenausriss oder ein Leck, der Verfall einer biologischen Prothese, das Auftreten einer Prothesenendokarditis und das Auftreten eines kompletten AV-Blocks durch intraoperative Verletzung des His-Bündels. Weiteres gehören neurologische Komplikationen durch Kalk- oder Luftembolien, bei mechanischen Klappen die Klappenthrombosierung mit Embolie trotz Antikoagulation und die antikoagulationsbedingte Blutung dazu. Eine Thrombosierung ist jedoch auch bei biologischen Prothesen möglich. Außerdem kann es zu einer bakterielle Besiedlung und einer daraus folgenden bakteriellen Endokarditis kommen. [20]

### **1.2.9 Ziele und Prognose:**

Die Ziele von Aortenklappen Interventionen sind vorhandene Symptome zu verbessern, das Überleben zu verlängern und das Risiko von Komplikationen wie irreversiblen ventrikulären Schädigungen, pulmonalem Hypertonus, Schlaganfällen und Vorhofflimmern zu reduzieren. [21]

Ein Aortenklappenersatz ist einer der risikoärmsten Eingriffe in der modernen Herzchirurgie. Bei isoliertem Aortenklappenersatz liegt die Operationsletalität unter 4%, die 5-Jahres-Überlebensrate beträgt 75% und die 10-Jahres Überlebensrate beträgt 60%. [18, 20]

## **2 Methoden**

### **2.1 Statistische Methoden**

#### **2.1.1 ROC- Kurven**

##### **2.1.1.1 Grundlagen**

Um die Qualität von medizinischen Handlungen messen zu können werden unterschiedliche Tests verwendet. Einer dieser Tests ist die sogenannte ROC-Kurve, welche die Genauigkeit einer Diagnose mittels Spezifität und Sensitivität vergleichbar macht. Dieser Vergleich sollte nicht vom Zufall abhängen und mit blindem Raten erreichbar sein, weswegen verschiedene Probleme berücksichtigt werden müssen. Bei der Diagnose einer seltenen Erkrankung könnte eine sehr hohe Genauigkeit erzielt werden, wenn immer behauptet wird die untersuchte Person sei nicht erkrankt, weil dies meist stimmen würde. Wenn also nur 1 von 100 Personen in einer Gruppe an einer Erkrankung leidet und die Aussage, dass sie gesund sei bei jeder einzelnen Person getroffen wird, ergibt sich eine Genauigkeit von 99%. Alle bis auf eine Person werden somit richtig zugeordnet. Die Genauigkeit wird berechnet indem die Zahl der richtigen Entscheidungen durch die Zahl der Fälle dividiert wird. Dieses Problem kann durch das Anwenden der Sensitivität und Spezifität gemeinsam, oder durch Einbringen der Prävalenz beseitigt werden. Ein weiteres Problem besteht darin, welche falsche Entscheidung getroffen wird. Es besteht ein Unterschied ob einer Person mit einer Erkrankung gesagt wird, sie ist gesund oder ob einer gesunden Person gesagt wird sie ist erkrankt. Daraus ergibt sich ein großer Unterschied in der Nützlichkeit und den Anwendungsgebieten von Tests. [22]

Um diese Probleme zu beheben, werden bei ROC-Kurven Sensitivität und Spezifität verwendet. Die Sensitivität gibt die Genauigkeit für die richtig erkannten positiven Fälle an und die Spezifität gibt die Genauigkeit für die richtig erkannten negativen an. Es muss dabei eine klare binäre Unterscheidung zwischen genau zwei unterschiedlichen Zuständen möglich sein. Positiv bedeutet das Merkmal, der Zustand oder die Krankheit ist vorhanden und negativ bedeutet das Merkmal, der

Zustand oder die Krankheit ist nicht vorhanden. Die Sensitivität wird berechnet indem die Anzahl der richtig positiven Entscheidungen durch die Gesamtzahl der positiven Fälle dividiert wird. Die Spezifität wird ähnlich berechnet, mit dem Unterschied dass die Anzahl der richtig negativen Entscheidungen durch die Gesamtzahl der negativen Fälle dividiert wird. Für das Beispiel mit den 100 Personen ergibt sich eine Spezifität von 100%, aber eine Sensitivität von 0%. [22]

$$\text{Sensitivität} = \frac{\text{Anzahl der richtig positiven Entscheidungen}}{\text{Anzahl der positiven Fälle}}$$

$$\text{Spezifität} = \frac{\text{Anzahl der richtig negativen Entscheidungen}}{\text{Anzahl der negativen Fälle}}$$

Die Sensitivität kann auch als Richtig-positiv-Rate und die Spezifität als Richtig-negativ-Rate angegeben werden. Im Gegensatz dazu gibt es die Falsch-negativ-Rate und die Falsch-positiv-Rate. Die Falsch-negativ-Rate wird berechnet indem die Anzahl der falsch negativen Entscheidungen durch die Gesamtzahl der positiven Fälle dividiert wird und bildet mit der Sensitivität ein Paar. Die Falsch-positiv-Rate wird berechnet indem die Anzahl der falsch positiven Entscheidungen durch die Gesamtzahl der negativen Fälle dividiert wird und bildet mit der Spezifität ein Paar. Es ist nur nötig jeweils einen Wert der beiden Paare zu wissen, weil sie zusammen 1 ergeben. Richtig-positiv-Rate + Falsch-negativ-Rate = 1 und Richtig-negativ-Rate + Falsch-positiv-Rate = 1. [22]

		Test Resultat:		
		(T+)	(T-)	Total
Krankheitsstatus:	(D+)	a	b	a + b
(Disease state)	(D-)	c	d	c + d

Tabelle 9: Konfusionsmatrix

a = Personen mit der Krankheit, welche ein positives Testresultat erhalten

b = Personen mit der Krankheit, welche ein negatives Testresultat erhalten

c = Personen ohne der Krankheit, welche ein positives Testresultat erhalten

d = Personen ohne der Krankheit, welche ein negatives Testresultat erhalten

Eine Schreibform die übersichtlich ist und in welche die Ergebnisse eines Tests einfach eingetragen werden können, ist die Kontingenztafel oder auch Kreuztafel. In diesem besonderen Fall handelt es sich um eine Wahrheits- oder Konfusionsmatrix, welche anhand eines binären Klassifikators, erkrankt und nicht erkrankt, in zwei Klassen unterteilt. Diese Aussage, ob eine Krankheit vorhanden ist, kann richtig sein oder nicht. Mit so einer Wahrheitsmatrix können, aus den jeweils vier Gruppen die durch den Krankheitsstatus und das Testresultat definiert sind, sämtliche Raten berechnet werden. Die Sensitivität ist a durch a + b, die Spezifität ist d durch c + d, die Falsch-negativ-Rate ist b durch a + b und die Falsch-positiv-Rate ist c durch c + d. [23]

Zum Vergleich von unterschiedlichen Tests kann die Genauigkeit der Sensitivität und Spezifität verwendet werden. Die Genauigkeit eines Tests wird berechnet indem die Sensitivität, multipliziert mit dem Anteil der Bevölkerung der positiv ist und die Spezifität, multipliziert mit dem Anteil der Bevölkerung der negativ ist, addiert werden. Dies ist möglich, weil die Genauigkeit die Anzahl der richtigen Entscheidungen durch die Anzahl der Fälle ist. Tests mit unterschiedlicher Sensitivität und Spezifität können allerdings dieselbe Genauigkeit aufweisen, weshalb ein Vergleich nur anhand der Genauigkeit nicht unbedingt sinnvoll ist. Die Anwendbarkeit von Tests ist auch davon

abhängig wie schwerwiegend falsche Entscheidungen sind. Wenn das Nicht-Entdecken einer Erkrankung, also falsch negative Entscheidungen, möglichst ausgeschlossen werden sollen, ist eine hohe Sensitivität von Nöten. Wenn hingegen fälschlich positive Annahmen unerwünscht sind, ist eine hohe Spezifität notwendig. Somit haben Tests unabhängig von ihrer Genauigkeit unterschiedliche Nutzbarkeit. [22]

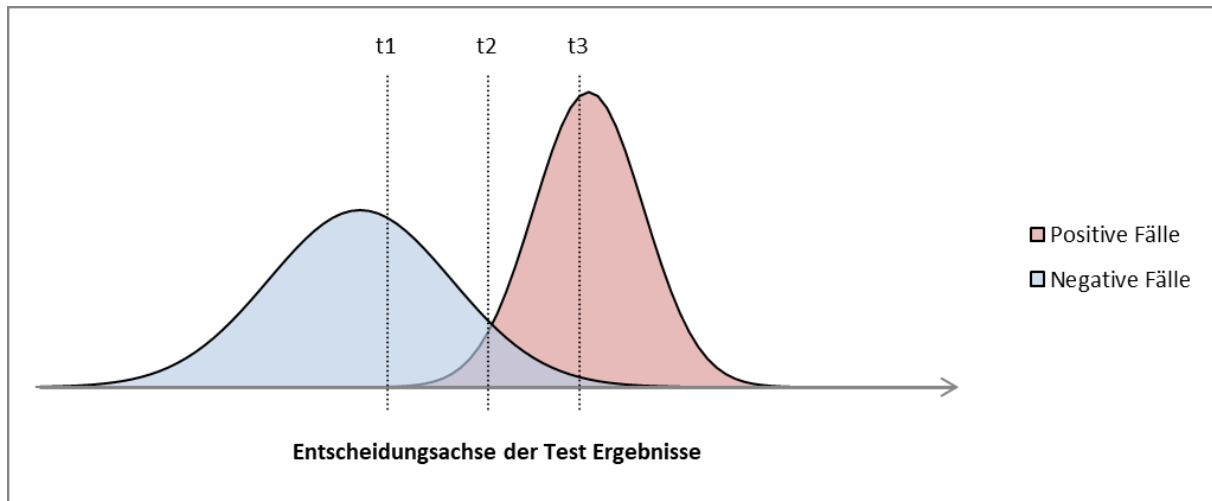


Abbildung 1: Entscheidungsachse eines Tests mit drei möglichen Grenzwerten (t1, t2, t3) die bestimmen, ab welchem Wert ein Test Resultat positiv ist.

Viele Untersuchungen haben aber nicht nur zwei mögliche Zustände oder Werte, wie unter anderem Laborwerte. Bei Ihnen muss künstlich eine Linie als Grenzwert gesetzt werden, die die Ergebnisse in zwei Gruppen aufteilt, wie in Abbildung 1 dargestellt. Dies stellt einerseits ein Problem dar, weil sich die Werte von Gesunden und Kranken überschneiden, ermöglicht andererseits die Beeinflussung der Sensitivität und Spezifität durch das Setzen des Grenzwertes. Dies kann in gewissem Maße auch auf Tests, die subjektiv bewertet werden, angewendet werden. Wenn ein Radiologe oder eine Radiologin ein Röntgenbild auf eine Erkrankung untersucht, kann er dazu veranlasst werden strenger oder milder zu entscheiden, womit ebenso eine Verschiebung des Grenzwertes erreicht wird. Wenn der Grenzwert eines Tests verändert wird ändern sich Sensitivität, Spezifität, die Falsch-positiv-Rate und die Falsch-negativ-Rate. Ein Erniedrigen des Grenzwertes führt zu einer Zunahme der Sensitivität und der Falsch-positiv-Rate und ein Erhöhen führt zu

einer Abnahme. Diese zwei Raten steigen oder sinken immer gemeinsam bei einer Änderung des Grenzwertes, gleiches gilt für die Spezifität und die Falsch-negativ-Rate. Diese müssen nicht bestimmt werden, weil sie durch die jeweils anderen schon gegeben sind und werden darum meistens auch nicht angegeben. Diese Veränderungen können auch graphisch dargestellt werden. Dazu werden die Sensitivität und die Falsch-positiv-Rate als x und y in einem Koordinatensystem eingetragen. Die Achsen reichen von Null bis Eins, weil dies die einzig möglichen Werte für die Sensitivität und Falsch-positiv-Rate sind. Wenn dies jetzt für jeden möglichen Grenzwert wiederholt wird und für jede Kombination ein Punkt eingetragen wird ergibt sich eine Kurve. Diese Kurve ist eine ROC (receiver operating characteristic) Kurve. [22]

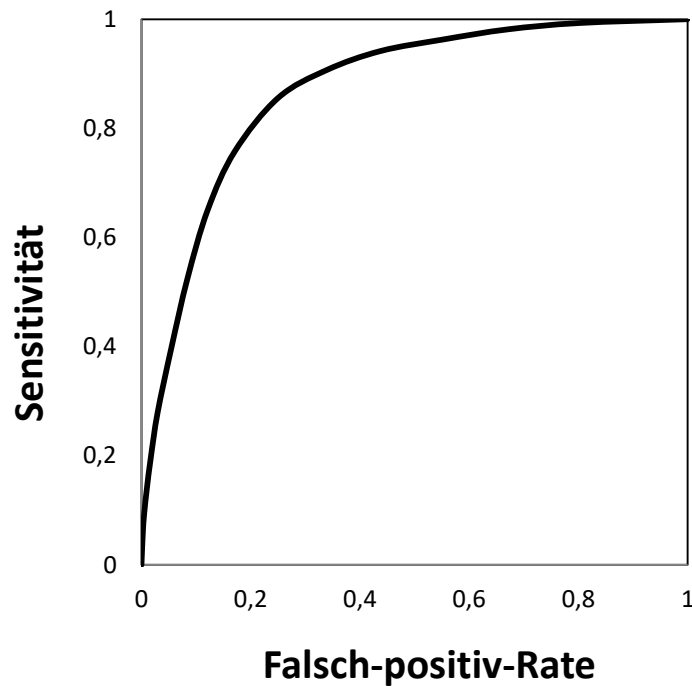


Abbildung 2: ROC-Kurve

Alle ROC-Kurven beginnen am linken unteren Ende der Darstellung bei einer Sensitivität und Falsch-positiv-Rate von 0. Dies kommt dadurch zustande, dass eine Verschiebung des Grenzwertes nach oben alle Fälle negativ werden lässt. Während dem Verschieben des Grenzwertes nach unten werden immer mehr Fälle positiv bis schließlich alle positiv sind. Sensitivität und Falsch-positiv-Rate erreichen den Wert 1.

Die ROC-Kurve sollte sich über der Diagonale zwischen der linken unteren Ecke und der rechten oberen Ecke befinden. Die Wahrscheinlichkeit, dass eine positive Entscheidung bei einem positiven Fall getroffen wird, ist über dieser Diagonale höher, als dass eine positive Entscheidung bei einem negativen Fall getroffen wird. Dadurch lässt sich schlussfolgern, dass ROC-Kurven anfangs steil ansteigen müssen, um sich über der Diagonale zu befinden und gegen Ende abflachen, sodass Sensitivität und Falsch-positiv-Rate beide den Wert 1 erreichen. [22]

Die ROC-Kurve beschreibt die Veränderung der Sensitivität und Falsch-positiv-Rate eines Tests wenn sich der Grenzwert ändert. Dies ermöglicht es den bestmöglichen Kompromiss zwischen Sensitivität und Falsch-positiv-Rate zu finden. Dies ist unabhängig von der Prävalenz, die jedoch einen anderen Grenzwert sinnvoll machen könnte. Wenn die Prävalenz niedrig ist sollte die Falsch-positiv-Rate auch niedrig gehalten werden, weil ansonsten die meisten positiven Entscheidungen falsch sein werden. Dies wäre eine unnötige Belastung für das Gesundheitssystem und Patienten und Patientinnen. Ebenso sollte die Falsch-positiv-Rate niedrig gewählt werden, wenn die Folgen einer falsch positiven Entscheidung sehr schlecht sind, auch wenn dies eine niedrige Sensitivität zur Folge hat. Wenn die Prävalenz im Gegensatz dazu sehr hoch ist oder das Übersehen von wirklich Positiven sehr schlecht ist, sollte der Grenzwert so gewählt werden, dass die Sensitivität auf Kosten einer hohen Falsch-positiv-Rate hoch ist. Die ROC-Kurve zeigt wie sehr die Falsch-positiv-Rate gehoben werden muss, um die gewünschte Sensitivität zu erreichen. [22]

### **2.1.1.2 Vergleich von ROC-Kurven**

Es gibt mehrere Möglichkeiten ROC-Kurven und somit Tests miteinander zu vergleichen. Um zwei Tests miteinander zu vergleichen, kann eine Kombination von Sensitivität und Falsch-positiv-Rate gemessen werden und dabei der Grenzwert so angepasst werden, dass die Sensitivität bei beiden Tests gleich ist. Dies ermöglicht den Vergleich der Falsch-positiv-Rate, um zu bestimmen welcher Test besser ist.

Genauso könnte die Falsch-positiv-Rate angeglichen werden und über die Sensitivität bestimmt werden welcher Test eine bessere Vorhersagekraft hat. Um festzustellen, ob ein Test eine höhere diagnostische Genauigkeit hat, können auch die zwei Kurven und nicht nur ein Punkt, miteinander verglichen werden. Wenn eine Kurve klar über der anderen liegt verfügt dieser Test über eine bessere Vorhersagekraft. Beim Vergleich zweier Kurven kann es jedoch auch dazu kommen, dass die Kurven sich kreuzen, wodurch nicht klar ersichtlich ist, welcher Test besser ist. In so einem Fall ist die diagnostische Genauigkeit situationsabhängig. Für die Anwendung spielen dabei Prävalenz, Vorteile und Nachteile der Entscheidungen eine Rolle. [22]

Angenommen man will zwei Tests, Test A und Test B, vergleichen. Diese Tests wurden bei jeweils einem willkürlich ausgewählten Grenzwert durchgeführt und Sensitivität, Falsch-positiv-Rate und Genauigkeit werden bei diesem Grenzwert bestimmt. Mit diesen Informationen allein kann grundsätzlich keine Aussage getroffen werden, welcher Test eine höhere diagnostische Genauigkeit hat, weil die Wahl des Grenzwertes nicht berücksichtigt wurde. Es könnten beide Tests gleich gut sein, einer könnte generell besser oder in bestimmten Bereichen besser sein. Beim Vergleich der ROC-Kurven wird dies ersichtlich. Dazu ein Beispiel. [22]

		Test Resultat:		
		(T+)	(T-)	Total
Krankheitsstatus:	(D+)	96	4	100
(Disease state)	(D-)	70	30	100

Tabelle 10: Konfusionsmatrix Test A

		Test Resultat:		
		(T+)	(T-)	Total
Krankheitsstatus:	(D+)	76	24	100
(Disease state)	(D-)	30	70	100

Tabelle 11: Konfusionsmatrix Test B

Mit den Daten aus den Konfusionsmatrizen können die Sensitivität, die Falsch-positiv-Rate und die Genauigkeit berechnet werden. Die Prävalenz und Abwesenheit der Erkrankung zur Berechnung der Genauigkeit wird aus der totalen Anzahl der Erkrankten und der totalen Anzahl der Gesunden kalkuliert. In diesem Beispiel hat die Erkrankung eine Prävalenz von 50%. Test A hat eine Sensitivität von 0,96, eine Falsch-positiv-Rate von 0,70 und eine Genauigkeit von 0,63. Test B hat hingegen eine Sensitivität von 0,76, eine Falsch-positiv-Rate von 0,30 und eine Genauigkeit von 0,73. Anhand dieser Daten allein kann jedoch kein Vergleich zwischen den Tests durchgeführt werden. Ein Vergleich bei nur einem Grenzwert kann sehr trügerisch sein, deswegen sollten immer vollständige ROC-Kurven für einen Vergleich verwendet werden. Es gibt verschiedene Möglichkeiten inwieweit sich Test A und Test B unterscheiden können, es ist sogar auch möglich, dass sie identisch sind. In Abbildung 3 ist zu sehen, dass Test A und Test B identisch sind, jedoch könnte Test A auch in allen Bereichen Test B überlegen sein.

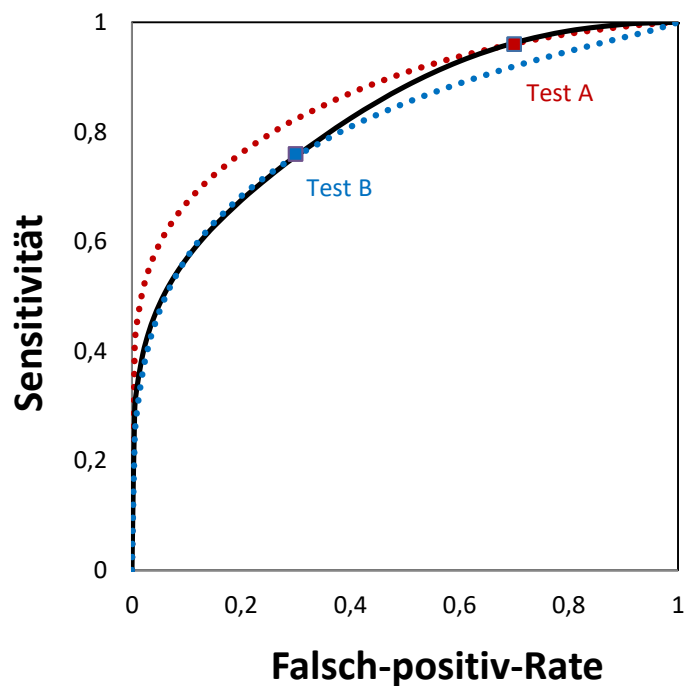


Abbildung 3: ROC-Kurve im Vergleich Test A und Test B

### 2.1.1.3 Indices zur diagnostischen Genauigkeit von ROC-Kurven

Reale ROC-Kurven werden durch 2 Parameter bestimmt um jeweils die Höhe und Stärke einer Kurve wiederzugeben. Deswegen ist es schwer mit nur einer Zahl einen Vergleich zwischen ROC-Kurven durchzuführen und Indices die nur auf einer Zahl beruhen sind fehleranfällig. Die Unterscheidung zwischen besser und schlechter ist jedoch komplett eindimensional und erfordert einen Einzelzahl Index um einen Vergleich zu erstellen. Darum muss ein Einzelzahl Index gewählt werden der aussagekräftig und möglichst zuverlässig im praktischen Sinne ist. [24]

Wenn sich zwei ROC-Kurven nicht überschneiden, wie bereits gezeigt, kann klar definiert werden, welche besser ist. Mit der AUC, area under the Curve, also der Fläche unter der Kurve oder auch Flächenindex, kann eine Wertung von ROC-Kurven erfolgen. Diese Fläche kann auf drei Arten interpretiert werden. Erstens als gemittelte Sensitivität auf alle möglichen Spezifitäten, womit die durchschnittliche

Höhe der Kurve ausgedrückt wird. Zweitens als gemittelte Spezifität über alle möglichen Sensitivitäten, was aussagt wie weit links sich die Kurve im Einheitsquadrat befindet. Drittens als Wahrscheinlichkeit wirklich positive Fälle richtig zu identifizieren, wenn eine Person zwischen wirklich positiven und wirklich negativen Fällen differenzieren muss. Dabei kann der Flächenindex 0,5, dem niedrigsten Wert, der einer Diagonale zwischen der linken untern und rechten oberen Ecke im Einheitsquadrat entspricht, betragen und die Wahrscheinlichkeit eine positive Entscheidung zu treffen ist, unabhängig ob ein Fall wirklich positiv oder negativ ist, immer gleich. Wenn der Flächenindex 1 ist, stellt dies die bestmögliche ROC-Kurve dar, die senkrecht ansteigt bis die Sensitivität 1 erreicht und erst dann waagrecht bis zur rechten oberen Ecke geht. [24]

Die AUC beschreibt immer eine gesamte ROC-Kurve und fasst deren Verlauf zusammen, womit grundsätzlich ein guter Parameter zum Vergleich besteht. Zwei überschneidende Kurven, können jedoch denselben Flächenindex aufweisen, obwohl sie sich in einigen Bereichen unterscheiden. Dabei kann es vorkommen, dass zwei Kurven mit dergleichen AUC gänzlich anders verlaufen. Dies ist besonders gefährlich, wenn eine Kurve im klinisch wichtigen Bereich eine hohe Falsch-positiv-Rate hat. Eine Kurve mit einem niedrigeren Flächenindex könnte im klinisch wichtigen Bereich eine niedrige Falsch-positiv-Rate aufweisen und somit besser sein als eine Kurve mit höherem Flächenindex. Um diese Einschränkung zu beheben wurden zwei Indices, die Teil-Flächen beschreiben, vorgeschlagen. Der eine gibt die AUC in einem bestimmten Bereich der Falsch-positiv-Rate an und der andere gibt die Fläche rechts von der Kurve in einem bestimmten Bereich der Sensitivität an. Dabei muss immer genau angegeben werden, um welchen Bereich es sich handelt, der aufgrund von klinischen Überlegungen gewählt werden sollte. Diese Indices basieren jeweils auf einem Teil der Sensitivität und falsch-positiv-Rate ohne auf das Verhältnis von Sensitivität und Spezifität zu achten, welches ebenso eine wichtige Rolle spielt. In einem ausgewählten Bereich kann mit Teilindices dennoch die diagnostische Genauigkeit verglichen werden. Der einfachste Vergleich von ROC-Kurven entsteht, wenn die Sensitivität bei einem vorher festgelegten Wert der Falsch-positiv-Rate verglichen wird oder umgekehrt, die Falsch-positiv-Rate bei einer vorher festgelegten Sensitivität verglichen wird. Dieser simple Vergleich gilt zwar nur für den festgelegten

Wert, ist jedoch einfach und intuitiv anwendbar. Solche Vergleiche sind dennoch nicht weit verbreitet, weil keine Einigungen für einheitliche Werte der Sensitivität und Falsch-positiv-Rate gefunden werden können, anhand deren ROC-Kurven verglichen werden sollen. Außerdem besagt die Entscheidungstheorie, dass ROC-Kurven anhand der Likelihood-Ratio verglichen werden sollten und nicht anhand einzelner Werte. Dieses Problem kann jedoch gelöst werden, indem der relative Nutzen und die relativen Kosten von richtig positiven, richtig negativen, falsch positiven und falsch negativen Entscheidungen berücksichtigt werden. Der Grenzwert eines Tests wird dabei verschoben, bis das bestmögliche Gleichgewicht zwischen den vier Entscheidungen gefunden ist und der Nettovorteil am größten ist. Eine Einigung über den relativen Nutzen und die relativen Kosten von Entscheidungen ist jedoch schwer zu erzielen, weshalb dieser Index ebenso kaum Verwendung findet. [24]

#### **2.1.1.4 Generalisierte ROC-Analyse**

Konventionelle ROC-Analysen beschreiben, bei Untersuchungen mit genau zwei möglichen Zuständen, immer alle Kombinationen von Sensitivität, Spezifität, Falsch-positiv-Rate und Falsch-negativ-Rate, die erreicht werden können. Dabei ist die Aufteilung in zwei Gruppen essentiell. Meistens wird zwischen positiven und negativen Fällen unterschieden, es ist jedoch auch jede mögliche andere Unterscheidung in zwei Gruppen zulässig. Dabei muss nur beachtet werden, dass mittels Goldstandard die Zugehörigkeit zu den Gruppen klar festgestellt werden kann. Vor allem in der medizinischen Bildgebung können die zwei Gruppen jedoch gemischt sein, sodass beide Gruppen unterschiedliche Zustände beinhalten. Wenn eine Erkrankung verschiedene Zustände aufweist, kann es dazu kommen, dass unterschiedliche Zustände beiden Gruppen zugeordnet werden. Patienten und Patientinnen können sowohl frühe Stadien sowie späte Stadien oder andere Krankheiten mit ähnlichen Symptomen aufweisen, was eine Zuordnung erschwert. Darum sind ROC-Kurven auch abhängig von den möglichen Zuständen und Stadien einer Erkrankung in denen sich Patienten und Patientinnen befinden können. Von denen alle in einer Test Gruppe vorhanden sein müssen um die diagnostische Genauigkeit zuverlässig abschätzen zu können. Theoretisch ist es sogar möglich und zulässig zwischen normalen und abnormen Patienten und Patientinnen zu

unterscheiden, jedoch ist dies aufgrund der zahlreichen Auswahlaspekte praktisch nicht durchführbar. [24]

Die klinische Relevanz solcher Unterscheidungen hängt von den gerade durchzuführenden Aufgaben und dem weiteren Prozedere ab. In der medizinischen Bildgebung muss ein Radiologe oder eine Radiologin unter anderem entscheiden, ob nach einer Mammographie eine Biopsie durchgeführt werden soll oder die Entscheidung zwischen An- und Abwesenheit einer Malignität treffen. Meistens ist der Diagnoseweg einer Erkrankung die Abfolge von solchen Unterscheidungen in zwei Gruppen und kann durch konventionelle ROC-Analysen dargestellt werden. Die bloße Entdeckung einer Erkrankung ist jedoch nicht ausreichend. Die genaue Lokalisation ist ebenso wichtig, welche allerdings durch eine konventionelle ROC-Analyse nicht dargestellt werden kann. Mit einer konventionellen ROC-Analyse sind die Fähigkeiten eines diagnostischen Systems die genaue Lokalisation eines Tumors, die Art eines Tumors, die Abwesenheit von Tumoren und die Anwesenheit von multiplen Tumoren festzustellen, nicht quantifizierbar. Eine Quantifizierung der Leistung eines diagnostischen Systems bei Unterscheidung von drei oder mehr Gruppen gleichzeitig, wie Abwesenheit einer Raumforderung, benigne Raumforderung oder maligne Raumforderung, ist ebenso nicht möglich. Hier kann nur schrittweise, zuerst die Entscheidung zwischen der An- und Abwesenheit einer Raumforderung und dann die Unterscheidung zwischen einer benignen oder malignen Raumforderung erfolgen. Um diese Aufgaben zu bewältigen sind einige generalisierte Formen von ROC-Analysen vorgeschlagen worden, die einige Vorteile jedoch auch ungelöste Probleme und Einschränkungen aufweisen. [24]

Um das Problem der Lokalisation in den Griff zu bekommen sind generalisierte LROC (localization receiver operating characteristics) Analysen entworfen worden, welche bei der Unterscheidung der An- und Abwesenheit einer Läsion und der Angabe der genauen Lokalisation bei Anwesenheit der Läsion zur Anwendung kommen. Dabei werden bestimmte Konfidenzintervalle, welche die Wahrscheinlichkeit der Anwesenheit einer Läsion, wie in einer konventionellen ROC-Kurve und die Lokalisation, die in Quadranten oder mit Koordinaten angegeben

werden kann, ermittelt. Die Kurve wird dann mit den Daten der Konfidenzintervalle, ähnlich wie eine konventionelle ROC-Kurve erstellt. Richtig positive Entscheidungen werden dabei nur anerkannt, wenn die Angabe der Lokalisation sich nah genug an der Position der tatsächlichen Läsion befindet, sofern eine vorhanden ist. Die Bewertung richtig negativer Entscheidungen erfolgt gleich wie bei einer konventionellen ROC-Kurve, womit die horizontale Achse einer generalisierten Kurve der einer konventionellen Kurve gleicht. Dies entspricht der Wahrscheinlichkeit eine Läsion in einem negativen Bild zu entdecken. Die vertikale Achse einer generalisierten LROC-Kurve ist jedoch, zusätzlich zur Wahrscheinlichkeit eine Läsion in einem positiven Bild zu entdecken, abhängig von der Lokalisation der Läsion und wird somit strenger bewertet. Darum liegt bei einer LROC-Kurve jeder Punkt, bei dem die Falsch-positiv-Rate größer als 0 ist, niedriger als in der entsprechenden konventionellen ROC-Kurve, welche aus denselben Daten, durch Nichtbeachten der Lokalisation, errechnet werden kann. In dem Einheitsquadrat auf dem ROC-Kurven aufgetragen werden, erreichen LROC-Kurven darum in der Regel nie die rechte obere Ecke. Um eine Sensitivität von 1 zu erreichen, reicht es nämlich nicht jedes Bild als positiv zu bezeichnen. Das Ergebnis einer LROC-Analyse hängt auch von den erlaubten Lokalisationsfehlern ab. Diese Toleranz bestimmt, ob das Bild, bei einer Einstufung über dem Entscheidungsgrenzwert eines tatsächlich positiven Bildes, als richtig positiv oder falsch negativ gewertet wird. Bildgebungssysteme werden, aufgrund der Beziehung zwischen LROC-Kurven und ihren entsprechenden ROC-Kurven für jede festgelegte Toleranz von Lokalisationsfehlern, von beiden auf die gleiche Art bewertet. Darum können in Vergleichsstudien durch konventionelle ROC-Analysen, welche die Genauigkeit der Lokalisation nicht bewerten, keine irreführenden Ergebnisse erzielt werden. Einige Überlegungen lassen jedoch darauf schließen, dass die statistische Aussagekraft von den Lokalisationsdaten aus LROC Studien größer ist als von konventionellen ROC-Analysen. [24]

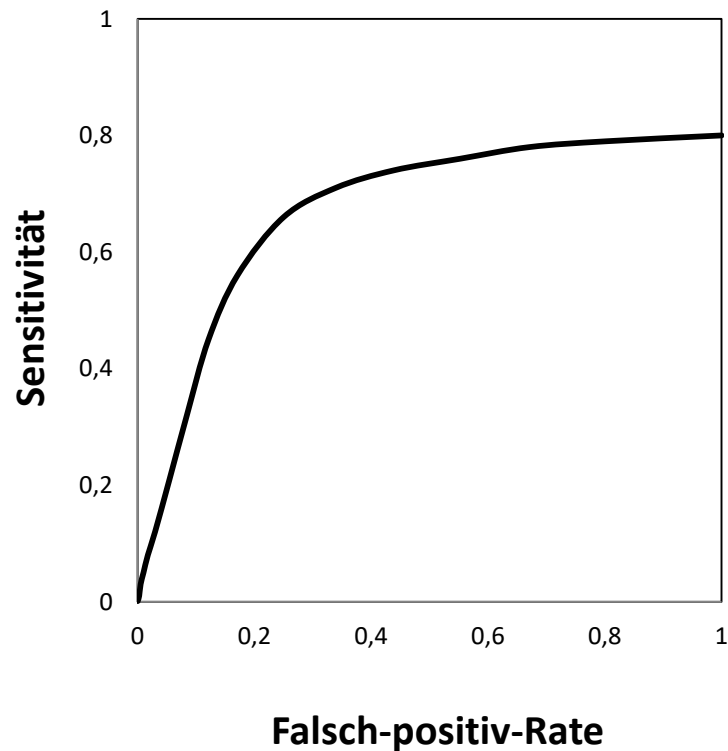


Abbildung 4: LROC-Kurve

Eine noch weiter generalisierte Form von LROC-Analysen stellen FROC (Free-response operating characteristic) Analysen dar. Dabei kann mehr als eine Läsion pro Bild angegeben werden. Es werden wiederum Konfidenzintervalle für die Anwesenheit jeder angegeben Läsion ermittelt. Unter Berücksichtigung der möglichen Positionen des Entscheidungsgrenzwertes, der zwischen positiven und negativen Berichten unterscheidet, wird eine Kurve anhand der Konfidenzintervalle und Lokalisationsdaten erstellt. Diese Kurve zeigt die Beziehung zwischen dem Anteil der korrekt entdeckten Läsionen und der Anzahl der falsch positiven Berichte pro Bild. Die horizontale Achse ist hierbei nicht mit dem maximalen Wert von 1 normiert, sondern muss erweitert werden, weil die Anzahl an falsch positiven Berichten pro Bild unbekannt ist. Wie bei LROC-Analysen ist die statistische Aussagekraft von den Lokalisationsdaten potentiell größer als bei konventionellen ROC-Kurven. Das Ergebnis hängt wiederum von den erlaubten Lokalisationsfehlern ab. Der allgemeine Fortschritt bei generalisierten ROC-Analysen, welche Systeme bewerten, die zwischen mehr als zwei Gruppen unterscheiden können, ist jedoch gering. [24]

### 2.1.1.5 Kurvenanpassung

Um eine durchgängige Kurve zu erstellen und damit auch die Fläche unter der Kurve bestimmen zu können ist eine Kurvenanpassung notwendig. Durch sie werden die nicht experimentell ermittelten Punkte auf einer ROC-Kurve geschätzt. Dies ist erforderlich, weil die Fallzahl für jedes Experiment begrenzt ist und bei einem subjektiven Entscheidungsgrenzwert durch einen Beobachter nur wenige Punkte einer ROC-Kurve, durch mildere oder strengere Entscheidungen beeinflusst, aufgetragen werden können. Eine allein subjektiv mit dem Auge ergänzte ROC-Kurve stellt meist schon eine gute Schätzung dar. Wenn eine objektive Kurvenanpassung erfolgen soll, müssen einige Annahmen bezüglich der Funktion der Kurve gemacht werden. [22]

Grundsätzlich werden zwei Methoden zur Erstellung von Kurven verwendet. Nicht-parametrische und parametrische Verfahren. Für die nicht-parametrischen Methoden sind keine Annahmen notwendig. Sie benötigen kein bestimmtes Verteilungsmuster der einzelnen Variablen und erzeugen empirische ROC-Kurven. Dazu wird die Trapezregel genutzt und alle beobachteten Punkte mit einer geraden Linie verbunden. Es entstehen mehrere Trapeze, deren Flächeninhalt einfach ausgerechnet und addiert werden kann um die AUC zu erhalten. Eine weitere nicht parametrische Methode, die zur Bestimmung der Fläche genutzt werden kann, benutzt den „Mann-Whitney-U-Test“ und den c-index. [25]

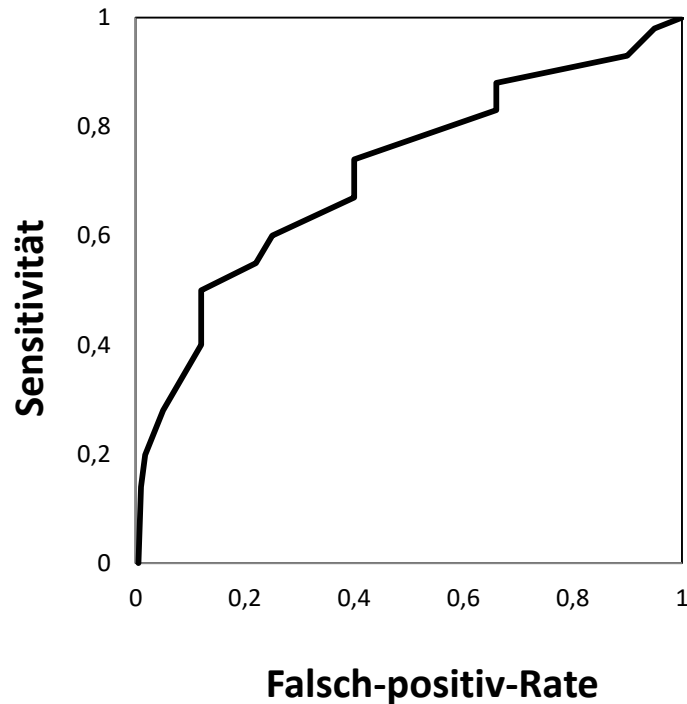


Abbildung 5: Empirische ROC-Kurve

Wenn die statistische Verteilung der Erkrankten und nicht Erkrankten bekannt ist oder angenommen werden kann, können parametrische Verfahren verwendet werden. Für diesen Zweck wird zumeist angenommen, dass eine binormale Verteilung der Erkrankten und nicht Erkrankten vorliegt oder durch eine Transformation erreicht werden kann. Somit können konventionelle ROC-Kurven anhand des Binormal-Standardmodells angepasst werden. Dieses nimmt an, dass die wahre Form der Kurve durch ein Paar normalverteilter Wahrscheinlichkeitsverteilungen an unterschiedlichen Positionen der Entscheidungsachse mit möglicherweise unterschiedlicher Standardverteilung ausgedrückt werden kann. Diese Annahme ist am gängigsten und gibt das am weitesten verbreitete Modell, wie auch in Abbildung 1 zu sehen ist, wieder. Mit der Zeit zeigte sich, dass dieses Modell eine gute Anpassung für statistisch verlässliche ROC-Kurven erstellt. Wenn der verwendete Datensatz jedoch nicht ausreichend ist, können die erstellten Kurven unpassende Formen annehmen. [22, 24, 25]

Um dies zu verhindern wurden weitere Modelle zum Anpassen von ROC-Kurven entwickelt. Unter den entwickelten Systemen hat sich das echte binormale Modell als das verlässlichste herausgestellt. Ein weiterer Vorteil dieses Modells ist, dass die Anpassungen nahezu identisch zu denen des Binormal-Standardmodells sind, wenn dieses eine geeignete Kurve erzeugt. [24]

### **2.1.1.6 Statistische Signifikanz und Abweichungen**

Um auch sicher davon auszugehen, dass Unterschiede zwischen geschätzten ROC-Kurven und deren AUC real und nicht Zufallsprodukte sind, ist es notwendig diese einer statistischen Signifikanztestung zu unterziehen. Damit soll bewiesen werden, dass es sich bei dem Ergebnis nicht um Abweichungen durch zufällige Schwankungen handelt. Um die statistische Signifikanz zwischen geschätzten ROC-Kurven zu bestimmen sind zahlreiche Tests entwickelt worden. Bei diesen Tests sind zwei Punkte zu berücksichtigen um die Signifikanz einer ROC-Kurve bewerten zu können. [24]

Erstens ist es wichtig welche Werte verglichen werden, weil dies verschiedene Tests erfordern kann. Handelt es sich um die ganze Kurve, die mit einem globalen Index, wie die AUC verglichen wird, oder um einen regionalen oder lokalen Index? Der Unterschied, der verglichen werden soll, muss vor der Berechnung der statistischen Signifikanz gewählt werden. Sonst könnte das signifikanteste oder am wenigsten signifikanteste Ergebnis verschiedener Tests gewählt werden und dadurch die berichtete Signifikanz verfälscht werden. [24]

Zweitens muss beachtet werden wie die statistische Schwankung verschiedener Tests zustande kommt. Mit dem p-Wert wird ausgedrückt wie wahrscheinlich es ist, einen in einem Experiment gezeigten Unterschied oder einen größeren, wenn kein Unterschied zu der Population von der die Daten stammen besteht, festzustellen. Er drückt also aus wie wahrscheinlich es ist, ein solches Ergebnis oder ein Extremes unter Annahme der Nullhypothese zu erhalten. Diese Wahrscheinlichkeit, die durch

mehrmaliges Wiederholen des Tests unter denselben Bedingungen zustande kommt, kann als Teil der Ergebnisse verstanden werden. Die Bedingungen unter denen ein Test wiederholt werden soll, sind jedoch nicht klar definiert. Ein Experiment mit unabhängig ausgewählten Untersuchern und unabhängig ausgewählten Fällen könnte auf verschiedene Arten wiederholt werden. Die Untersucher und Fälle könnten wieder unabhängig ausgewählt werden, die Untersucher könnten unabhängig ausgesucht werden und die gleichen Fälle bewerten, die gleichen Untersucher könnten neue unabhängige Fälle bewerten oder die gleichen Untersucher könnten die gleichen Fälle bewerten. Mit diesen verschiedenen Möglichkeiten kann es zu großen Unterschieden der Ergebnisse bei der Wiederholung von Experimenten kommen. Die meisten Unterschiede sollten bei einem Austausch von Untersuchern und Fällen auftreten und die geringsten Unterschiede wenn Fälle und Untersucher gleich bleiben. [24]

Aus diesem Grund müssen bei der Auswahl eines statistischen Tests die verschiedenen Arten wie ein Ergebnis zustande kommen kann berücksichtigt werden. Einige statistische Tests berücksichtigen nur ein Wechseln der Untersucher, während andere nur eine Änderung der Fälle berücksichtigen. In die Berechnung der statistischen Signifikanz sollten darum beide Variationen einbezogen werden. Bei der Wiederholung von solchen Experimenten in verschiedenen Labors werden nämlich sowohl andere Untersucher als auch andere Fälle verwendet werden. Dies entspricht Multi-reader, Multi-case (MRMC) ROC-Analysen, welche immer bedeutender werden und in den meisten Studien zum Vergleich von ROC-Analysen genutzt werden. [24]

Es kann natürlich auch dazu kommen, dass kein statistisch signifikanter Unterschied zwischen ROC-Kurven bewiesen werden kann. Entweder weil tatsächlich kein Unterschied besteht oder weil die Fallzahl zu gering war um statistische Aussagekraft zu erzielen. Wenn kein statistisch signifikanter Unterschied entdeckt werden kann, soll das aber nicht bedeuten, dass kein Unterschied existiert. Darum sollen Maßstäbe angegeben werden, in denen sich der Unterschied befinden könnte. Diese Konfidenzintervalle dienen, wenn keine statistische Signifikanz gezeigt werden kann, als Hilfe bei der Interpretation von p-Werten. [24]

### 2.1.1.7 Idealer Grenzwert und Youden-Index

Um den idealen Grenzwert für einen Test zu bestimmen sind ROC-Kurven essentiell. Ziel ist es den Punkt mit der bestmöglichen richtigen Klassifizierung der Gruppen zu finden. Dabei gibt es unterschiedliche Ansätze den idealen Punkt dafür auf einer ROC-Kurve zu bestimmen. Die drei Methoden, die dafür genutzt werden, sind der  $d^2$ -Wert, die Kosten-Nutzen Rechnung und der Youden-Index. [25]

Der ideale Grenzwert oder „cut-off“ Punkt, der am besten zwischen positiven und negativen Fällen unterscheiden kann und die geringste Fehlerquote hat, ist der Punkt, welcher sich am nächsten zur linken oberen Ecke befindet. Dieser Punkt kann durch einen Vergleich der Abstände zwischen dem Punkt (0,1) und jedem Punkt auf der ROC-Kurve bestimmt werden. Der Abstand muss für jeden Punkt einer ROC-Kurve berechnet werden und der Punkt mit dem geringsten Abstand entspricht dem bestmöglichen „cut-off“ Wert. Bei einer idealen ROC-Kurve, welche durch die linke obere Ecke geht, wäre dieser Abstand somit 0. Die meisten Programme zur ROC-Analyse berechnen die Sensitivität und Spezifität für jeden Punkt einer ROC-Kurve, dadurch lässt sich der ideale Grenzwert leicht feststellen. Die Berechnung erfolgt folgendermaßen: [25, 26]

$$d^2 = [(1 - S_N)^2 + (1 - S_P)^2]$$

d = der Abstand zwischen dem Punkt (0,1) und einem beliebigen Punkt der Kurve

$S_N$  = die Sensitivität

$S_P$  = die Spezifität

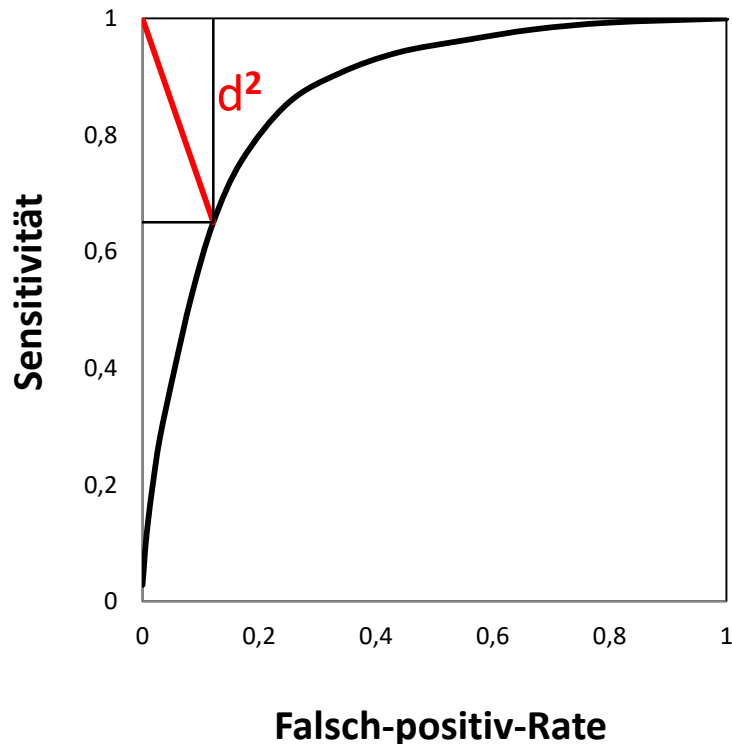


Abbildung 6: „cut-off“ Punkt  $d^2$

Kosten-Nutzen Rechnungen sind die zweite Möglichkeit den idealen „cut-off“ Punkt einer ROC-Kurve zu bestimmen. Dazu müssen die Kosten von allen möglichen Konsequenzen bekannt sein. Die Kosten beinhalten die Behandlungskosten aber auch Kosten von anderen Einflüssen auf die Gesundheit. Der Nutzen einer Konsequenz kann als negative Kosten angegeben werden. Wenn diese „Kosten“ für alle Möglichkeiten, richtig positive, richtig negative, falsch positive und falsch negative Entscheidungen bekannt sind, können zusammen mit den Durchführungskosten des Tests die Durchschnittskosten errechnet werden. Die jeweiligen Kosten einer Entscheidung werden dabei mit ihrer Wahrscheinlichkeit multipliziert und zu den Grundkosten addiert. Die Kosten eines Tests sind somit von der Sensitivität und Spezifität des Tests, der Prävalenz in der Bevölkerung und den Konsequenzen der Entscheidungen abhängig. [22]

$$\bar{C} = C_0 + C_{TP} \times P(TP) + C_{FP} \times P(FP) + C_{TN} \times P(TN) + C_{FN} \times P(FN)$$

$\bar{C}$  = durchschnittliche Kosten eines Tests

$C_0$  = Durchführungs- oder Grundkosten eines Tests

$C_{TP}$  = Kosten für Konsequenzen einer richtig positiven Entscheidung

$C_{FP}$  = Kosten für Konsequenzen einer falsch positiven Entscheidung

$C_{TN}$  = Kosten für Konsequenzen einer richtig negativen Entscheidung

$C_{FN}$  = Kosten für Konsequenzen einer falsch negativen Entscheidung

$P(TP)$  = Wahrscheinlichkeit dass eine richtig positive Entscheidung getroffen wird

$P(FP)$  = Wahrscheinlichkeit dass eine falsch positive Entscheidung getroffen wird

$P(TN)$  = Wahrscheinlichkeit dass eine richtig negative Entscheidung getroffen wird

$P(FN)$  = Wahrscheinlichkeit dass eine falsch negative Entscheidung getroffen wird

Der ideale Grenzwert einer solchen Kosten-Nutzen Rechnung wird erreicht, wenn die durchschnittlichen Kosten minimal und der durchschnittliche Nutzen maximal sind. Für die Berechnung des Grenzwertes sind die Grundkosten eines Tests jedoch nicht relevant. Sie spielen nur eine Rolle beim Vergleich von Tests, weil die durchschnittlichen Kosten durch sie erhöht werden. Der Punkt mit den geringsten Kosten und größtem Nutzen einer ROC-Kurve ist bei folgender Steigung gegeben:  
[22]

$$m = \frac{P(D-)}{P(D+)} \times \frac{[C_{FP} - C_{TN}]}{[C_{FN} - C_{TP}]}$$

$m$  = Steigung einer ROC-Kurve am Punkt mit idealen Grenzwert

$P(D+)$  = Prävalenz einer Erkrankung in der Bevölkerung

$P(D-)$  = Anteil der Bevölkerung ohne die Erkrankung

Für diese Berechnung müssen die Sensitivität und die Spezifität nicht beachtet werden, die Prävalenz einer Erkrankung in der Bevölkerung und die Kosten der Entscheidungen reichen vollkommen aus. Um zu verstehen wo sich der ideale „cut-off“ Punkt einer ROC-Kurve befindet, muss nur die Form einer ROC-Kurve beachtet werden. Die Steigung einer konventionellen ROC-Kurve ist anfangs in der linken unteren Ecke groß und nimmt gegen rechts oben immer weiter ab. Wenn die Prävalenz einer Erkrankung in einer Bevölkerung sehr hoch ist, die Erkrankung also häufig ist, wird die Rate  $P(D-)/P(D+)$  kleiner als 1 und die Steigung, bei welcher der Punkt mit den geringsten Kosten und größtem Nutzen einer ROC-Kurve liegt, gering. Solche Steigungen erreicht eine ROC-Kurve in der rechten oberen Ecke. Die Sensitivität und Falsch-positiv-Rate müssen dabei dementsprechend hoch sein. Wenn im Gegensatz dazu die Prävalenz einer Erkrankung sehr niedrig ist, die Erkrankung also selten ist, wird die  $P(D-)/P(D+)$  Rate größer als 1 und die Steigung, bei welcher der Punkt mit den geringsten Kosten und größtem Nutzen einer ROC-Kurve liegt, groß. Die größte Steigung hat eine ROC-Kurve am Anfang in der linken unteren Ecke, wo die Sensitivität und Falsch-positiv-Rat gering sind. [22]

Der ideale „cut-off“ Punkt hängt aber auch von den einzelnen Kosten der Konsequenzen ab. Beim genauen Betrachten der Formel wird klar, dass hohe Kosten von falsch negativen Entscheidungen und geringe Kosten von richtig negativen Entscheidung ebenfalls zu einer großen Steigung führen. Wenn  $C_{FP} - C_{TN}$  groß ist, also eine Behandlung und weitere diagnostische Tests teuer oder sogar schädigend für gesunde Patienten und Patientinnen sind, wird ein Punkt mit geringer Sensitivität und geringer Falsch-positiv-Rate ideal sein. Ein Punkt mit hoher Sensitivität und hoher Falsch-positiv-Rate wird hingegen ideal sein, wenn  $C_{FN} - C_{TP}$  größer ist. Dies ist der Fall, wenn die Behandlung und weitere diagnostische Untersuchungen günstig sind und keinen Schaden bei gesunden Patienten und Patientinnen anrichten. Das Nichtentdecken von Erkrankten stellt dabei einen großen Schaden dar und der Nutzen durch das Entdecken der Erkrankung ist sehr groß. Dies kann mit negativen Kosten für  $C_{TP}$  ausgedrückt werden. [22]

Mit dem ermittelten Punkt, bei dem die Kosten minimal sind, können diese auch ausgerechnet werden. Die minimalen Durchschnittskosten eines Tests  $\bar{C}_{\min}$  können dann zum Vergleich von unterschiedlichen Tests als Index verwendet werden. Außerdem könnte der Nettonutzen eines Tests ausgerechnet werden. [22]

Die dritte und gängigste Möglichkeit zur Bestimmung eines „cut-off“ Punktes ist der Youden-Index, der nach dem Erfinder benannt ist. Er wird mit J angegeben und zeigt, wenn er den maximalen Wert bei einer ROC-Kurve erreicht, den idealen „cut-off“ Punkt. Er wird berechnet indem die Falsch-negativ-Rate von der Sensitivität abgezogen wird, um den Erfolg eines Tests bei Erkrankten zu bemessen. Genauso wird die Falsch-positiv-Rate von der Spezifität abgezogen, um den Erfolg des Tests bei Gesunden zu bemessen. Dann wird der Mittelwert dieser zwei Werte berechnet was den Youden-Index ergibt. [27]

$$J = \frac{1}{2} \times \left[ \frac{a - b}{a + b} + \frac{d - c}{c + d} \right] = \frac{a}{a + b} + \frac{d}{c + d} - 1$$

a, b, c und d stehen wie in Tabelle 9 gezeigt für jeweils eine der vier Gruppen die durch den Krankheitsstatus und das Testresultat definiert sind.

Dabei zeigt sich  $J = \text{Sensitivität} + \text{Spezifität} - 1$ . Der Index reicht theoretisch von -1 bis 1 wobei nur Werte über 0 sinnvoll sind. Wenn die Falsch-positiv-Rate und die Sensitivität gleich hoch sind, ist der Youden-Index 0. Er ist außerdem unabhängig von den relativen und absoluten Größen der Gruppen. Tests mit dem gleichen Youden-Index machen gleich viele Fehler pro 100 Patienten und Patientinnen. [27]

Der Punkt einer ROC-Kurve mit dem größten Youden-Index befindet sich dort, wo der größte vertikale Abstand zwischen der Kurve und der Diagonalen, die einem zufälligen Entscheiden entspricht, erreicht wird. [25]

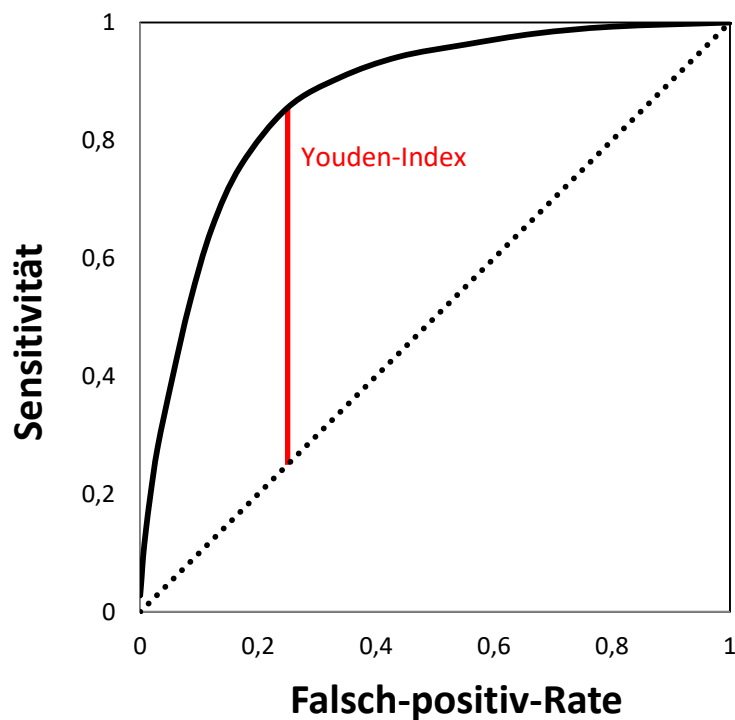


Abbildung 7: Youden-Index

### 2.1.2 Variable life-adjusted display (VLAD) und Leistungsverhältnis

Das Variable life-adjusted display oder kurz VLAD stellt eine Methode zur Qualitätssicherung oder genauer zur Leistungsbewertung dar, bei der das individuelle Risiko von Patienten und Patientinnen berücksichtigt wird. Dabei wird der Unterschied zwischen erwarteten Todesfällen und tatsächlichen Todesfällen bei einer Therapie oder Operation bestimmt. Davor muss allerdings ermittelt werden, ob die erwartete Mortalität verlässlich ist, ansonsten ist die Bewertung nicht sinnvoll. Dies kann unter anderem durch ROC-Analysen festgestellt werden. Die kumulative Summe des Unterschieds zwischen erwarteter und tatsächlicher Mortalität wird dann graphisch auf der Y-Achse gegen die individuellen Fälle auf der X-Achse, in der Reihenfolge des Auftretens, dargestellt. Es muss sich jedoch nicht immer um den Unterschied zwischen erwarteten Todesfällen und tatsächlichen Todesfällen handeln. Das VLAD kann genauso gut für jeden anderen Vergleich zwischen binären

Outcomes genutzt werden, wie Auftreten von Infektionen oder Aufhalten auf der Intensivstation. [28, 29]

Um den Unterschied zwischen erwarteten Todesfällen und tatsächlichen Todesfällen zu berechnen, muss die erwartete Mortalität für jeden Patienten und jede Patientin bekannt sein oder kalkuliert werden. Sie liegt immer zwischen 0 und 1, ein Mortalitätsrisiko von 11 % entspricht dabei einem Wert von 0,11. Von diesem Wert wird dann das Outcome abgezogen. Das Outcome beträgt 0 wenn es nicht eintritt, also bei einem Überleben und 1 wenn es eintritt, es zu einem Todesfall kommt. Diese Berechnung wird für jeden Patienten und jede Patientin durchgeführt und alle Werte addiert. [30, 29]

$$C_n = \sum_{i=1}^n Y_i - \sum_{i=1}^n X_i$$

$C_n$  = kumulativer Unterschied der Mortalität

$Y_i$  = erwartetes Mortalitätsrisiko

$X_i$  = Outcome einer Operation, 1 für einen Todesfall und 0 für das Überleben

Zusammengerechnet ergibt dies den kumulativen Unterschied zwischen erwarteter und tatsächlicher Mortalität, welcher graphisch in der Reihenfolge des Auftretens dargestellt wird. Die Nulllinie entspricht hierbei der erwarteten Mortalität. Die tatsächliche Mortalität zeigt, über den beobachteten Zeitraum, wie viele Leben gerettet werden oder auch verloren gehen, sie gibt sozusagen im Vergleich die Leistung wieder. Wenn ein Patient oder eine Patientin mit niedrigem Risiko zu sterben überlebt, wird dies den Graphen nur um ein kleines Stück heben. Hingegen wird der Graph stärker angehoben, wenn ein Patient oder eine Patientin mit hohem Risiko zu sterben überlebt. Umgekehrt wenn ein Patient oder eine Patientin mit

hohem Risiko stirbt wird der Graph dementsprechend gesenkt werden und wenn ein Patient oder eine Patientin mit niedrigem Risiko stirbt wird der Graph stark gesenkt werden. [30]

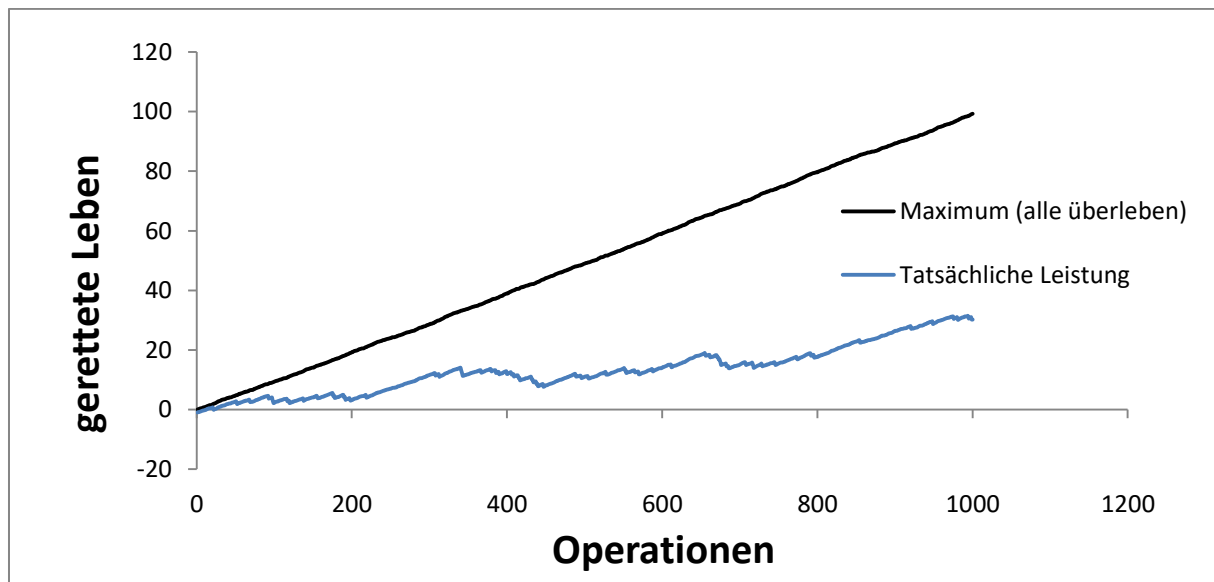


Abbildung 8: VLAD

Das Leistungsverhältnis auch performance ratio ist ein Vergleichsmaßstab, der anhand einer Zahl wiedergibt, wie gut eine Krankenanstalt, eine Abteilung oder ein Operateur oder eine Operateurin im Vergleich ist. Die Berechnung erfolgt indem die tatsächliche Mortalität durch die mittlere erwartete Mortalität dividiert wird. Wenn die performance ratio unter 1 ist, sterben weniger Patienten und Patientinnen als erwartet, dies entspricht einer guten Leistung. Bei einem Wert über 1 sterben mehr Patienten und Patientinnen als erwartet, die Leistung ist unterdurchschnittlich.

$$performance\ ratio = \frac{beobachtete\ Mortalität}{mittlere\ erwartete\ Mortalität}$$

### 2.1.3 Datenreview, Zulassung und verwendete statistische Methoden

Alle Daten stammen von Operationen im Universitätsklinikum Graz, und wurden aus dem QS II, einem Qualitätssicherungssystem, nachdem sie von Operateuren und Operateurinnen eingetragen wurden, entnommen. Standardmäßig wird die Qualität der Daten zuerst hausintern überprüft und dann durch Herrn Dr. Roland Schistek in Salzburg, welcher Beauftragter der Österreichischen Gesellschaft für Herz-Thoraxchirurgie ist. Als letzter Schritt werden die Daten durch die Arbeitsgruppe des ÖBIG/GÖG in einem externen Audit validiert, bevor sie zur Verwendung freigegeben werden.

Diese Studie beschreibt die Performance des logistischen EuroSCORE und des EuroSCORE II mit Hilfe der Diskriminierungsstärke und Kalibrierung. Die risikobereinigte Mortalität wurde errechnet, indem die beobachtete Mortalität durch die erwartete Mortalität dividiert wurde. Die Genauigkeit der verschiedenen Scores wurde durch ROC-Analysen veranschaulicht. Diese sind mit ihrer AUC und einem Konfidenzintervall von 95% angegeben. Durch den Youden-Index und den  $d^2$ -Wert wurde die Diskriminationskraft bestimmt und die idealen „cut-off“ Werte mit der jeweiligen Sensitivität und Spezifität berechnet.

Quantitative Variablen wurden anhand von Mittelwerten und deren Standardabweichung sowie Median, Minimum und Maximum angegeben. Kategoriale Variablen wurden mit ihren anteiligen Prozentwerten und absoluten Häufigkeiten angegeben.

Die meisten Berechnungen wurden mit IBM SPSS, einem Statistikprogramm, durchgeführt und dann zur besseren Formatierung in Microsoft Excel übertragen. Der Youden-Index, der  $d^2$ -Wert und die performance ratio wurden in Microsoft Excel berechnet, die variable life-adjusted displays wurden ebenso mit Microsoft Excel erstellt.

Die Ethikkommission der Medizinischen Universität Graz hat die Studie geprüft und mit der EK-Nummer 29-516 ex 16/17 unter dem Studientitel „Validation of EuroSCORE II risk model for aortic valve replacement – a retrospective study“ approbiert (siehe Appendix 1 – Ethikvotum).

### 3 Ergebnisse

#### 3.1 Demographische Daten

Das gesamte Patienten- und Patientinnenkollektiv bestand aus 843 Patienten und Patientinnen, welche jeweils einen Aortenklappenersatz am LKH Graz zwischen 2012 und 2016 erhielten. Das Patienten- und Patientinnenalter variierte stark, mit einem Minimum von 14 und einem Maximum von 92 Jahren. Von den 843 Patienten und Patientinnen waren 499 (59,20%) männlich und 344 (40,80%) weiblich. Eine bakterielle Endokarditis hatten 38 (4,50%), an chronischen Lungenerkrankungen litten 282 (33,50%) und periphere vaskuläre Verschlusskrankheiten konnten bei 80 (9,50%) Patienten und Patientinnen nachgewiesen werden. Bei 319 (37,80%) konnte eine Form von Angina Pectoris festgestellt werden und 17 (2,00%) hatten einen kardiogenen Schock.

		MW	SD	Med	Min	Max	N	(%)
Anzahl							843	
Alter		70	12	72	14	92		
Geschlecht	männlich						499	59,20%
	weiblich						344	40,80%
Bakterielle Endokarditis	nein						805	95,50%
	ja						38	4,50%
Chronische Lungenerkrankung/COPD	nein						561	66,50%
	schwach						131	15,50%
	moderat						122	14,50%
	schwer						29	3,40%
Periphere vaskuläre Verschlusskrankungen/PAVK	nein						763	90,50%
	ja						80	9,50%
Angina Pectoris	keine						524	62,20%
	stabil						303	35,90%
	instabil						16	1,90%
Kardiogener Schock	nein						826	98,00%
	ja						17	2,00%

Tabelle 12: Risikofaktoren I

Nur 22 (2,60%) wurden nach NYHA – Klassifikation in Stadium I eingeteilt, 264 (31,30%) wurden in Stadium II, die Mehrheit mit 506 (60,00%) in Stadium III und 51 (6,00%) in Stadium IV eingeteilt. Eine Aortenklappeninsuffizienz war bei 430 (51,00%) und eine Aortenklappenstenose bei 736 (87,30%) nachweisbar. Die durchschnittliche linksventrikuläre Auswurffraktion betrug  $55\% \pm 11\%$ . Elektive Operationen waren mit 764 (90,60%) der größte Anteil, 42 (5,00%) waren dringende Eingriffe, 27 (3,20%) Notfälle und bei 10 (1,20%) Operationen war die Dringlichkeit lebensbedrohend. Eine biologische Aortenklappe erhielten 669 (79,40%) und eine mechanische Aortenklappe 174 (20,60%). Von den verwendeten Aortenklappen waren 62 (7,40%) sutureless.

		MW	SD	Med	Min	Max	N	(%)
NYHA - Klassifikation	I						22	2,60%
	II						264	31,30%
	III						506	60,00%
	IV						51	6,00%
Aortenklappeninsuffizienz	keine						413	49,00%
	minimal						163	19,30%
	leicht						122	14,50%
	mittelgradig						85	10,10%
	hochgradig						60	7,10%
Aortenklappenstenose	nein						107	12,70%
	ja						736	87,30%
Auswurffraktion (%)		55	11	60	15	77		
Dringlichkeit	elektiv						764	90,60%
	dringend						42	5,00%
	Notfall						27	3,20%
	lebensbedr.						10	1,20%
Implantat Typ	biologisch						669	79,40%
	mechanisch						174	20,60%
Sutureless	nein						781	92,60%
	ja						62	7,40%

Tabelle 13: Risikofaktoren II

		N	(%)
OP-Jahr	2012	167	19,8(%)
	2013	168	19,9(%)
	2014	172	20,4(%)
	2015	156	18,5(%)
	2016	180	21,4(%)
30-Tages-Status	lebend	825	97,9(%)
	tot	18	2,1(%)
Status	lebend	799	94,8(%)
	tot	32	3,8(%)
	unbekannt	12	1,4(%)

Tabelle 14: Status und 30-Tage-Status

Von 2012 bis 2016 wurden pro Jahr im Schnitt 169 Patienten und Patientinnen operiert, wobei die Anzahl im Jahr 2016 mit 180 am höchsten war. Innerhalb des Beobachtungszeitraumes von 30 Tagen nach erfolgter Operation verstarben 18 Patienten und Patientinnen, dies entspricht 2,1%. Insgesamt ist das Versterben von 32 Patienten und Patientinnen bekannt, der Status von 12 konnte nicht erhoben werden.

	MW	SD	Med	Min	Max
Alter	70	12	72	14	92
Add. EuroSCORE	6,65	2,88	7	2	20
Log. EuroSCORE	9,07	11,09	6,2	1,5	88,5
EuroSCORE II	3,64	6,88	1,8	0,5	82,9

Tabelle 15: Alter und Scores

Der durchschnittliche additive EuroSCORE betrug  $6,65\% \pm 2,88\%$ , der durchschnittliche logistische EuroSCORE  $9,07\%$  mit der höchsten Standardabweichung der drei Scores von 11,09 und der durchschnittliche EuroSCORE II  $3,64\% \pm 6,88\%$ .

		OP-Jahr									
		2012		2013		2014		2015		2016	
		N	(%)	N	(%)	N	(%)	N	(%)	N	(%)
30-Tages-Status	lebend	164	98,2	166	98,8	164	95,3	155	99,4	176	97,8
	tot	3	1,8	2	1,2	8	4,7	1	0,6	4	2,2

Tabelle 16: 30-Tages-Status nach OP-Jahr

Den höchsten Wert hatte der mittlere additive EuroSCORE im Jahr 2014 mit  $7\% \pm 3,2\%$  und die niedrigsten Werte im Jahr 2012 mit  $6,4\% \pm 2,8\%$  und im Jahr 2015 mit  $6,4\% \pm 2,9\%$ . Der mittlere logistische EuroSCORE zeigte eine ähnliche Verteilung mit dem höchsten Wert im Jahr 2014 mit  $10,2\% \pm 14,2\%$  und dem niedrigsten im Jahr 2012 mit  $8\% \pm 9,1\%$ . Ebenso war der EuroSCORE II im Jahr 2014 am höchsten mit  $4,6\% \pm 9,7\%$  und hatte den niedrigsten Wert im Jahr 2015 mit  $2,9\% \pm 4,6\%$ . Dies stimmt mit der höchsten 30-Tages-Mortalität im Jahr 2014 von 4,70% überein. Die niedrigste 30-Tages-Mortalität war im Jahr 2015 mit 0,6% zu verzeichnen. Ein Zusammenhang der höchsten Mittelwerte mit dem durchschnittlichen Alter der Patienten und Patientinnen ist gegeben. Das Alter war ebenso am höchsten im Jahr 2014 gleich wie im Jahr 2013 und betrug  $71 \pm 12$ . Jedoch war das niedrigste durchschnittliche Alter im Jahr 2016 bei  $68 \pm 13$ , was mit keinem der Scores übereinstimmt.

		MW	SD	Med	Min	Max
Alter	2012	70	12	72	20	90
	2013	71	12	74	24	88
	2014	71	12	73	19	92
	2015	69	12	71	29	87
	2016	68	13	72	14	91
Add. EuroSCORE	2012	6,4	2,8	6	2	18
	2013	6,6	2,7	7	2	20
	2014	7	3,2	7	2	20
	2015	6,4	2,9	6	2	16
	2016	6,8	2,8	7	2	16
Log. EuroSCORE	2012	8	9,1	5,6	1,5	78
	2013	8,6	10,5	6,2	1,5	87,2
	2014	10,2	14,2	6,4	1,5	88,5
	2015	8,6	9,8	5,5	1,5	55,6
	2016	9,8	10,8	6,4	1,5	62,6
EuroSCORE II	2012	3	4,5	1,7	0,6	37,6
	2013	3,2	5,6	1,8	0,6	59,7
	2014	4,6	9,7	1,9	0,6	82,9
	2015	2,9	4,6	1,5	0,5	33,6
	2016	4,4	8,1	1,9	0,6	63,1

Tabelle 17: Alter und Scores nach OP-Jahr

### 3.2 ROC-Analysen: logistischer EuroSCORE und EuroSCORE II für den Zeitraum 2012 bis 2016

Zur Bestimmung des Vorhersagewertes des logistischen EuroSCORE und des EuroSCORE II über die 30-Tages-Mortalität wurden die jeweiligen ROC-Analysen erstellt. Der logistische EuroSCORE weist dabei eine AUC von 0,847 mit einer Untergrenze von 0,761 und einer Obergrenze von 0,933 bei einem 95% Konfidenzintervall. Für den EuroSCORE II ergibt sich bei einem 95% Konfidenzintervall eine AUC von 0,829 mit einer Untergrenze von 0,733 und einer Obergrenze von 0,926. Beide Ergebnisse zeigen über den gesamten Zeitraum von 2012 bis 2016 statistische Signifikanz mit einem p-Wert unter 0,001.

	AUC	Untergrenze	Obergrenze	p-Wert
Add. EuroSCORE	0,837	0,752	0,921	<0,001
Log. EuroSCORE	0,847	0,761	0,933	<0,001
EuroSCORE II	0,829	0,733	0,926	<0,001

Tabelle 18: AUC-Werte der Scores

Bei der Berechnung des idealen Grenzwertes ergibt sich für den logistischen EuroSCORE ein Youden-Index von 0,61, was einem „cut-off“-Wert von 8,95, einer Sensitivität von 0,889 und einer Spezifität von 0,721 entspricht. Der höchste Youden-Index von 0,554 für den EuroSCORE II ist bei einem „cut-off“-Wert von 2,65 erreicht, was einer Sensitivität von 0,889 und einer Spezifität von 0,665 entspricht.

	Youden-Index	cut-off	Sensitivität	Spezifität
Add. EuroSCORE	0,571	7,5	0,889	0,682
Log. EuroSCORE	0,61	8,95	0,889	0,721
EuroSCORE II	0,554	2,65	0,889	0,665

Tabelle 19: Youden-Index der Scores

Der geringste Abstand zum Punkt (0,1) der ROC-Kurve des logistischen EuroSCORE beträgt 0,284 und wird bei einem „cut-off“-Wert von 10,05 mit einer Sensitivität von 0,833 und einer Spezifität von 0,77 erreicht. Für den EuroSCORE II ist der geringste Abstand zum Punkt (0,1) 0,341, welcher bei einem „cut-off“-Wert von 2,95 mit einer Sensitivität von 0,833 und einer Spezifität von 0,703 erreicht wird

	d-Wert	cut-off	Sensitivität	Spezifität
Add. EuroSCORE	0,337	7,5	0,889	0,682
Log. EuroSCORE	0,284	10,05	0,833	0,77
EuroSCORE II	0,341	2,95	0,833	0,703

Tabelle 20: d-Wert der Scores

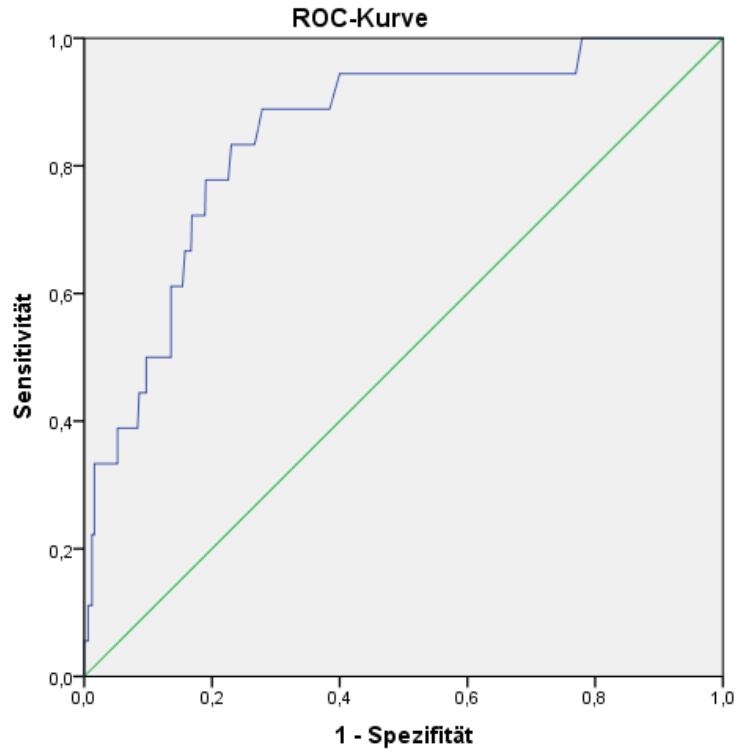


Abbildung 9: ROC-Kurve logistischer EuroSCORE 2012-2016

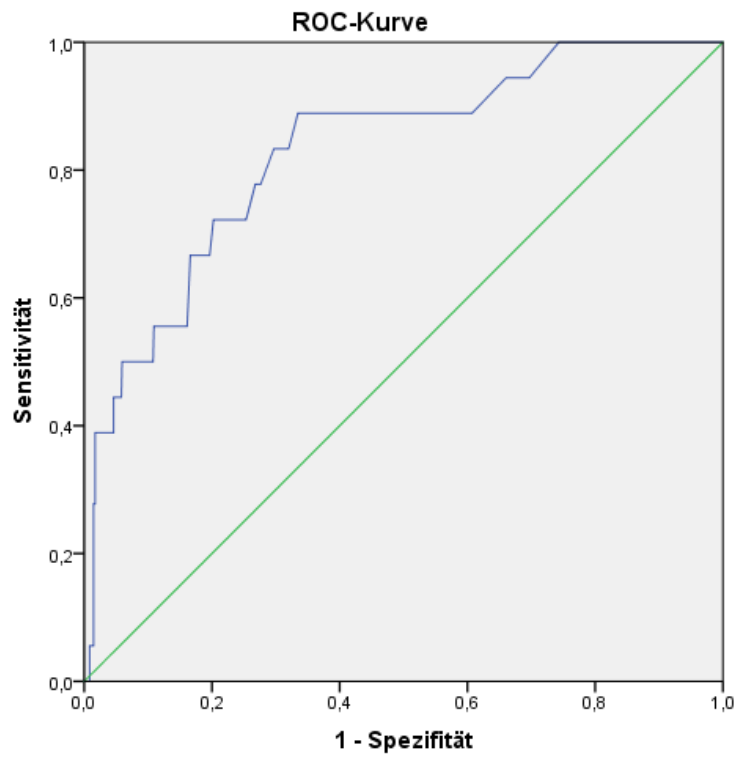


Abbildung 10: ROC-Kurve EuroSCORE II 2012-2016

Die Aufschlüsselung der Ergebnisse nach Einzeljahren:

		AUC	Untergrenze	Obergrenze	p-Wert
2012	Log. EuroSCORE	0,609	0,3	0,918	0,519
	EuroSCORE II	0,555	0,226	0,884	0,745
2013	Log. EuroSCORE	0,925	0,823	1	0,039
	EuroSCORE II	0,875	0,705	1	0,069
2014	Log. EuroSCORE	0,922	0,85	0,994	<0,001
	EuroSCORE II	0,907	0,821	0,992	<0,001
2015	Log. EuroSCORE	1	1	1	0,085
	EuroSCORE II	0,994	0,981	1	0,089
2016	Log. EuroSCORE	0,782	0,718	0,846	0,054
	EuroSCORE II	0,789	0,695	0,883	0,048

Tabelle 21: AUC-Werte nach OP-Jahr

Die Aufschlüsselung nach Einzeljahren zeigt eine klare statistische Signifikanz für 2014, das Jahr mit der höchsten 30-Tages-Mortalität. Zusätzlich ist eine statistische Signifikanz für den logistischen EuroSCORE im Jahr 2013 und den EuroSCORE II im Jahr 2016 nachweisbar. Die restlichen p-Werte befinden sich knapp über dem Signifikanzniveau von 5%. Eine Ausnahme stellt das Jahr 2012 dar, bei dem der p-Wert sowohl für den logistischen EuroSCORE als auch den EuroSCORE II weit über dem Signifikanzniveau liegt. Ebenso beträgt die AUC für das Jahr 2012 für den logistischen EuroSCORE und den EuroSCORE II annähernd 0,5 und entspricht fast der Diagonale zwischen der linken untern und rechten oberen Ecke. Die Schwankungsbreite der beiden Scores ist im Jahr 2012 ebenfalls sehr groß. Die Untergrenze der AUC für den logistischen EuroSCORE im Jahr 2012 bei einem Konfidenzintervall von 95% liegt bei 0,3 und die Obergrenze bei 0,918. Für den EuroSCORE II liegt die Untergrenze der AUC bei einem Konfidenzintervall von 95% bei 0,226 und die Obergrenze bei 0,884.

Die ROC-Kurve des logistischen EuroSCORE im Jahr 2015, Abbildung 17, entspricht einer idealen ROC-Kurve mit einer AUC von 1. Der p-Wert beträgt 0,085 womit allerdings keine statistische Signifikanz bewiesen werden kann. Einen ähnlichen Verlauf zeigt die ROC-Kurve des EuroSCORE II im Jahr 2015, Abbildung 18, welche eine AUC von 0,994 bei einem p-Wert von 0,089 aufweist. Der Grund dafür ist, dass 2015 nur eine Person verstorben ist.

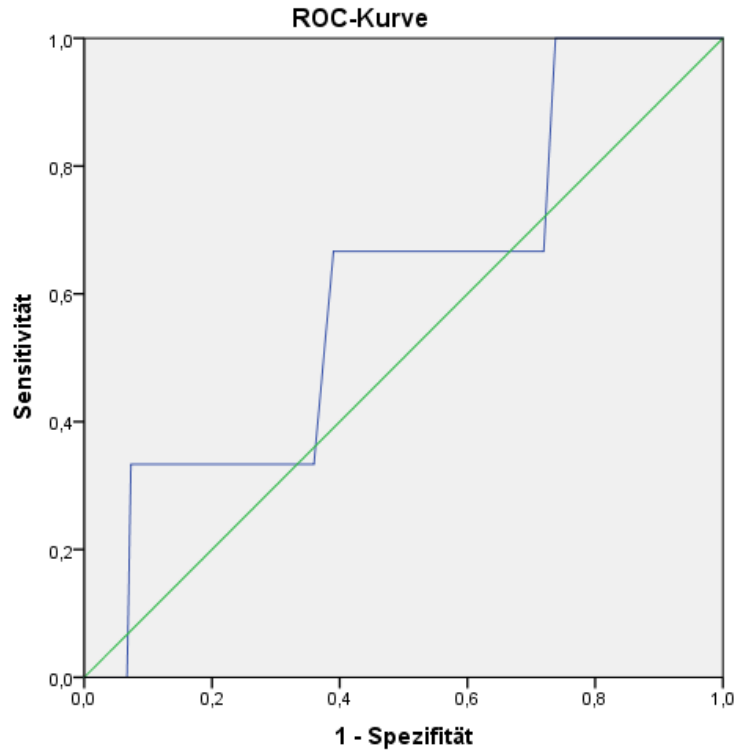


Abbildung 11: ROC-Kurve logistischer EuroSCORE 2012

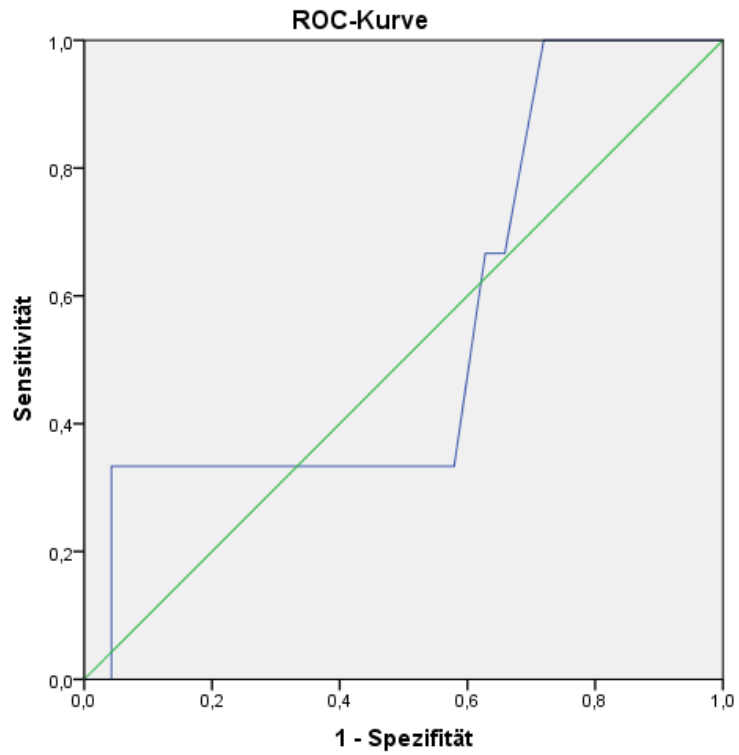


Abbildung 12: ROC-Kurve EuroSCORE II 2012

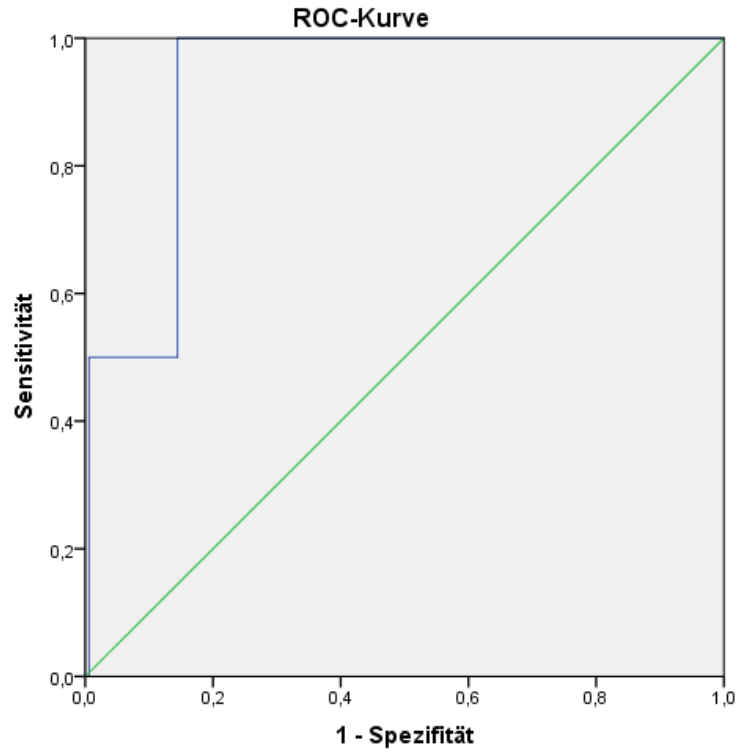


Abbildung 13: ROC-Kurve logistischer EuroSCORE 2013

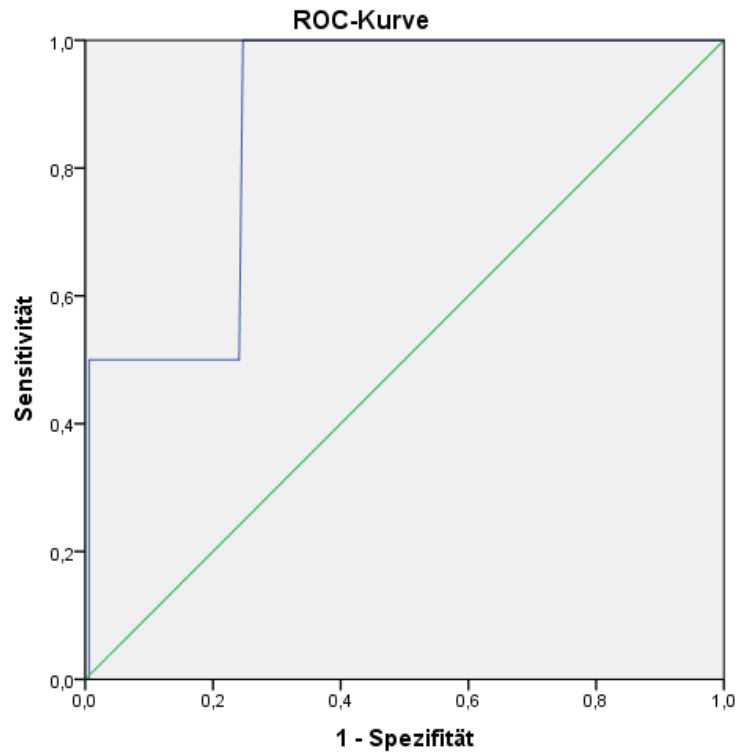


Abbildung 14: ROC-Kurve Euroscore II 2013

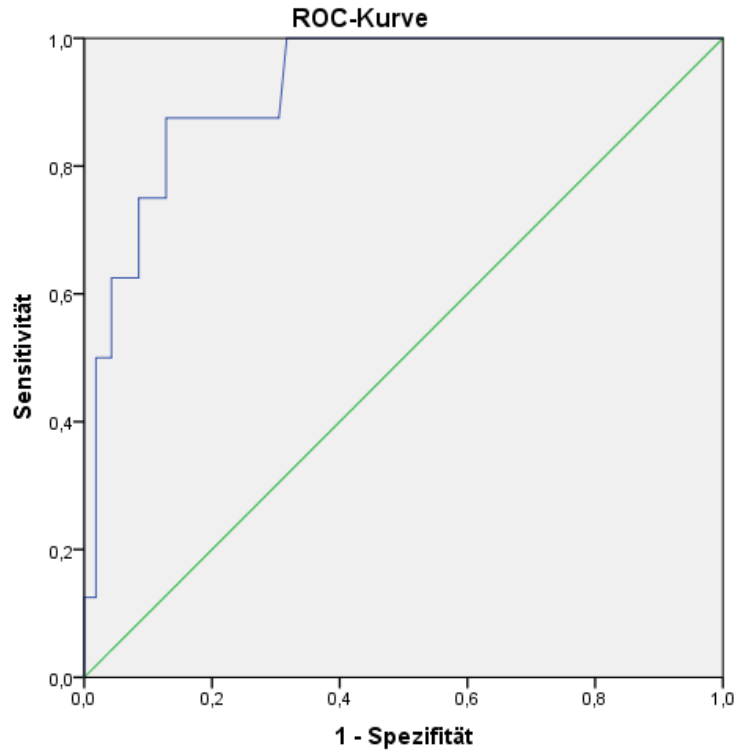


Abbildung 15: ROC-Kurve logistischer EuroSCORE 2014

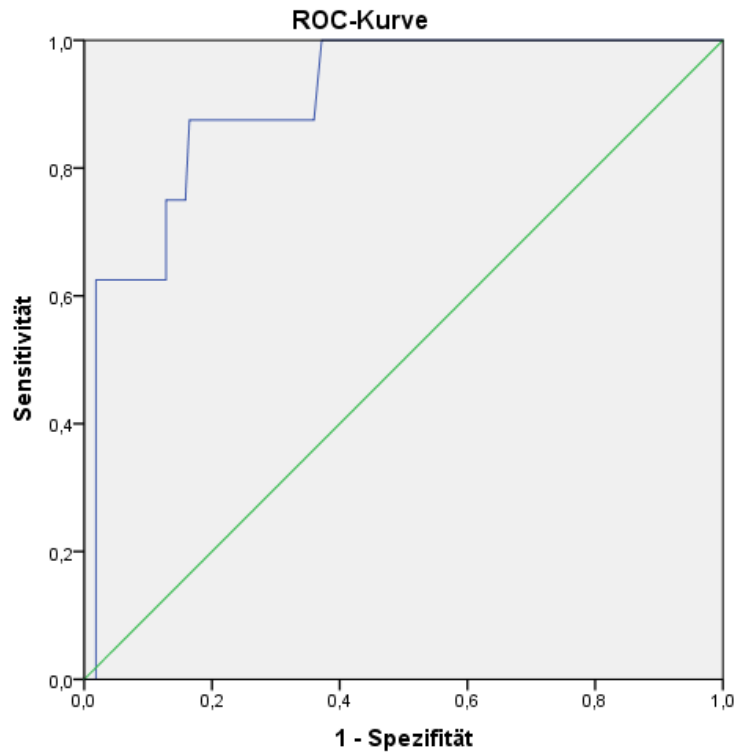


Abbildung 16: ROC-Kurve EuroSCORE II 2014

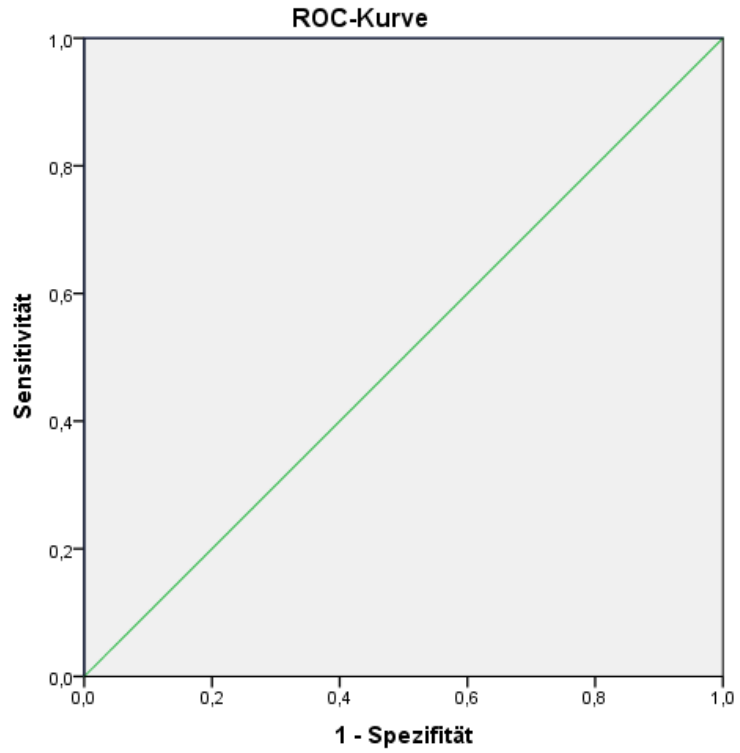


Abbildung 17: ROC-Kurve logistischer EuroSCORE 2015

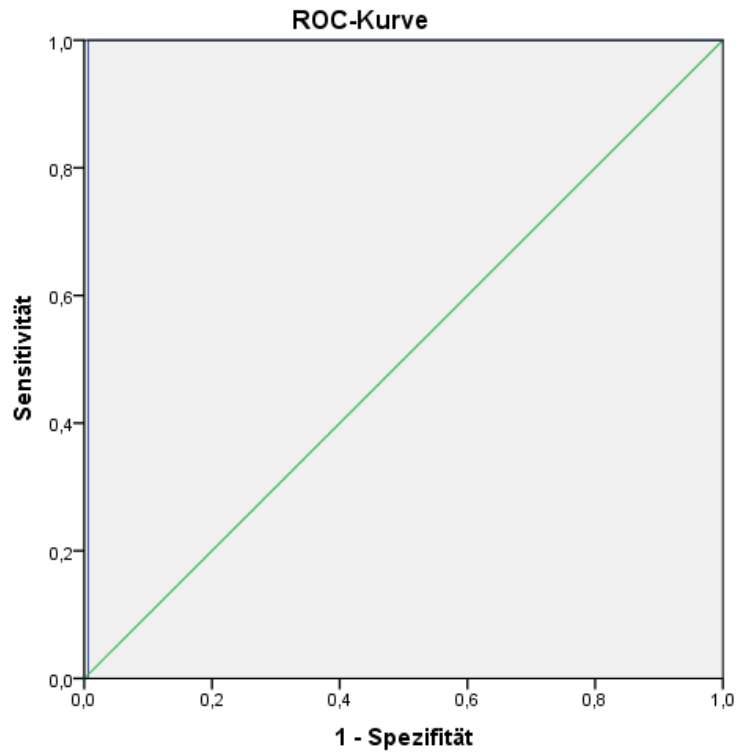


Abbildung 18: ROC-Kurve EuroSCORE II 2015

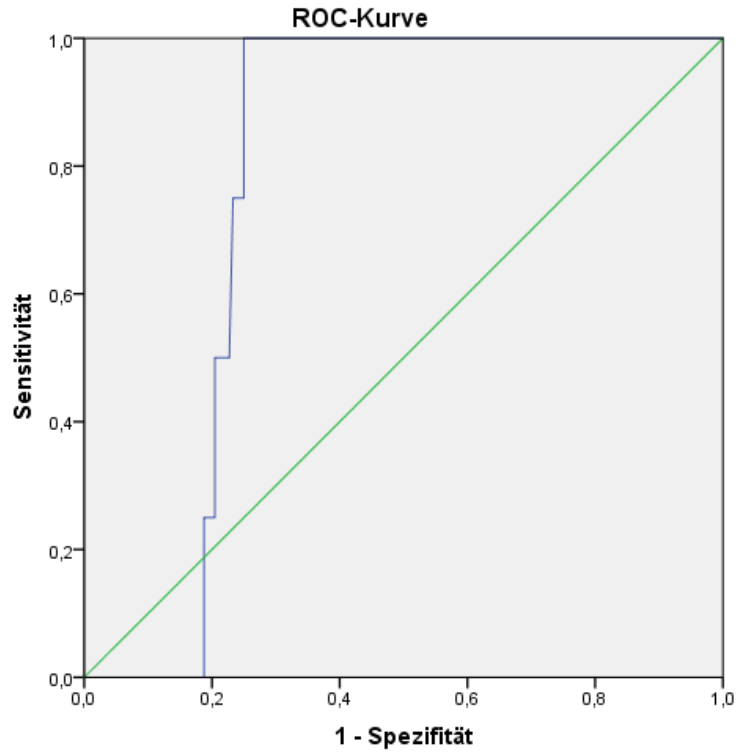


Abbildung 19: ROC-Kurve logistischer EuroSCORE 2016

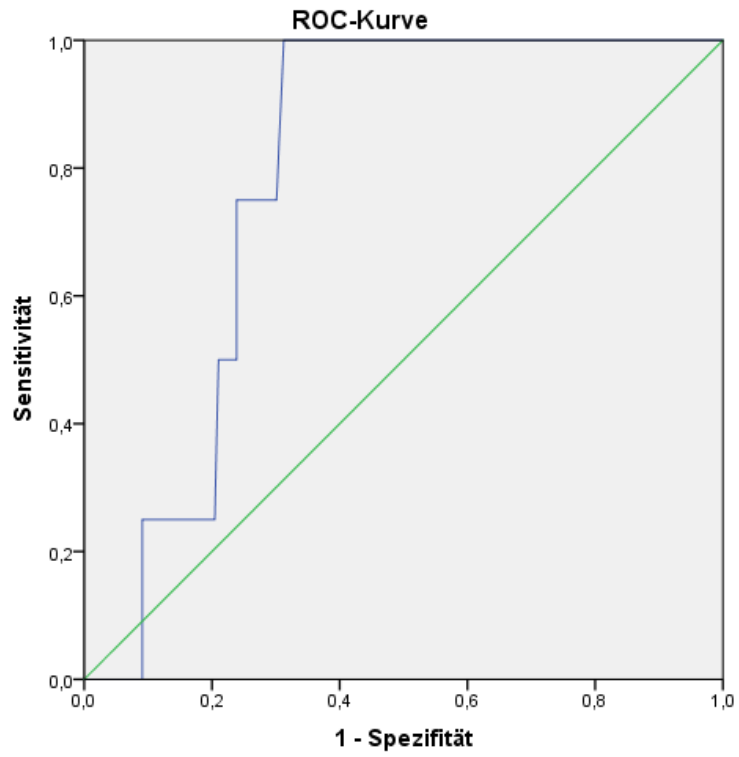


Abbildung 20: ROC-Kurve EuroSCORE II 2016

### 3.3 Variable life-adjusted display (VLAD) und performance ratio

Zur Übersicht ist nochmal der 30-Tages-Status, diesmal für Patienten und Patientinnen separat, für den Zeitraum von 2012-2016 angeführt.

		Gesamt	männlich	weiblich
30-Tages- Status	lebend	825	490	335
	tot	18	9	9
Summe		843	499	344

Tabelle 22: Übersicht 30-Tages-Status

#### 3.3.1 Variable life-adjusted display (VLAD) und performance ratio für den logistischen EuroSCORE (2012-2016)

Tabelle 23 zeigt die berechnete performance ratio für den logistischen EuroSCORE für den Zeitraum von 2012 bis 2016. Die performance ratio ist mit 0,23 sehr niedrig, was einem außerordentlich positiven Ergebnis entspricht.

	Gesamt	männlich	weiblich
OP-Anzahl	843	499	344
Verstorbene	18	9	9
Verstorbene (%)	2,10	1,80	2,60
durchschnittlicher Score (%)	9,07	8,19	10,34
performance ratio	0,23	0,22	0,25

Tabelle 23: Performance ratio des logistischen EuroSCORE

Das VLAD für den logistischen EuroSCORE für den Zeitraum von 2012-2016 ist in Abbildung 21 zu sehen. Es zeigt eine sehr gute Leistung, die der maximalen Leistung sehr nahe kommt.

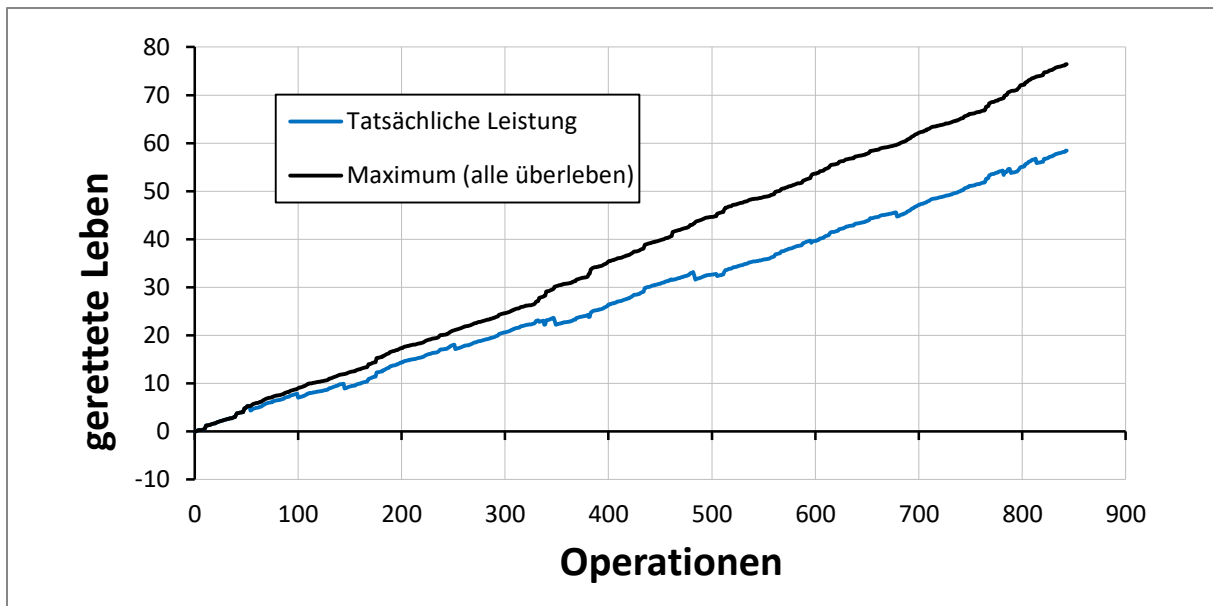


Abbildung 21: VLAD für den logistischen EuroSCORE 2012-2016

### 3.3.2 Variable life-adjusted display (VLAD) und performance ratio für den EuroSCORE II (2012-2016)

Tabelle 24 zeigt die berechnete performance ratio für den EuroSCORE II für den Zeitraum von 2012 bis 2016. Die performance ratio ist mit 0,58 ebenfalls sehr niedrig, zwar nicht so niedrig wie für den logistischen EuroSCORE, jedoch entspricht auch ein Wert von 0,58 einem außerordentlich positiven Ergebnis.

	Gesamt	männlich	weiblich
OP-Anzahl	843	499	344
Verstorbene	18	9	9
Verstorbene (%)	2,10	1,80	2,60
durchschnittlicher Score (%)	3,64	3,52	3,82
performance ratio	0,58	0,51	0,68

Tabelle 24: Performance ratio für den EuroSCORE II

Das VLAD für den EuroSCORE II für den Zeitraum von 2012-2016 ist in Abbildung 22 zu sehen. Es zeigt eine sehr gute Leistung, die stärker von der maximalen Leistung abweicht als die Leistung des logistischen EuroSCORE, trotzdem ist die Leistung immer noch bemerkenswert.

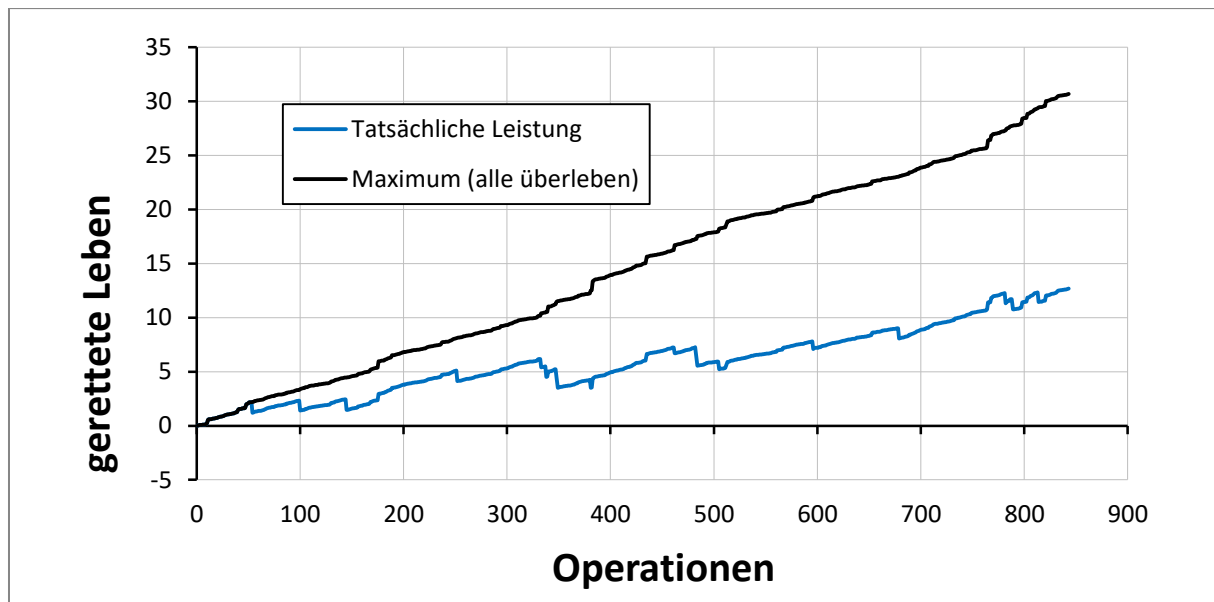


Abbildung 22: VLAD für den EuroSCORE II 2012-2016

## 4 Diskussion

### 4.1 Allgemeines

Qualitätsmanagement und dessen Weiterentwicklung zählen mittlerweile zu zentralen Themen in der Medizin. Es stellt ein Führungsinstrument innerhalb der Klinikleitung dar und betrifft alle Mitarbeiter und Mitarbeiterinnen gleich. Ein Erfolg und somit ein Halten oder eine Verbesserung der Qualität kann nur erzielt werden, wenn alle Berufsgruppen ihren Anteil leisten. Änderungsvorschläge, Richtlinien und Weisungen müssen von allen Berufsgruppen akzeptiert und auch durchgeführt werden, sodass eine Verbesserung erzielt werden kann. Qualitätsmanagement funktioniert aber nicht nur wie eine Direktive in einem hierarchischen System. Das Einhalten von Weisungen und die Mitarbeit in der Umsetzung von Richtlinien sind

zwar erforderlich, umgekehrt können aber auch Vorschläge und Gedanken zur Verbesserung von allen Mitarbeitern und Mitarbeiterinnen eingebracht werden.

In vielen Institutionen und Bereichen werden jedoch Systeme zum Qualitätsmanagement nicht von Personen innerhalb der Einrichtung ausgearbeitet, sondern kommen von außerhalb. Dabei gilt es bestimmte Qualitätsziele in einem definierten Zeitraum zu erfüllen. Die Erreichung jener Ziele wird dann anhand von festgelegten Kriterien gemessen. Eine klare Definition dessen, was erreicht werden soll, ist dabei notwendig und allgemeine Angaben sollen vermieden werden. Erfolg oder Misserfolg müssen durch messbare Kriterien klar ersichtlich sein. Wenn der vorher definierte Zeitraum endet, wird bewertet ob die gesetzten Ziele erreicht werden konnten oder nicht. Danach muss im Rahmen des Qualitätsmanagements der Erfolg oder Misserfolg objektiv analysiert werden um daraus Schlüsse ziehen zu können. Eine Analyse des Versagens kann genauso wie eine Analyse des Erfolges zu Erkenntnissen führen.

## **4.2 Scoring**

Zum Scoring werden weltweit verschiedene Risikomodelle verwendet. Sie sagen alle die Ergebnisse von Behandlungen voraus und dienen so einerseits der Operationsvorbereitung und Risikoabschätzung vor Eingriffen und andererseits dem Qualitätsmanagement nach stattgefundenen Operationen. Die Qualität wird dabei unter Verwendung der tatsächlichen Mortalität, der geschätzten Mortalität und der AUC von ROC-Kurven bestimmt. Ein Vorteil solcher Modelle ist dann gegeben, wenn sie leicht zu verwenden sind, einfache eindeutige Definitionen beinhalten und mit wenigen Risikofaktoren auskommen. Der wichtigste Punkt ist allerdings, dass sie der Realität möglichst nahe kommen, also die vorausgesagten Ergebnisse mit den tatsächlichen übereinstimmen.

Die Erfassung von Daten ist jedoch genauso wichtig wie die Systeme an sich, weil die Scores nur durch eine korrekte Sammlung von Daten berechnet werden können. Die notwendigen Daten müssen vor Operationen und für den Beobachtungszeitraum

danach erfasst werden. Auf die Vollständigkeit aller relevanten Daten muss ein besonderes Augenmerk gelegt werden, weil ansonsten eine Auswertung nur begrenzt sinnvoll oder sinnlos ist. In Graz ist dies durch das QS II, ein Qualitätssicherungssystem, einfach und kann mit minimalem Aufwand durchgeführt werden. Durch standardisierte Eingaben von beteiligten Mitarbeitern und Mitarbeiterinnen können aussagekräftige Statistiken erstellt werden. Außerdem werden durch eine Standardisierung weder zu wenige noch zu viele Daten erfasst. Die interne Vergleichbarkeit, die Vergleichbarkeit mit anderen inländischen Zentren und die Vergleichbarkeit mit ausländischen Zentren, die andere Systeme verwenden, ist ein zusätzliches Asset.

Durch die Abschätzung der Mortalität ermöglichen Risikomodelle eine bessere Vorbereitung auf Operationen, bei der alle beteiligten Personen integriert werden können. Der mündige Patient oder die mündige Patientin hat somit auch mehr Einblick oder eine bessere Vorstellung, welches Risiko eine Operation beinhaltet. Scores bieten somit ein einfach verständliches objektives Kriterium, welches bei der Entscheidung ob eine Operation stattfinden soll oder nicht sehr hilfreich sein kann.

Im Rahmen der Qualitätssicherung werden Scores verwendet um die Ergebnisqualität zu vergleichen. Ein gut kalibrierter Score ist somit für eine faire Bewertung essentiell. Bei einer schlechten Kalibrierung kann die Leistung missinterpretiert und schlechter oder besser, als sie eigentlich ist, geschätzt werden.

Die zwei am weitesten verbreiteten Modelle zur Einschätzung des Risikos sind der STS-Score( Risikoscore der Society of Thoracic Surgeons) und der EuroSCORE, es gibt jedoch auch andere Modelle wie den ACEF-Score( Age, Kreatinin, Ejection Fraction).

Der STS wird vor allem in Nordamerika verwendet und hat sich dort als guter Prädiktor für die 30-Tages-Mortalität bewiesen. Er wurde im Verlauf der Jahre mehrmals überarbeitet und die aktuelle Version wurde 2007, aus Daten die im

Zeitraum von 2002 bis 2006 gesammelt wurden, entwickelt. Im Gegensatz zum EuroSCORE trifft der STS Aussagen über multiple Outcomes und nicht nur über die 30-Tages-Mortalität. Dies beinhaltet unter anderem das Risiko einen Schlaganfall oder ein Nierenversagen zu erleiden und gibt auch Auskunft über die Dauer des Krankenhausaufenthaltes. Er bietet somit ein größeres Vorhersagepaket hat jedoch auch Limitationen. Der Score ist nur für spezifische Herzoperationen einsetzbar und nicht allgemein verwendbar. Er ist verwendbar für isolierte Bypassoperationen der Koronararterien, isolierten Ersatz der Aorten- oder Mitralklappe, isolierte Mitralklappenrekonstruktionen und Kombinationen aus Klappenoperationen mit Bypassoperationen. [31]

Der EuroSCORE wurde 1999 entwickelt und ermöglichte eine sehr effiziente Beurteilung des Operationsrisikos. Die Mortalität konnte erwiesenermaßen an unterschiedlichen Orten und bei verschiedenen Populationen ausgezeichnet eingeschätzt werden. Außerdem konnte das Mortalitätsrisiko bei neuen Operationstechniken, die zum Zeitpunkt des Erscheinens des EuroSCORE noch nicht existierten, ebenfalls gut eingeschätzt werden. Mit dem medizinischen Fortschritt und einer Veränderung des Patienten- und Patientinnenguts wurden jedoch im Laufe der Zeit die anfangs guten Ergebnisse schlechter. Mängel traten auf, das Sterberisiko wurde überschätzt und besonders im Bereich der Hochrisikopatienten und Hochrisikopatientinnen, der tendenziell größer wurde, wichen die Schätzungen von der Realität ab. In vielen Studien überschätzte der logistische EuroSCORE das durchschnittliche Mortalitätsrisiko, wies aber in ROC-Analysen eine gute AUC auf. Somit zeigte der logistische EuroSCORE eine schlechte Kalibrierung mit guter Diskriminationskraft. [32–38]

Aus diesen Erfordernissen heraus wurde 2011 der EuroSCORE II entwickelt, obwohl der ursprüngliche EuroSCORE trotz Mängeln immer noch eine gute Beurteilung des Operationsrisikos ermöglichte. Risikofaktoren wurden entfernt, hinzugefügt, verändert und mit neuen Koeffizienten versehen, wie unter anderem die Nierenfunktion, die statt mit dem Serumkreatinin über die Kreatininclearance besser beurteilt werden

kann. Diese Änderungen waren notwendig um eine Anpassung an hinzugewonnene Erkenntnisse und den Technologiefortschritte zu ermöglichen. [12, 38]

Eine der wichtigsten rezenten Arbeiten zur Vorhersagekraft von Risikomodellen bei herzchirurgischen Eingriffen erschien 2016 im American Journal of Cardiology. In dieser Metaanalyse wurden drei Prädiktionsmodelle von Sullivan et al verglichen, der STS-Score, der EuroSCORE II und der ACEF-Score. Alle drei Modelle werden in Guidelines zur Risikobeurteilung von Herzoperationen empfohlen, andere Modelle wurden von der Metaanalyse ausgeschlossen. Ein Vergleich der Diskriminationskraft wurde mit der AUC durchgeführt. Dabei stellte sich heraus, dass der EuroSCORE II und der STS-Score eine annähernd gleiche Diskriminationskraft zeigen, die tendenziell höher als die des ACEF-Score ist. Beide Modelle eignen sich somit gut zur Risikoabschätzung. Bei der Kalibrierung der drei Scores war kein statistischer Unterschied ersichtlich. Ebenso konnte kein Beweis gefunden werden, dass andere Modelle, welche nicht in die Analyse inkludiert wurden, besser als die drei Inkludierten sind. Der Grund weswegen der ACEF-Score eine schlechtere Diskriminationskraft als die anderen beiden zeigte, könnte darauf zurückzuführen sein, dass er nur drei Risikofaktoren nutzt und für elektive Operationen entwickelt wurde. Außerdem basiert der ACEF-Score auf einer Kohorte aus einem Zentrum und nicht wie der EuroSCORE II auf einer multizentrischen Studie oder der STS-Score auf der STS Datenbank. [39]

Kirmani et al veröffentlichten 2013 im European Journal of Cardio-Thoracic Surgery eine Studie zum Vergleich des EuroSCORE II und des STS-Score. Dabei wurden die Daten von 15499 Eingriffen verwendet und ein Vergleich der Modelle anhand der Mortalität und der AUC erstellt. Insgesamt gab es eine beobachtete Mortalität von 3,5%. Der durchschnittliche EuroSCORE II war 2,53 und der durchschnittliche STS-Score 2,39. Die AUC betrug für den EuroSCORE II 0,818 und für den STS-Score 0,805. Beide Scores zeigten eine hohe Diskriminationskraft jedoch war die Kalibrierung beider Scores schlecht. Der EuroSCORE II und der STS-Score konnten als gute Möglichkeiten zur Risikoabschätzung validiert werden, jedoch scheinen beide das Risiko im Hochrisikobereich zu unterschätzen. [40]

In Interactive Cardiovascular and Throacic Surgery publizierten Poullis et al 2015 eine Arbeit zur Validität des logistischen EuroSCORE und EuroSCORE II bei Patienten und Patientinnen über 70 Jahre. Eine AUC von  $\geq 0,7$  wurde, unabhängig von der Sensitivität, Spezifität und Kalibrierung, als Kriterium gesetzt um einen Score als klinisch akzeptabel zu erachten. Es wurden 4584 Fälle analysiert und bewertet. Sowohl für den EuroSCORE als auch den EuroSCORE II war die AUC bei Patienten und Patientinnen über 70 Jahre unter 0,7. Bei einem Alter unter 70 Jahren war die AUC über 0,7. Beide Modelle zeigten bei Patienten und Patientinnen unter 70 Jahren eine gute Diskriminationskraft mit hoher Sensitivität und Spezifität jedoch mit Schwächen bei älteren Patienten und Patientinnen. Die Validität des logistischen EuroSCORE und EuroSCORE II konnte bei Patienten und Patientinnen unter 70 Jahren bestätigt werden. [41]

Rabbani et al verglichen 2014 den EuroSCORE, EuroSCORE II und den STS-Score bei Klappenoperationen mit und ohne Bypass in Pakistan. Bei 576 Patienten und Patientinnen mit zumindest einer Klappenoperation wurde eine Mortalität von 4,86% erhoben. Der logistische EuroSCORE überschätzte die Mortalität mit 6,96%, der EuroSCORE II ebenfalls aber sehr gering mit 4,94% und der STS-Score unterschätzte die Mortalität mit 2,13%. Alle Modelle hatten jedoch eine sehr ähnliche gute Diskriminationskraft von über 0,8. Im Vergleich zum logistischen EuroSCORE zeigte der EuroSCORE II eine bessere Kalibrierung. In einer gesonderten Auswertung von einerseits Klappenoperationen und andererseits Klappenoperationen mit Bypass ging der EuroSCORE II bei Klappenoperationen als bestes Modell hervor, wohingegen der STS-Score bei Klappenoperationen mit Bypass besser als andere Modelle abschnitt. [42]

Einen Vergleich der klinischen Leistung des EuroSCORE II, des logistischen EuroSCORE und des additiven EuroSCORE publizierten Velicki et al 2014 in The Thoracic and Cardiovascular Surgeon. Dabei wurden 1247 Patienten und Patientinnen bei welchen eine Herzoperation erfolgte inkludiert. Die Mortalitätsraten, die AUC und die performance ratio wurden jeweils verglichen. Bei einer Mortalität von 3,45% im Krankenhaus ergab sich eine durchschnittliche Mortalität von 4,43%

für den additiven EuroSCORE, 5,27% für den logistischen EuroSCORE und 2,12% für den EuroSCORE II. Die performance ratio war für den additiven EuroSCORE 0,78, für den logistischen EuroSCORE 0,65 und für den EuroSCORE II 1,63. Alle AUC-Werte waren ähnlich mit 0,754 für den additiven EuroSCORE, 0,759 für den logistischen EuroSCORE und 0,743 für den EuroSCORE II. Der logistische EuroSCORE überschätzte die Mortalität signifikant und der EuroSCORE II unterschätze die Mortalität signifikant. Obwohl beide Modelle keine idealen Vorhersagen erstellt haben, konnten sie die 30-Tages-Mortalität mit einem zufriedenstellenden Ergebnis vorhersagen. Beide Scores konnten somit validiert werden, jedoch konnte kein Vorteil eines Scores über den anderen gezeigt werden. [43]

### **4.3 ROC-Analysen und performance ratio**

Für die hier durchgeführte retrospektive Studie ergab sich eine 30-Tages-Mortalität von 2,1% mit einem durchschnittlichen logistischen EuroSCORE von 9,07 und einem EuroSCORE II von 3,64. Bei einem Vergleich der tatsächlichen Mortalität mit der erwarteten Mortalität ist ersichtlich, dass der EuroSCORE II der Realität näher kommt. Die performance ratio beträgt für den logistischen EuroSCORE 0,23 und für den EuroSCORE II 0,58. Dies kann für eine schlechte Kalibrierung der Scores, besonders für den logistischen EuroSCORE oder für eine hervorragende Leistung des Krankenhauses und der Chirurgen und Chirurginnen sprechen. Eine Kombination aus beiden Faktoren ist ebenfalls möglich.

Die AUC der ROC-Kurven wurden ebenfalls analysiert und alle Scores zeigen sehr gute Werte. Der logistische EuroSCORE hat eine AUC von 0,847 bei einem 95% Konfidenzintervall mit Untergrenze von 0,761 und Obergrenze von 0,933 und der EuroSCORE II hat eine AUC von 0,829 bei einem 95% Konfidenzintervall mit Untergrenze von 0,733 und Obergrenze von 0,926. Beide Scores zeigen eine statistische Signifikanz. In anderen Studien sind AUC Werte über 0,7 klinisch akzeptabel, was für beide Scores hier erfüllt wird. Sowohl der logistische

EuroSCORE als auch der EuroSCORE II zeigen eine außerordentlich gute Diskriminationskraft.

Um zu sehen welcher Score besser klassifizieren kann, wurden die idealen Grenzwerte bestimmt. Der Youden-Index für den logistischen EuroSCORE ist 0,61 und für den EuroSCORE II 0,554. Der geringste Abstand zur linken oberen Ecke, dem Punkt (0,1), ist für den logistischen EuroSCORE 0,284 und für den EuroSCORE II 0,341. Somit hat der logistische EuroSCORE einen höheren Youden-Index und einen geringeren Abstand zur linken oberen Ecke als der EuroSCORE II.

Im Jahr 2015 veröffentlichten Wang et al eine retrospektive Studie in The Journal of Thoracic and Cardiovascular Surgery zum Vergleich vier verschiedener Risikomodelle bei isoliertem Aortenklappenersatz. Dabei wurden der EuroSCORE, EuroSCORE II, STS-Score und der AUS-AVR-Score (aortic valve replacement nach australischem Modell) bezüglich ihrer Kalibrierung und Diskriminationskraft verglichen. Bei 620 Patienten und Patientinnen, die einen Aortenklappenersatz erhielten, betrug die Mortalität 2,9%. Der durchschnittliche EuroSCORE war 8,7%, der EuroSCORE II 3,8%, der STS-Score 2,8% und der AUS-AVR-Score 3,2%. Das ergibt eine performance ratio von 0,33 für den EuroSCORE, 0,77 für den EuroSCORE II, 1,05 für den STS-Score und 0,90 für den AUS-AVR-Score. Die AUC der ROC-Kurven war für den EuroSCORE 0,752, für den EuroSCORE II 0,711, für den STS-Score 0,716 und für den AUS-AVR-Score 0,684. Die Diskriminationskraft aller vier untersuchten Scores wurde somit als gleich gut bestätigt. Mit Ausnahme des EuroSCORE zeigten alle Scores eine gute Kalibrierung, der STS-Score schnitt jedoch am besten ab. Für den EuroSCORE war die Kalibrierung schlecht, die Mortalität wurde stark überschätzt. [37]

Barili et al publizierten in The Annals of Thoracic Surgery 2013 eine Studie zum Vergleich des STS-Score, des logistischen EuroSCORE, des EuroSCORE II und des ACEF-Score bei isoliertem Aortenklappenersatz. Die Kalibrierung und Diskriminationskraft der Scores wurde aus den Daten von 1758 Patienten und Patientinnen bestimmt. Die beobachtete Mortalität im Krankenhaus war 1,4%. Die

erwartete Mortalität betrug für den STS-Score 2,0% für den logistischen EuroSCORE 7,25%, für den EuroSCORE II 1,88% und für den ACEF-Score 0,46%. Die AUC der entsprechenden ROC-Kurven war für den STS-Score 0,85, für den logistischen EuroSCORE 0,80, für den EuroSCORE II 0,81 und für den ACEF-Score 0,75. Die Diskriminationskraft aller Scores war somit sehr gut, wobei der STS-Score am besten abschnitt. Die Kalibrierung des EuroSCORE II war jedoch am besten, mit einer Unterschätzung des Risikos im Hochrisikobereich. Der logistische EuroSCORE überschätzte das Risiko und wies eine schlechte Kalibrierung auf. [44]

Im Journal of Cardiothoracic and Vascular Anesthesia veröffentlichten Biancari et al 2014 eine Metaanalyse zur Leistung des EuroSCORE II und des STS-Score bei Aortenklappenersatz. Es wurden nur Studien inkludiert die Daten über den EuroSCORE II und den STS-Score bei Patienten und Patientinnen mit chirurgischem Aortenklappenersatz oder Transkatheter-Aortenklappenersatz beinhalten. Beide Scores zeigten eine gute Kalibrierung mit einer performance ratio nahe 1,0 und konnten validiert werden. Bei Patienten und Patientinnen mit Transkatheter-Aortenklappenersatz unterschätzten jedoch beide Scores das Risiko signifikant. Für Patienten und Patientinnen mit niedrigem Risiko, die einen chirurgischen Aortenklappenersatz erhielten, konnte der EuroSCORE II sehr präzise die Mortalität voraussagen. [45]

#### **4.4 Zusammenfassung**

Möglichkeiten zur Risikoeinschätzung sind auf chirurgischen Abteilungen nicht mehr wegzudenken. Einerseits zur Vorbereitung auf Operationen und andererseits zum Qualitätsmanagement. Patienten und Patientinnen können besser über ihr Risiko aufgeklärt und über Eingriffe informiert werden, je besser das System oder der Score, der zur Verfügung steht, ist. Ein guter Score ist ebenfalls für jegliche Statistik essentiell, weil es ansonsten zu einer starken Abweichung von der Realität kommen kann.

Die Überarbeitung des EuroSCORE nach langer Zeit war sinnvoll um eine Anpassung an die veränderte Bedingung zu gewährleisten. Andere Komorbiditäten, wie ein höheres Patienten- und Patientinnenalter, neue Erkenntnisse, Methoden und Techniken wie die Kreatininclearance zur besseren Bestimmung der Nierenfunktion und Transkatheter Operationsmethoden, veränderten die medizinischen Umstände unter denen Operationen stattfinden. Das Umfeld in dem der ursprüngliche EuroSCORE geschaffen wurde hat sich verändert.

Obwohl schon viele Jahre seit der Erstellung des logistischen EuroSCORE vergangen sind, ist er dem EuroSCORE II in Bezug auf die Diskriminationskraft in dieser Studie trotzdem nicht unterlegen. Die vorausgesagte Mortalität des EuroSCORE II entspricht aber eher der Realität, womit er einer besseren Kalibrierung aufweist, wobei die niedrige performance ratio im Vergleich zu anderen Studien auf eine gute Leistung zurückzuführen sein kann. Die extrem niedrige performance ratio des logistischen EuroSCORE wird auf einen Mix aus guter Leistung des Krankenhauses und schlechter Kalibrierung mit Überschätzung des Risikos zurückzuführen sein.

Der perfekte Score zur Risikoeinschätzung existiert leider nicht. Dazu müsste ein Score alle Risikofaktoren beinhalten und in allen Situationen anwendbar sein. Es wird allerdings immer Bereiche geben in denen ein Score schlechter als ein anderer abschneidet und das Berücksichtigen aller Risikofaktoren ist nicht plausibel. Außerdem verändern sich die Bedingungen unter denen Scores entwickelt wurden laufend, weswegen immer Anpassungen erfolgen müssen. Der EuroSCORE II ist jedoch klar eine Verbesserung des ursprünglichen EuroSCORE und kommt somit dem idealen Score einen Schritt näher.

## 5 Literatur

1. Markgraf, D. Qualität. <http://wirtschaftslexikon.gabler.de/Archiv/55799/qualitaet-v6.html> (letzter Zugriff am: 31.07.2017)
2. Qualitätsmanagement. <http://www.wirtschaftslexikon24.com/d/qualitaetsmanagement/qualitaetsmanagement.htm> (letzter Zugriff am: 31.07.2017)
3. Hellmann W. Qualitätsmanagement für Chirurgen – Kenntnis grundlegender Zusammenhänge und innovativer Strategien fördert die Wettbewerbsfähigkeit der Fachabteilung. Zentralblatt für Chirurgie 2017
4. Australasian Medical Publishing Company, Duckett SJ, Coory M, Sketcher-Baker K. The Medical Journal of Australia // Identifying variations in quality of care in Queensland hospitals. The Medical journal of Australia 2007; 187: 571–575
5. What is euroSCORE ? [http://www.euroscore.org/what\\_is\\_euroscore.htm](http://www.euroscore.org/what_is_euroscore.htm) (letzter Zugriff am: 31.07.2017)
6. Garcia-Valentin A, Mestres CA, Bernabeu E, et al. Validation and quality measurements for EuroSCORE and EuroSCORE II in the Spanish cardiac surgical population. European journal of cardio-thoracic surgery : official journal of the European Association for Cardio-thoracic Surgery 2016; 49: 399–405
7. Guida P, Mastro F, Scrascia G, Whitlock R, Paparella D. Performance of the European System for Cardiac Operative Risk Evaluation II. The Journal of thoracic and cardiovascular surgery 2014; 148: 3049-57.e1
8. Roques F, Nashef SA, Michel P, et al. Risk factors and outcome in European cardiac surgery. European journal of cardio-thoracic surgery : official journal of the European Association for Cardio-thoracic Surgery 1999; 15: 816-22; discussion 822-3
9. Roques F, Michel P, Goldstone AR, Nashef SAM. The logistic EuroSCORE. European Heart Journal 2003; 24: 881–882
10. euroSCORE scoring system. [http://euroscore.org/euroscore\\_scoring.htm](http://euroscore.org/euroscore_scoring.htm) (letzter Zugriff am: 03.08.2017)
11. How to calculate the logistic euroSCORE. <http://euroscore.org/logisticEuroSCORE.htm> (letzter Zugriff am: 03.08.2017)

12. Nashef SAM, Roques F, Sharples LD, et al. EuroSCORE II. European journal of cardio-thoracic surgery : official journal of the European Association for Cardio-thoracic Surgery 2012; 41: 734-44; discussion 744-5
13. Behrends JC. [Thie.] Duale Reihe Physiologie (2010) // Physiologie. Stuttgart: Thieme, 2010
14. Gerd Herold, Herold G. Innere Medizin // Innere Medizin 2014. Köln: Selbstverl., 2014
15. Hahn J-M, Adler G. Checkliste innere Medizin. Stuttgart: Thieme, 2010
16. Vahanian A, Alfieri O, Andreotti F, et al. Guidelines on the management of valvular heart disease (version 2012). European Heart Journal 2012; 33: 2451–2496
17. Baumgartner H, Falk V, Bax JJ, et al. 2017 ESC/EACTS Guidelines for the management of valvular heart disease. European Heart Journal 2017
18. Siewert JR, Stein HJ. Chirurgie. Berlin, Heidelberg: Springer, 2012
19. Autschbach R, Jacobs M, Neumann UP. Chirurgie in 5 Tagen. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012
20. Henne-Bruns D, Barth E. Chirurgie. Stuttgart: Thieme, 2012
21. Nishimura RA, Otto CM, Bonow RO, et al. 2014 AHA/ACC Guideline for the Management of Patients With Valvular Heart Disease. Circulation 1997 // 2014; 129: e521-643
22. Metz CE. Basic principles of ROC analysis. Seminars in Nuclear Medicine 1978; 8: 283–298
23. Swets JA. Swets\_1988 // Measuring the accuracy of diagnostic systems. Science (New York, N.Y.) 1988; 240: 1285–1293
24. Metz CE. Receiver operating characteristic analysis. Journal of the American College of Radiology : JACR 2006; 3: 413–422
25. Kumar R, Indrayan A. Receiver operating characteristic (ROC) curve for medical researchers. Indian Pediatr 2011; 48: 277–287
26. Streiner DL, Cairney J. What's under the ROC? An introduction to receiver operating characteristics curves. Canadian journal of psychiatry. Revue canadienne de psychiatrie 2007; 52: 121–128
27. YODEN WJ. Index for rating diagnostic tests. Cancer 1950; 3: 32–35
28. O'Neill S, Wigmore SJ, Harrison EM. Debate. BMC surgery 2015; 15: 102

29. Coory M, Duckett S, Sketcher-Baker K. Using control charts to monitor quality of hospital care with administrative data. *International journal for quality in health care : journal of the International Society for Quality in Health Care* 2008; 20: 31–39
30. Williams H, Gwyn R, Smith A, Dramis A, Lewis J. Variable life-adjusted display (VLAD) for hip fracture patients. *European journal of orthopaedic surgery & traumatology : orthopedie traumatologie* 2015; 25: 1047–1050
31. About the STS Risk Calculator v.81.  
<http://riskcalc.sts.org/stswebriskcalc/views/About%20the%20STS%20Risk%20Calculator%20v2%2081.pdf> (letzter Zugriff am: 30.11.2017)
32. Nashef SAM, Roques F, Hammill BG, et al. Validation of European System for Cardiac Operative Risk Evaluation (EuroSCORE) in North American cardiac surgery. *European journal of cardio-thoracic surgery : official journal of the European Association for Cardio-thoracic Surgery* 2002; 22: 101–105
33. Basraon J, Chandrashekar YS, John R, et al. Comparison of risk scores to estimate perioperative mortality in aortic valve replacement surgery. *The Annals of thoracic surgery* 2011; 92: 535–540
34. Grossi EA, Schwartz CF, Yu P-J, et al. High-risk aortic valve replacement. *The Annals of thoracic surgery* 2008; 85: 102-6; discussion 107
35. Leontyev S, Walther T, Borger MA, et al. Aortic valve replacement in octogenarians. *The Annals of thoracic surgery* 2009; 87: 1440–1445
36. Wendt D, Osswald BR, Kayser K, et al. Society of Thoracic Surgeons score is superior to the EuroSCORE determining mortality in high risk patients undergoing isolated aortic valve replacement. *The Annals of thoracic surgery* 2009; 88: 468-74; discussion 474-5
37. Wang TKM, Choi DHM, Stewart R, Gamble G, Haydock D, Ruygrok P. Comparison of four contemporary risk models at predicting mortality after aortic valve replacement. *The Journal of thoracic and cardiovascular surgery* 2015; 149: 443–448
38. Lebreton G, Merle S, Inamo J, et al. Limitations in the inter-observer reliability of EuroSCORE. *European journal of cardio-thoracic surgery : official journal of the European Association for Cardio-thoracic Surgery* 2011; 40: 1304–1308

39. Sullivan PG, Wallach JD, Ioannidis JPA. Meta-Analysis Comparing Established Risk Prediction Models (EuroSCORE II, STS Score, and ACEF Score) for Perioperative Mortality During Cardiac Surgery. *The American journal of cardiology* 2016; 118: 1574–1582
40. Kirmani BH, Mazhar K, Fabri BM, Pullan DM. Comparison of the EuroSCORE II and Society of Thoracic Surgeons 2008 risk tools. *European journal of cardiothoracic surgery : official journal of the European Association for Cardio-thoracic Surgery* 2013; 44: 999-1005; discussion 1005
41. Poullis M, Pullan M, Chalmers J, Mediratta N. The validity of the original EuroSCORE and EuroSCORE II in patients over the age of seventy. *Interactive cardiovascular and thoracic surgery* 2015; 20: 172–177
42. Rabbani MS, Qadir I, Ahmed Y, Gul M, Sharif H. Heart valve surgery. *Heart international* 2014; 9: 53–58
43. Velicki L, Cemerlic-Adjic N, Pavlovic K, et al. Clinical performance of the EuroSCORE II compared with the previous EuroSCORE iterations. *The Thoracic and cardiovascular surgeon* 2014; 62: 288–297
44. Barili F, Pacini D, Capo A, et al. Reliability of new scores in predicting perioperative mortality after isolated aortic valve surgery. *The Annals of thoracic surgery* 2013; 95: 1539–1544
45. Biancari F, Juvonen T, Onorati F, et al. Meta-analysis on the performance of the EuroSCORE II and the Society of Thoracic Surgeons Scores in patients undergoing aortic valve replacement. *Journal of cardiothoracic and vascular anesthesia* 2014; 28: 1533–1539

# Appendix 1 –Ethikvotum

## Ethikvotum 1

Ethikkommission



Medizinische Universität Graz

Auenbruggerplatz 2, A-8036 Graz  
ethikkommission@medunigraz.at  
Tel.: +43 / 316 / 385-13928, Fax: -14348

### VOTUM

gültig bis 08.08.2018

**EK-Nummer:** 29-516 ex 16/17  
**Studientitel:** Validation of EuroSCORE II risk model for aortic valve replacement - a retrospective study  
**Prüfer:** Ao. Univ.-Prof. Dr. med. univ. Igor Knez  
Klinische Abteilung für Herzchirurgie Universitätsklinik Chirurgie Medizinische U  
**Sponsor:** Klinische Abteilung für Herzchirurgie Universitätsklinik Chirurgie Medizinische  
Universität Graz  
**Ansprechpartner:** Ao. Univ.-Prof. Dr. med. univ. Igor Knez, 8036 Graz, Auenbruggerplatz 29  
**CRO:** -  
**Antragsteller:** Medizinische Universität Graz  
**Ansprechpartner:** cand. med. Yannik Snieder

Die o.a. Studie wurde von der Ethikkommission erstmals im 'expedited Review' am 29.06.2017 behandelt.  
Die Ethikkommission ist zu folgendem Schluss gekommen:

**Es besteht kein Einwand gegen die Durchführung der Studie in der vorliegenden Form.**

Kommissionsmitglieder, die für diesen Tagesordnungspunkt als befangen anzusehen waren und daher gemäß Geschäftsordnung an der Entscheidungsfindung und Abstimmung nicht teilgenommen haben:  
keine

#### Zur Beurteilung vorliegende Dokumente:

<b>Dokumente eingegangen am 16.06.2017, begutachtet im 'expedited Review' am 29.06.2017</b>	
✓ Antragsformular ECS	16.06.2017
Originalprotokoll YS_Studienprotokoll 1.1	16.06.2017
<b>Dokumente eingegangen am 30.06.2017 (in der nächsten Begutachtung mitbegutachtet)</b>	
✓ Cover Letter	30.06.2017
✓ Originalprotokoll 1.2	30.06.2017
<b>Dokumente eingegangen am 22.07.2017 (in der nächsten Begutachtung mitbegutachtet)</b>	
✓ Letter of Authorization	22.07.2017
<b>Dokumente eingegangen am 07.08.2017, begutachtet im 'expedited Review' am 08.08.2017</b>	
✓ Antragsformular ECS Unterschriftenseiten	07.08.2017

Die Ethikkommission geht - rechtlich unverbindlich - davon aus, dass es sich um keine klinische Prüfung nach AMG bzw. MPG handelt.

Es handelt sich um eine Studie im Rahmen einer Diplomarbeit.

Das Votum der Ethikkommission berührt in keiner Weise die alleinige Verantwortung der Prüferin / des Prüfers / der Prüfer für die ordnungsgemäße Durchführung der Studie unter Einhaltung aller einschlägiger gesetzlicher Bestimmungen und Richtlinien.

Weiters machen wir darauf aufmerksam, dass der Kommission unverzüglich zu melden sind:

- Abweichungen vom Protokoll aus Sicherheitsgründen oder Protokolländerungen

EK-Nummer: 29-516 ex 16/17

Votum (08.08.2017)

Seite 1 von 2

Medizinische Universität Graz, Auenbruggerplatz 2, A-8036 Graz. [www.medunigraz.at](http://www.medunigraz.at)

Rechtsform: Juristische Person öffentlichen Rechts gem. Universitätsgesetz 2002. Information: Mitteilungsblatt der Universität und [www.medunigraz.at](http://www.medunigraz.at). DVR-Nr. 210 9494.  
UID: ATU 575 111 79. Bankverbindung: Bank Austria Creditanstalt BLZ 12000 Konto-Nr. 500 948 400 04, Raiffeisen Landesbank Steiermark BLZ 38000 Konto-Nr. 49510.

# Appendix 1 –Ethikvotum

## Ethikvotum 2

- Änderungen, die das Risiko der Teilnehmer/-innen erhöhen oder die Durchführung der Studie wesentlich beeinflussen
- Mutmaßliche unerwartete schwerwiegende Nebenwirkungen - SUSARs (AMG-Studien ab 1.5.2004) oder schwerwiegende unerwünschte Ereignisse - SAEs (andere Studien)
- Jegliche Information über sonstige Umstände, die die Sicherheit der Teilnehmer/-innen oder die Durchführung der Studie beeinträchtigen können

Dieses Votum gilt für ein Jahr ab dem Datum der Ausstellung. Bei längerer Studiendauer ist rechtzeitig vor Ablauf der Gültigkeit des Votums ein Zwischenbericht vorzulegen (Berichtsformular), um eine etwaige Verlängerung zu erlangen.

Graz, 08. August 2017

  
\_\_\_\_\_  
Univ. Prof. DI Dr. Josef Haas  
Vorsitzender

  
\_\_\_\_\_  
i. V. Univ. Prof. Dr. Hermann Toplak  
Stv. Vorsitzender

**Achtung:** Bitte bei allen das Projekt betreffende Schreiben oder telefonischen Anfragen die EK-Nummer angeben!

## Appendix 2 –performance ratio nach Jahren

### Performance ratio des logistischen EuroSCORE 2012

	Gesamt	männlich	weiblich
OP-Anzahl	164	87	77
Verstorbene	3	1	2
Verstorbene (%)	1,80	1,14	2,53
durchschnittlicher Score (%)	8,04	5,91	10,41
performance ratio	0,22	0,19	0,24

### Performance ratio des EuroSCORE II 2012

	Gesamt	männlich	weiblich
OP-Anzahl	164	87	77
Verstorbene	3	1	2
Verstorbene (%)	1,80	1,14	2,53
durchschnittlicher Score (%)	3,00	2,09	4,02
performance ratio	0,60	0,55	0,63

### Performance ratio des logistischen EuroSCORE 2013

	Gesamt	männlich	weiblich
OP-Anzahl	166	103	63
Verstorbene	2	1	1
Verstorbene (%)	1,19	0,96	1,56
durchschnittlicher Score (%)	8,64	7,33	10,76
performance ratio	0,14	0,13	0,14

### Performance ratio des EuroSCORE II 2013

	Gesamt	männlich	weiblich
OP-Anzahl	166	103	63
Verstorbene	2	1	1
Verstorbene (%)	1,19	0,96	1,56
durchschnittlicher Score (%)	3,23	2,65	4,18
performance ratio	0,37	0,36	0,37

## Appendix 2 –performance ratio nach Jahren

### Performance ratio des logistischen EuroSCORE 2014

	Gesamt	männlich	weiblich
OP-Anzahl	164	99	65
Verstorbene	8	6	2
Verstorbene (%)	4,65	5,71	2,99
durchschnittlicher Score (%)	10,20	10,52	9,70
performance ratio	0,46	0,54	0,31

### Performance ratio des EuroSCORE II 2014

	Gesamt	männlich	weiblich
OP-Anzahl	164	99	65
Verstorbene	8	6	2
Verstorbene (%)	4,65	5,71	2,99
durchschnittlicher Score (%)	4,55	4,87	4,05
performance ratio	1,02	1,17	0,74

### Performance ratio des logistischen EuroSCORE 2015

	Gesamt	männlich	weiblich
OP-Anzahl	155	97	58
Verstorbene	1	0	1
Verstorbene (%)	0,64	0,00	1,69
durchschnittlicher Score (%)	8,61	6,71	11,74
performance ratio	0,07	0,00	0,14

### Performance ratio des EuroSCORE II 2015

	Gesamt	männlich	weiblich
OP-Anzahl	155	97	58
Verstorbene	1	0	1
Verstorbene (%)	0,64	0,00	1,69
durchschnittlicher Score (%)	2,90	2,44	3,66
performance ratio	0,22	0,00	0,46

## Appendix 2 –performance ratio nach Jahren

### Performance ratio des logistischen EuroSCORE 2016

	Gesamt	männlich	weiblich
OP-Anzahl	176	104	72
Verstorbene	4	1	3
Verstorbene (%)	2,22	0,95	4,00
durchschnittlicher Score (%)	9,75	10,01	9,39
performance ratio	0,23	0,09	0,43

### Performance ratio des EuroSCORE II 2016

	Gesamt	männlich	weiblich
OP-Anzahl	176	104	72
Verstorbene	4	1	3
Verstorbene (%)	2,22	0,95	4,00
durchschnittlicher Score (%)	4,38	5,21	3,22
performance ratio	0,51	0,18	1,24