

Diplomarbeit

In-Hospital-Qualitätssicherung in der Herzchirurgie

Ist der neue EuroSCORE II im Vergleich zum bereits verwendeten EuroSCORE das approbate Mittel zur Qualitätssicherung bei aortokoronaren Bypasspatienten in der Herzchirurgie?

eingereicht von:

Alissa Strugger

zur Erlangung des akademischen Grades

Doktorin 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, 21.02.2017

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 21.02.2017

Alissa Strugger eh

Danksagungen

An dieser Stelle möchte ich nachfolgend stehenden Personen meinen großen Dank aussprechen, ohne deren Hilfe und Unterstützung diese Arbeit nie zustande gekommen wäre.

Zuallererst danke ich Herrn Univ.-Prof. Dr. Igor Knez für die Bereitstellung dieses Themas und die Betreuung meiner Arbeit, für seine Anleitung und Hilfestellung. Die vielen Zusammentreffen, bei denen er jederzeit ein offenes Ohr für mich hatte und mir mit guten Ratschlägen zur Seite stand, waren für mich stets eine große Bereicherung. Außerdem bedanke ich mich bei Herrn cand. med. Jovan Vujic für die Bereitstellung seines Fachwissens und die kritische Auseinandersetzung mit meiner Arbeit.

Ferner danke ich auch Frau Dipl.-Ing. Dr. Regina Riedl für ihre Unterstützung in statistischen Belangen.

Ich möchte mich auch bei meiner besten Freundin Jennifer bedanken, der das Themengebiet der vorliegenden Arbeit nicht fern liegen könnte, die sich aber doch immer ausführlich über meinen Fortschritt informiert hat und mir mit unterstützenden Worten zur Seite stand. Ich danke auch ihrem Freund Jakob, der mir mit seinen ausgezeichneten Computerkenntnissen immer eine große Hilfe war und ohne den die Bewältigung der Statistik und der Formatierung für mich um einiges schwieriger gewesen wäre.

Des Weiteren danke ich meinem Freund Christoph für seine Ermunterungen, seine Kritik, seine Geduld und seine Unterstützung. Selbst in den stressigsten Zeiten konnte ich dank ihm zur Ruhe kommen, mich neu sammeln und mich wieder auf das Wesentliche konzentrieren.

Abschließend gilt mein besonderer Dank meiner Mutter, die mir mein Studium erst ermöglicht hat und die mir die ganzen Jahre zur Seite gestanden hat und immer an mich glaubt. Auch während der Erstellung der Arbeit hatte sie immer aufmunternde Worte für mich übrig und ohne ihre sprachliche Expertise hätte ich manche Passagen nicht so flüssig meistern können.

Zusammenfassung

Hintergrund: Der EuroSCORE („European System for Cardiac Operative Risk Evaluation“) wurde entwickelt, um das Risiko eines Patienten oder einer Patientin, innerhalb von 30 Tagen nach einer Herzoperation zu versterben anhand seiner oder ihrer individuellen Risikofaktoren zu berechnen. Aufgrund des sich verändernden Risikoprofils von Herzpatienten und -patientinnen und der neuen Techniken und Möglichkeiten in der Herzchirurgie machte sich der Bedarf nach einem überarbeiteten, moderneren System bemerkbar. Dabei handelt es sich um den EuroSCORE II. In unserer Studie wurde die prädiktive Aussagekraft des logistischen EuroSCORE mit der des EuroSCORE II verglichen.

Methodik: Anhand der Daten von 1542 koronarerteriellen Bypasspatienten und -patientinnen, die von 2012 bis 2016 an der Abteilung für Herzchirurgie des Universitätsklinikums Graz operiert wurden, wurden die Ergebnisse der beiden Scores überprüft. Receiver Operating Characteristic (ROC) Analysen wurden in Bezug auf die 30-Tages Mortalität durchgeführt, um Aussagen über Sensitivität und Spezifität zu erhalten. Variable-Life-Adjusted-Display (VLAD) und Funnel Plots (FP) wurden verwendet um die Performance Ratio darzustellen.

Resultate: Die tatsächliche 30-Tage Sterblichkeit für den gesamten Beobachtungszeitraum betrug 3,8%. Die durchschnittliche vorausgesagte 30-Tage Mortalität für den logistischen EuroSCORE und den EuroSCORE II betrug 7,33% (Performance Ratio 0,56) bzw. 4,79% (Performance Ratio 0,86). Die Diskriminationskraft für die 30-Tage Mortalität – ermittelt mittels Area Under the Curve (AUC) – für den logistischen EuroSCORE und den EuroSCORE II 0,778 (0,719-0,837 95% KI) bzw. 0,789 (0,732-0,846 95% KI). In weiteren Analysen wurden der logistische EuroSCORE und der EuroSCORE II verglichen, wobei eine hochsignifikante Korrelation ($p < 0.001$) festzustellen war. Der EuroSCORE II zeigte insgesamt eine etwas bessere absolute Voraussage- und Diskriminationskraft in Bezug auf den Faktor 30-Tage Mortalität.

Konklusion: Der EuroSCORE II zeigte eine etwas bessere prädiktive Diskrimination in Bezug auf die 30-Tage Mortalität bei der Gruppe der Bypass-Patienten. Die inklusive Natur des EuroSCORE II für verschiedenartige herzchirurgische Prozeduren zeigt eine

bessere Flexibilität als der logistische EuroSCORE. Als eine Konsequenz daraus kann der EuroSCORE II für die prädiktive Outcome-Kalkulation von komplexen herzchirurgischen Patienten uneingeschränkt verwendet werden.

Abstract

Background: The EuroSCORE („European System for Cardiac Operative Risk Evaluation“) was developed to calculate the risk of a patient to die within a period of 30 days after a cardiac operation based on his individual risk factors. The need for an overhauled modern system became noticeable because of the changing risk profile in cardiac patients and the new techniques and possibilities in cardiac surgery, Thereby we are talking about the EuroSCORE II. The predictive validity of the logistic EuroSCORE and the EuroSCORE II were compared in our study.

Methods: The results of both scores were reviewed based on the data of 1542 coronarartery bypass patients, who were operated at the department of cardiac surgery of the University Hospital Graz between 2012 and 2016. Receiver Operating Characteristic (ROC) analyses regarding mortality within 30 days were performed in order to determine sensitivity and specificity of both scores. Variable-Life-Adjusted-Display (VLAD) and Funnel Plots (FP) were used to express the performance ratio.

Results: Observed 30-day mortality for the whole observational period was 3,8%. Mean predicted mortality for logistic EuroSCORE and EuroSCORE II was 7,33% (performance ratio 0,56) and 4,79%, respectively (performance ratio 0,86). The discriminative ability for 30-day mortality, identified by means of Area Under the Curve (AUC), for logistic EuroSCORE and EuroSCORE II was 0,778 (0,719-0,837 95% CI) and 0,789 (0,732-0,846 95% CI), respectively. In secondary analyses comparing logistic EuroSCORE with EuroSCORE II, the risk scores were correlated with $p < 0.001$. However 30-day mortality for EuroSCORE II overall displayed a slightly better absolute prediction and discriminative ability.

Conclusions: EuroSCORE II displayed a slightly better predictive discrimination for 30-day mortality, in comparison to the logistic EuroSCORE, within the group of bypass patients. The inclusive nature of EuroSCORE II for numerous procedures provides more flexibility than the logistic EuroSCORE for complex high-risk procedures. As a consequence, EuroSCORE II should be considered for calculating risk scores for complex cardiac surgical patients.

Inhaltsverzeichnis

Eidesstattliche Erklärung	I
Danksagungen	II
Zusammenfassung	III
Abstract	V
Inhaltsverzeichnis	VI
Glossar und Abkürzungen	VIII
Abbildungsverzeichnis	IX
Tabellenverzeichnis.....	XI
1. Einführung	1
1.1. Qualität und Qualitätsmanagement im Allgemeinen	1
1.1.1. Definition von Qualität im medizinischen Bereich?	1
1.2. Der EuroSCORE	3
1.3. Bypasschirurgie.....	14
1.3.1. Grundlagen	14
1.3.2. Herz-Lungen-Maschine	16
1.3.3. Kardioprotektion.....	17
1.3.4. Indikation.....	17
1.3.5. Prozedere	17
1.3.6. Prognose.....	18
1.3.7. Risiken.....	19
2. Methoden.....	21
2.1. Statistische Methoden.....	21
2.1.1. ROC-Analysen	21
2.1.1.1.Grundlagen.....	21
2.1.2. VLAD und Funnel Plot	39

2.1.3. Youden-Index	43
2.1.4. Datenreview, verwendete statistische Methoden, Zulassung.....	43
3. Ergebnisse	45
3.1. Demographische Daten	45
3.2. ROC-Analyse: Logistischer EuroSCORE vs. EuroSCORE II für den Zeitraum von 2012 bis 2016	50
3.3. Funnelplot und VLAD-Analysen	58
3.3.1. Funnelplot und VLAD für den logistischen EuroSCORE (2012-2016).....	58
3.3.2. Funnelplot und VLAD für den EuroSCORE II (2012-2016).....	59
4. Diskussion	61
4.1. Allgemeines	61
4.2. Scoring.....	61
4.3. ROC-Analysen.....	65
4.4. Funnel-Plots und VLADS.....	68
4.5. Zusammenfassung	68
Literaturverzeichnis	70

Glossar und Abkürzungen

ACEF	Age, Creatinine, Ejection Fraction	max.	maximal
ACT	Activated Clot Time	MIDCAB	Minimally Invasive Direct Coronary Artery Bypass
AUC	Area Under the Curve	min.	minimal
bzw.	beziehungsweise	NYHA	New York Health Association
CABG	Coronararterieller Bypass Graft	PTCA	Perkutane Transluminale Coronarangiographie
CC	Creatinine Clearance	RCX	Ramus Circumflexus
CCS	Canadian Cardiovascular Society	RIVA	Ramus Interventricularis Anterior
FNF	Falsch Negative Fraktion	RIVP	Ramus Interventricularis Posterior
FPF	Falsch Positive Fraktion	RNF	Richtig Negative Fraktion
FROC	Free Response Receiver Operating Characteristics	ROC	Receiver Operating Characteristics
HLM	Herz-Lungen-Maschine	RPF	Richtig Positive Fraktion
i.v.	intravenös	VLAD	Variable Life Adjusted Display
IDDM	Insulin-Dependant Diabetes Mellitus	vs.	versus
IMA	Internal Mammary Artery	z.B.	zum Beispiel
LROC	Localization Receiver Operating Characteristics	ZNS	Zentrales Nervensystem
LV	Linksventrikulär		
LVEF	Left Ventricular Ejection Fraction		

Abbildungsverzeichnis

Abbildung 1: Maske im Onlinekalkulator für den logistischen EuroSCORE	12
Abbildung 2: ROC-Kurve	13
Abbildung 3: Der Effekt des Entscheidungsgrenzwertes bzw. der Abnormitätsschwelle auf Sensitivität und Spezifität. Mit einer Entscheidungs-Schwelle auf t_3 auf der Entscheidungs-Achse links, ergibt das Verteilungspaar eine Kombination aus Sensitivität (hellgrau) und Spezifität (dunkelgrau), welche rechts im Sensitivität-Spezifität Graphen als t_3 aufgetragen wird. (Gleiche Wirkung bei t_1 , t_2 und t_4)	25
Abbildung 4: Die linke Seite zeigt die kontinuierliche Beziehung zwischen Sensitivität und Spezifität, abgeleitet aus Abb. 3 bei kontinuierlicher Variation des Entscheidungsgrenzwertes. Die rechte Seite, welche eine typische ROC-Kurve repräsentiert, ist ein Spiegelbild, da die Sensitivität gegen $1 - \text{Spezifität}$ aufgetragen wird.	27
Abbildung 5: Das Problem, die diagnostische Genauigkeit zweier Systeme zu vergleichen, wenn nur die Sensitivität oder die Spezifität bekannt ist.....	27
Abbildung 6: Das Problem kann gelöst werden, wenn die ROC-Kurven beider Tests bestimmt werden.	29
Abbildung 7: Das Problem kann gelöst werden, wenn die ROC-Kurven beider Tests bestimmt werden.	29
Abbildung 8: Eine konventionelle ROC-Kurve. Die Aufgabe des Entscheiders ist es, zwischen tatsächlich positiven Fällen und tatsächlich negativen Fällen zu unterscheiden. Es wird keine Lokalisation der Läsion verlangt ⁽³¹⁾	35
Abbildung 9: Festlegung der Schwellenwerte in einer Beispielkurve	39
Abbildung 10: Funnel Plot und VLAD	41
Abbildung 11: Statistik im Funnel Plot.....	42
Abbildung 12: ROC-Kurve logistischer EuroSCORE für 2012-2016	51
Abbildung 13: ROC-Kurve EuroSCORE II für 2012-2016.....	51
Abbildung 14: logistischer EuroSCORE im Beobachtungsjahr 2012	53
Abbildung 15: EuroSCORE II im Beobachtungsjahr 2012.....	53
Abbildung 16: logistischer EuroSCORE im Beobachtungsjahr 2013	54
Abbildung 17: logistischer EuroSCORE im Beobachtungsjahr 2013	54
Abbildung 18: logistischer EuroSCORE im Beobachtungsjahr 2014	55
Abbildung 19: EuroSCORE II im Beobachtungsjahr 2014.....	55
Abbildung 20: logistischer EuroSCORE im Beobachtungsjahr 2015	56

Abbildung 21: EuroSCORE II im Beobachtungsjahr 2015	56
Abbildung 22: logistischer EuroSCORE im Beobachtungsjahr 2016	57
Abbildung 23: EuroSCORE II im Beobachtungsjahr 2016	57
Abbildung 24: Funnelpplot und VLAD für den logistischen EuroSCORE	59
Abbildung 25: Funnelpplot und VLAD für den EuroSCORE II	60
Abbildung 26: Ethikvotum 1	80
Abbildung 27: Ethikvotum 2	81
Abbildung 28: Maske im Onlinekalkulator für den EuroSCORE II	82
Abbildung 29: Funnelpplot logistischer EuroSCORE 2012	83
Abbildung 30: Funnelpplot EuroSCORE II 2012.....	83
Abbildung 31: Funnelpplot logistischer EuroSCORE 2013	84
Abbildung 32: Funnelpplot EuroSCORE II 2013.....	84
Abbildung 33: Funnelpplot logistischer EuroSCORE 2014	85
Abbildung 34: Funnelpplot EuroSCORE II 2014.....	85
Abbildung 35: Funnelpplot logistischer EuroSCORE 2015	86
Abbildung 36: Funnelpplot EuroSCORE II 2015.....	86
Abbildung 37: CABG 2012-2015 relative Mortalität in %.....	87
Abbildung 38: CABG 2012-2015 durchschnittlicher logistischer EuroSCORE in %	88
Abbildung 39: CABG 2012-2015 Performance Ratio logistischer EuroSCORE.....	88
Abbildung 40: CABG 2012-2015 durchschnittlicher EuroSCORE II in %	89
Abbildung 41: CABG 2012-2015 Performance Ratio EuroSCORE II.....	89

Tabellenverzeichnis

Tabelle 1: Die patientenbezogenen Risikofaktoren und deren Gewichtung im numerischen sowie logistischen EuroSCORE	7
Tabelle 2: Die herzbezogenen Risikofaktoren und deren Gewichtung im numerischen sowie logistischen EuroSCORE	8
Tabelle 3: Die operationsbezogenen Risikofaktoren und deren Gewichtung im numerischen sowie logistischen EuroSCORE	8
Tabelle 4: Die patientenbezogenen Risikofaktoren und deren Gewichtung im EuroSCORE II	10
Tabelle 5: Die herzbezogenen Risikofaktoren und deren Gewichtung im EuroSCORE II	11
Tabelle 6: Die operationsbezogenen Risikofaktoren und deren Gewichtung im EuroSCORE II	11
Tabelle 7: Patientengut	45
Tabelle 8: Risikofaktoren	46
Tabelle 9: Status und 30-Tages-Status	47
Tabelle 10: Alter und Scores der Patienten und Patientinnen	47
Tabelle 11: Status und 30-Tages-Status nach OP-Jahr	48
Tabelle 12: Alter und Scores nach OP-Jahr	49
Tabelle 13: Anzahl der Überlebenden und Toten nach 30 Tagen	50
Tabelle 14: Die AUC-Werte zeigen statistische Signifikanz für den Zeitraum von 2012-2016	50
Tabelle 15: 30-Tages-Status nach OP-Jahr	52
Tabelle 16: Scores und AUC mit KI nach OP-Jahr	52
Tabelle 17: Gesamtübersicht über die Ergebnisse	58
Tabelle 18: Performance Ratio des logistischen EuroSCORE	58
Tabelle 19: Performance Ratio des EuroSCORE II	59

1. Einführung

1.1. Qualität und Qualitätsmanagement im Allgemeinen

Die modernen Systeme des Gesundheitswesens sind durch den wachsenden Kosten- und Leistungsdruck und die damit verbundene Konkurrenzsituation der einzelnen Branchenteilnehmer gekennzeichnet. Sämtliche in der Medizin angewandte Verfahren müssen sich auszahlen, zuallererst natürlich zu Gunsten der verringerten Mortalität, im weiteren Verlauf aber auch im Sinne eines Krankenhausaufenthaltes, der sich so kurz wie möglich gestaltet, einer raschen Remobilisation und Rehabilitation und einer schnellen Wiedereingliederung in das Berufsleben der Patienten und Patientinnen.

Daraus ergibt sich eine Verdrängungs- und Neuordnungssituation am Markt „Gesundheit gegen Krankheit“, basierend auf Parametern, die für die gute Vergleichbarkeit verschiedener Krankenanstalten sorgen. Hieraus resultiert mehr der Bedarf an optimaler Strukturierung und qualitätsorientierter Leistungserbringung als die reine Quantitätsorientierung der einzelnen Krankenhausträger. Daher erlangt medizinisches Qualitätsmanagement, als eine Form der Klinikleitung, mit dem Ziel einer flächendeckenden Implementierung von medizinischen Qualitätssystemen zum Zwecke der transparenten Verbesserung der Patientensicherheit und zur Steigerung von Effizienz und Effektivität, nicht nur aufgrund gesetzlicher Vorgaben, wachsende Bedeutung.

1.1.1. Definition von Qualität im medizinischen Bereich?

Die an die ISO-Norm angelehnte Definition „Qualität ist die Gesamtheit von Eigenschaften und Merkmalen eines Produktes oder einer Dienstleistung, die sich auf seine/ihre Eignung zur Erfüllung festgelegter und vorausgesetzter Bedürfnisse beziehen.“ bedarf weiterer Zusätze, wenn sie im medizinischen Bereich angewandt werden soll. ⁽¹⁵⁾

Primär kann Qualität eingeteilt werden in Infrastruktur-, Prozess- und Ergebnisqualität. Hier kann die Messung mittels verschiedenster Parameter objektiv gemessen werden.

Des Weiteren kann man Qualität auch in den Aspekten der gelieferten, wahrgenommenen und genutzten Qualität beschreiben. Hierbei handelt es sich allerdings nur um subjektiv wahrgenommene Parameter, was im Folgenden kurz anhand eines Beispiels aus dem Gebiet der Chirurgie erklärt wird:

Ein Patient oder eine Patientin mit vielen Komorbiditäten wird einem komplexen chirurgischen Eingriff unterzogen. Führt dieser nun zu einem guten Ergebnis, ist das Operationsteam zufrieden, da der Patient oder die Patientin durch die gute Arbeit gerettet werden konnte, das intensivmedizinische Team ist zufrieden, weil der Patient oder die Patientin durch die nachfolgende Therapie wieder remobilisiert und in sein Alltags- und Berufsleben eingegliedert werden kann und die Angehörigen und der oder die Betroffene selbst sind natürlich zufrieden, wenn ein solcher Eingriff und die darauffolgende Therapie reibungslos verläuft, was in erster Linie weiterleben bedeutet. Diese Form von Qualität oder einem guten Ergebnis lässt sich allerdings weniger leicht an klar definierten Parametern festmachen.

Nun kommt die Frage auf, wie die Verarbeitung und Implementierung jener Begriffe in den Klinikalltag funktioniert. Daher muss Qualitätsmanagement als Form der Klinikleitung mit der stetigen Bemühung der Verbesserung in den Bereichen der Struktur-, Prozess- und Ergebnisqualität als Priorität angesehen werden. Dies beinhaltet sowohl Punkte wie den optimierten Einkauf und eine klar strukturierte Abteilungsorganisation, als auch das Anstreben einer möglichst kurzen Verweildauer der Patienten und Patientinnen durch Optimierung der Abläufe.

Als messbare Qualitätsfaktoren gelten mittlerweile auch verbesserte klinische Ergebnisse bezüglich der Mortalität und Morbidität, die damit verbundene höhere Zufriedenheit von Patienten und Patientinnen und Angehörigen, aber auch die der eigenen Mitarbeiter und Mitarbeiterinnen.

An einem von Konkurrenzdruck geprägten Markt kann schließlich auch das Erreichen von Zertifikaten unter anderem zu Marketingzwecken einen klaren Vorteil darstellen.

Somit ist Qualitätsmanagement auch ein auf eine spezielle Organisation, wie eine Fachabteilung, bezogenes Verfahren, welches sich um stetige Verbesserung bemüht und sich dabei an den individuellen Interessen der Mitarbeiter und Mitarbeiterinnen, Patienten und Patientinnen und zuweisenden Parteien berücksichtigt.

Die statistische Prozesskontrolle existiert bereits seit Jahrzehnten und ist ein wichtiges Werkzeug zur Verbesserung der Qualität von Produkten oder Dienstleistungen. Nun besteht aber die Anforderung für die Anwendung derselben im Gesundheitswesen darin, dass eine Anpassung an das individuelle Risiko einzelner Patienten und Patientinnen erfolgt. Andernfalls wäre die Folge eine automatische schlechtere Beurteilung jener Kliniken, Abteilungen und Chirurgen und Chirurginnen, welche auch risikoreichere Patienten und Patientinnen annehmen und operieren, da diese aufgrund des Hochrisikopatientengutes natürlich mehr Todesfälle aufweisen. Solche Zahlen dürfen aber eben nur in Verbindung mit dem Risiko der Operierten interpretiert werden. Zu diesem Zwecke verwendet man verschiedene Methoden, eine davon ist das variable life-adjusted display (VLAD). (siehe Abschnitt 2.1.2)

Um der Anforderung nachzukommen, dass Daten nicht nur gesammelt und verarbeitet werden, sondern dass auch abweichende Muster erkannt und behandelt werden und in weiterer Folge zu Adjustierungen im Prozessablauf führen, müssen Schwellenwerte definiert werden, ab wann man nicht mehr von „Normalität“ spricht und ab welchen weitere Schritte eingeleitet werden. ⁽¹⁶⁾

Zwar ist der personelle Aufwand zur Einführung und Umsetzung eines suffizienten Qualitätssicherungssystems innerhalb einer Klinik sehr hoch, allerdings wird dieser durch die zu erreichende Verbesserung jedenfalls gerechtfertigt. ⁽¹⁷⁾

1.2. Der EuroSCORE

Das primär in der Herzchirurgie verwendete Qualitätsmerkmal einer Operation ist die 30-Tages-Mortalität. Diese zeigt auf, ob ein Patient bzw. eine Patientin 30 Tage nach dem Eingriff überlebt (positiver Ausgang), oder nicht (negativer Ausgang), unabhängig vom momentanen Zustand dieses Patienten bzw. dieser Patientin. Allerdings ist die 30-Tages-Mortalität isoliert betrachtet ein unbrauchbares Mittel, um die Qualität einer Operation objektiv bewerten zu können. Das Spektrum herzchirurgischer Eingriffe reicht von alltäglichen „Standardeingriffen“, wie zum Beispiel einem einfachen aortokoronaren Bypass, bis hin zu äußerst komplexen Eingriffen, welche natürlich auch ein unterschiedliches Risiko aufweisen. Neben der Art der Operation sind auch weitere Faktoren wie die Anamnese des Patienten oder der Patientin wie Alter,

körperliche Gesamtkonstitution und noch weitere Kriterien entscheidend für die Abschätzung des Risikos eines Eingriffes. Somit lässt sich erkennen, dass es kein objektiver Maßstab für die Qualität des Ergebnisses sein kann, die bloßen Zahlen der nach 30 Tagen überlebenden bzw. verstorbenen Patienten und Patientinnen heranzuziehen. Aus diesem Grunde gibt es weltweit eine Reihe statistischer Ansätze, die Sterbewahrscheinlichkeit eines Patienten bei herzchirurgischen Eingriffen unter Berücksichtigung seiner Risikofaktoren vorausszusagen. Einer dieser Ansätze ist der EuroSCORE.

EuroSCORE steht für European System for Cardiac Operative Risk Evaluation. Es handelt sich um ein System zur Berechnung der Wahrscheinlichkeit der 30-Tages-Mortalität im Rahmen von herzchirurgischen Eingriffen unter Einbeziehung verschiedenster relevanter Parameter. Die Idee dahinter ist, dass die Mortalität bei Herzoperationen ein guter Maßstab für die Qualität solcher ist, wenn man die Risikofaktoren der jeweiligen Patienten bzw. Patientinnen zur Auswertung heranzieht.^(4,6,21)

Aufgrund des sich verändernden Risikoprofils in der Herzchirurgie waren rohe perioperative Sterbezahlen nicht mehr ausreichend für die Aufklärung der Patienten und Patientinnen oder die Beurteilung der Qualität einzelner Krankenhäuser. Zwar gibt es das Argument, dass Sterbedaten zur Qualitätsmessung reichen, da sie auch die Entscheidung der wohlüberlegten Patientenwahl widerspiegeln, jedoch ist dies gerade in der Herzchirurgie aufgrund des Risikoparadox nicht anwendbar: Besonders bei Hochrisikopatienten und -patientinnen kann die Überlegenheit der operativen über die konservative Therapie gezeigt werden. Außerdem ist die Sterblichkeit nicht der einzig wichtige Faktor, da auch überlebende Patienten und Patientinnen mit hoher Morbidität und schlechten Langzeitergebnissen zu kämpfen haben. Trotzdem resultieren diese beiden Faktoren aus einem nicht zufriedenstellendem operativen Outcome, was wiederum der Hauptgrund für die Mortalität ist. Somit ist die risikobezogene Mortalität der Grundstein für eine adäquate Qualitätsbewertung.

Zur Entwicklung des Systems wurden die Daten von fast zwanzigtausend aufeinanderfolgenden OPs an herzchirurgischen Patienten und Patientinnen in 128 Krankenhäusern in acht europäischen Ländern herangezogen. Das Outcome (Überleben oder Tod) wurde dann in Relation zu den vorbestehenden Risikofaktoren gesetzt, wodurch 97 solcher ausfindig gemacht werden konnten. Die wichtigsten,

verlässlichsten und objektivsten Risikofaktoren wurden dann verwendet, um das Berechnungssystem vorzubereiten. Es wurde aus einem Teil der Daten entwickelt, der andere Teil diente der Überprüfung der Anwendbarkeit des Systems.

In verschiedenen Studien wurden die für das Outcome wichtigsten Risikofaktoren studiert und überprüft, was dann zur endgültigen Festlegung der im EuroSCORE verwendeten Parameter geführt hat. Es gibt drei Gruppen von Faktoren: Patientenbezogene, Herzbezogene und Operationsbezogene.

Zu den patientenbezogenen Faktoren, aufgeführt mit ihrer Gewichtung im numerischen EuroSCORE, zählen folgende: Alter: 1 (ab 60 bzw. in fünf-Jahres-Intervallen + 1), Geschlecht: 1 (weibliches Geschlecht), Chronische Lungenerkrankung: 1 (langfristige Therapie mit bronchodilatativen Medikamenten oder Steroiden), extrakardiale Arteriopathie: 2 (Claudicatio und/oder Carotisstenose >50% und/oder stattgefundene oder geplante Intervention an Aorta abdominalis, Extremitätenarterien oder Carotiden), neurologische Dysfunktion: 2 (starke Einschränkung des Gehens oder des täglichen Lebens), vorangegangene Herzoperation: 3 (Öffnen des Pericards), Serumkreatinin: 2 (>200 µmol/L präoperativ), aktive Endokarditis: 3 (antibiotische Behandlung zum Zeitpunkt der Operation), kritischer präoperativer Zustand: 3 (Ventrikuläre Tachykardie/Kammerflimmern/abgewendeter plötzlicher Herztod und/oder präoperative Herzdruckmassage und/oder präoperative Beatmung vor der Ankunft im Schockraum und/oder Verabreichung positiv inotroper Medikamente, intraaortale Ballongegenpulsation/präoperatives akutes Nierenversagen).

Die herzbezogenen Faktoren sind instabile Angina Pectoris: 2 (Ruheangina mit Verabreichung von i.v. Nitraten vor der Ankunft im Schockraum), LV-Dysfunktion: (LVEF 30-50%: 1, LVEF <30%: 3), Rezenter Herzinfarkt: 2 (<90 Tage), pulmonaler Hypertonus: 2 (Systolischer Pulmonalarteriendruck >60 mmHg).

Die operationsbezogenen Faktoren sind die folgenden: Notfall: 2 (Durchführung bei der Einweisung vor dem nächsten Arbeitstag), anderes als isolierter CABG: 2, Operation der thorakalen Aorta: 3, Septumruptur nach einem Herzinfarkt: 4.

Nach der Phase der Entwicklung wurde der EuroSCORE im Jahr 1999 erstmals veröffentlicht und in verschiedensten europäischen Krankenhäusern an den eigenen Patienten und Patientinnen überprüft und auch im Vergleich zu anderen, schon bestehenden Scoring-Systemen getestet. Hier wurde er bezüglich seiner Aussagekraft oft mit schon bestehenden oder hausinternen Systemen verglichen. ⁽²¹⁾

Ursprünglich wurde der EuroSCORE als anwenderfreundliches System, welches auch am Patientenbett verwendet werden kann, geplant, daher wurde zuerst nur die numerische Version publiziert, wenngleich die Berechnungen auf dem Boden der logistischen Regression beruhen. Aus diesem Grund konnte das System in weiten Teilen Europas, vor allem auch in technisch weniger weit entwickelten Ländern, angewendet werden.

Erst mit dem steigenden Interesse von KardiologInnen und HerzchirurgInnen an einem komplexeren System als dem doch sehr simplen additiven EuroSCORE wurden die Details der logistischen Regressionsgleichung des EuroSCORE veröffentlicht. Der logistische EuroSCORE produziert akkuratere Ergebnisse, vor allem bei Hochrisikopatienten und -patientinnen, der wichtigste Nachteil ist die kompliziertere Berechnung des Ergebnisses. ^(18,4)

Die Funktionsweise des Systems ist einfach: Jedem Risikofaktor ist eine Zahl bzw. eine Gewichtung zugeteilt. Wenn nun einer dieser Risikofaktoren vorhanden ist, werden die jeweiligen Punkte zugewiesen und im Fall des numerischen EuroSCORE addiert, wodurch man eine Zahl erhält, die die ungefähre Höhe des Mortalitätsrisikos in Prozent angibt.

Patientenbezogene Faktoren		Score/Beta
Alter	Pro fünf Jahre oder ab 60	1/0.0666354
Geschlecht	Weiblich	1/0.3304052
Chronische Lungenerkrankung	Langfristige Therapie mit bronchodilatativen Medikamenten oder Steroiden gegen die Lungenerkrankung	1/0.4931341
extrakardiale Arteriopathie	Claudicatio und/oder Carotisstenose >50% und/oder stattgefundene oder geplante Intervention an Aorta abdominalis, Extremitätenarterien oder Carotiden	2/0.6558917
neurologische Dysfunktion	Starke Einschränkung des Gehens oder des täglichen Lebens	2/0.841626
vorangegangene Herzoperation	Mit Öffnung des Perikards	3/1.002625
Serumkreatinin	>200 µmol/L präoperativ	2/0.6521653
aktive Endokarditis	Antibiotische Behandlung zum Zeitpunkt der Operation	3/1.101265
kritischer präoperativer Zustand	Ventrikuläre Tachykardie/Kammerflimmern/ abgewendeter plötzlicher Herztod und/oder präoperative Herzdruckmassage und/oder präoperative Beatmung vor der Ankunft im Schockraum und/oder Verabreichung positiv inotroper Medikamente, intraaortale Ballongegenpulsation/ präoperatives akutes Nierenversagen	3/0.9058132

Tabelle 1: Die patientenbezogenen Risikofaktoren und deren Gewichtung im numerischen sowie logistischen EuroSCORE ⁽²¹⁾

Herzbezogene Faktoren		Score/ Beta
Instabile Angina Pectoris	Ruheangina mit Verabreichung von i.v. Nitraten vor der Ankunft im Schockraum	2/0.5677075
LV-Dysfunktion	Moderat oder LVEF 30-50% Schlecht oder LVEF <30	1/0.4191643 3/1.094443
Rezenter Herzinfarkt	(<90 Tage)	2/0.5460218
pulmonaler Hypertonus	Systolischer Pulmonalarteriendruck >60 mmHg	2/0.7676924

Tabelle 2: Die herzbezogenen Risikofaktoren und deren Gewichtung im numerischen sowie logistischen EuroSCORE ⁽²¹⁾

Operationsbezogene Faktoren		Score/Beta
Notfall	Durchführung bei der Einweisung vor dem nächsten Arbeitstag	2/0.7127953
Anderes als isolierter CABG	Größere Herzoperation anstatt/zusätzlich zu CABG	2/0.5420364
Operation der thorakalen Aorta	Probleme mit Aorta ascendens, Arcus Aorticus, Aorta descendens	3/1.159787
Septumruptur nach einem Herzinfarkt	Ja	4/1.462009

Tabelle 3: Die operationsbezogenen Risikofaktoren und deren Gewichtung im numerischen sowie logistischen EuroSCORE ⁽²¹⁾

Verwendet man den logistischen EuroSCORE, erhält man ein prozentuelles Risiko der Mortalität, abgeleitet aus den Ergebnissen eines logistischen Regressionsmodells, in dem der Einfluss der einzelnen Faktoren auf das Risiko der 30-Tages-Mortalität modelliert wird. Dadurch erhält man ein sehr viel genaueres Ergebnis. ⁽²⁰⁾

Dieses wird wie folgt berechnet:

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

Hierbei ist e der natürliche Logarithmus, β_0 die Konstante der logistischen Regressionsgleichung, β_i ist der Koeffizient der Variable X_i in der logistischen Regressionsgleichung aus der oben angeführten Tabelle. X_i ist 1, wenn ein kategorischer Risikofaktor vorhanden ist und 0, wenn er fehlt. Beim Alter nimmt X_i ab 60 Jahren um 1 pro Jahr zu.

Welcher der beiden Scores nun verwendet wird, hängt vor allem vom Arzt oder der Ärztin selbst ab. Der numerische EuroSCORE ist sehr einfach zu berechnen, kann allerdings fallweise, besonders bei Hochrisikopatienten und -patientinnen, das Risiko unterschätzen, vor allem, wenn gewisse Kombinationen verschiedener Faktoren vorliegen. Der logistische EuroSCORE produziert ein genaueres Ergebnis, ist aber aufwendiger zu berechnen.

Seit 2011 gibt es nun einen neuen, überarbeiteten EuroSCORE II, welcher entwickelt wurde, um die Mängel der Vorgänger auszugleichen. Dieser wurde aus einem moderneren Datensatz erstellt und wurde so weiterentwickelt, dass er evidenzbasierte Verbesserungen und die Fortschritte in der modernen Herzchirurgie berücksichtigt. Er wurde aus einer Kohorte von 16.828 Patienten und Patientinnen aus 154 Zentren in 43 Ländern über eine Zeit von 12 Wochen berechnet und dann an einer anderen Patientengruppen von 5553 Personen getestet um seine Validität zu schätzen. ⁽²⁰⁾

Der EuroSCORE II wurde um einige Risikofaktoren ergänzt, andere dagegen wurden entfernt und es gibt zusätzlich neue Auswahlmöglichkeiten. So ist nun nicht mehr die Höhe des Serumkreatinins anzugeben, sondern mithilfe eines Rechners kann man die Kreatininclearance berechnen und dann aus vier Optionen die Stärke der renalen Beeinträchtigung wählen. Anstatt der neurologischen Dysfunktion ist nun eine allgemeine Beeinträchtigung der Mobilität ein Faktor und außerdem wurde auch insulinpflichtiger Diabetes hinzugefügt.

Patientenbezogene Faktoren		Beta
Alter		0,0285181
Geschlecht	weiblich	0,2196434
	Dialyse	0,6421508
Renale Beeinträchtigung	CC ≤ 50 (ml/min)	0,8592256
	CC 50 – 85 (ml/min)	0,303553
extrakardiale Arteriopathie	Claudicatio und/oder Carotisstenose>50% und/oder stattgefundene oder geplante Intervention an Aorta abdominalis, Extremitätenarterien oder Carotiden	0,5360268
Eingeschränkte Mobilität	Aufgrund muskuloskeletaler oder neurologischer Dysfunktion	0,2407181
vorangegangene Herzoperation	Mit Öffnung des Perikards	0,1118599
Chronisch obstruktive Lungenerkrankung	Langzeitmedikation mit Bronchodilatoren oder Steroiden	0,1886564
aktive Endokarditis	Antibiotische Behandlung zum Zeitpunkt der Operation	0,6194522
kritischer präoperativer Zustand	Ventrikuläre Tachykardie/Kammerflimmern/ abgewendeter plötzlicher Herztod und/oder präoperative Herzdruckmassage und/oder präoperative Beatmung vor der Ankunft im Schockraum und/oder Verabreichung positiv inotroper Medikamente, intraaortale Ballongegenpulsation/ präoperatives akutes Nierenversagen	1,086517
IDDM	Insulinabhängiger Diabetes Mellitus	0,3542749

Tabelle 4: Die patientenbezogenen Risikofaktoren und deren Gewichtung im EuroSCORE II ⁽²¹⁾

Herzbezogene Faktoren		Beta
NYHA	I - Keine Symptome	0
	II - Symptome bei moderater Belastung	0,1070545
	III - Symptome bei leichter Belastung	0,2958358
	IV - Symptome in Ruhe	0,5597929
CCS Klasse 4 Angina	Angina in Ruhe	0,2226147
	LVEF > 50 %	0
LV-Funktion	LVEF 31 – 50 %	0,3150652
	LVEF 21 – 30 %	0,8084096
	LVEF < 21%	0,9346919
Rezenter Herzinfarkt	(<90 Tage)	0,1528943
pulmonaler Hypertonus	31 – 55 mmHg	0,1788899
	>55 mmHg	0,3491475

Tabelle 5: Die herzbezogenen Risikofaktoren und deren Gewichtung im EuroSCORE II ⁽²¹⁾

Operationsbezogene Faktoren		Beta
Dringlichkeit der Operation	Elektiv	0
	Dringend: OP notwendig vor Entlassung	0,3174673
	Notfall: OP noch am Aufnahmetag	0,7039121
	Lebensbedrohend: Reanimation auf dem Weg in den OP oder vor Einleitung der Anästhesie	1,362947
Anderes als isolierter CABG	Eine größere Herz-OP außer CABG	0,0062118
	Zwei größere Herz-OP	0,5521478
	Drei oder mehr größere Herz-OP	0,9724533
Operation der thorakalen Aorta	Probleme mit Aorta ascendens, Arcus Aorticus, Aorta descendens	0,6527205

Tabelle 6: Die operationsbezogenen Risikofaktoren und deren Gewichtung im EuroSCORE II ⁽²¹⁾

Bei den herzbezogenen Faktoren ist nun zusätzlich zur Ruhe Angina Pectoris das NYHA Stadium anzugeben und bei den operationsbezogenen wurde der Septumdefekt nach Herzinfarkt gestrichen. (siehe auch nachfolgende Abbildung 1 in englischer Fachsprache) Verschiedene Studien überprüfen, ob das überarbeitete Modell besser ist, als seine Vorgänger, jedoch variiert das Ergebnis in verschiedenen Kliniken und Ländern.

Cardiac-related factors		
Unstable angina ⁶	No ▾	0
LV function	Select ▾	0
Recent MI ⁷	No ▾	0
Pulmonary hypertension ⁸	No ▾	0
Operation-related factors		
Emergency ⁹	No ▾	0
Other than isolated CABG	No ▾	0
Surgery on thoracic aorta	No ▾	0
Post infarct septal rupture	No ▾	0
Patient-related factors		
Age (years)	0	0
Gender	Select ▾	0
Chronic pulmonary disease ¹	No ▾	0
Extracardiac arteriopathy ²	No ▾	0
Neurological dysfunction ³	No ▾	0
Previous Cardiac Surgery	No ▾	0
Creatinine > 200 µmol/ L	No ▾	0
Active endocarditis ⁴	No ▾	0
Critical preoperative state ⁵	No ▾	0
Logistic ▾ EuroSCORE		0

Abbildung 1: Maske im Onlinekalkulator für den logistischen EuroSCORE
(21)

Die EuroSCORE Rechner können über einen Onlinekalkulator der Website www.euroscore.org abgerufen und ohne viel Aufwand verwendet werden. Durch diese

Art der Risikoabschätzung sind die Unterschätzung der „risikoarmen“ sowie die Überschätzung von Hochrisikopatienten und -patientinnen bereinigt.

Als ein entsprechendes Maß für die Voraussagbarkeit des EuroSCOREs gilt die AUC („Area under the Curve“, Fläche unter der Kurve) unter der ROC-Kurve („Receiver-Operating-Characteristic“). Im Idealfall ist diese 1,0, als aussagekräftig gelten Werte $>0,7$, darunterliegende Werte deuten auf einen Zufallsprozess hin. (siehe Abschnitt 2.1.1)

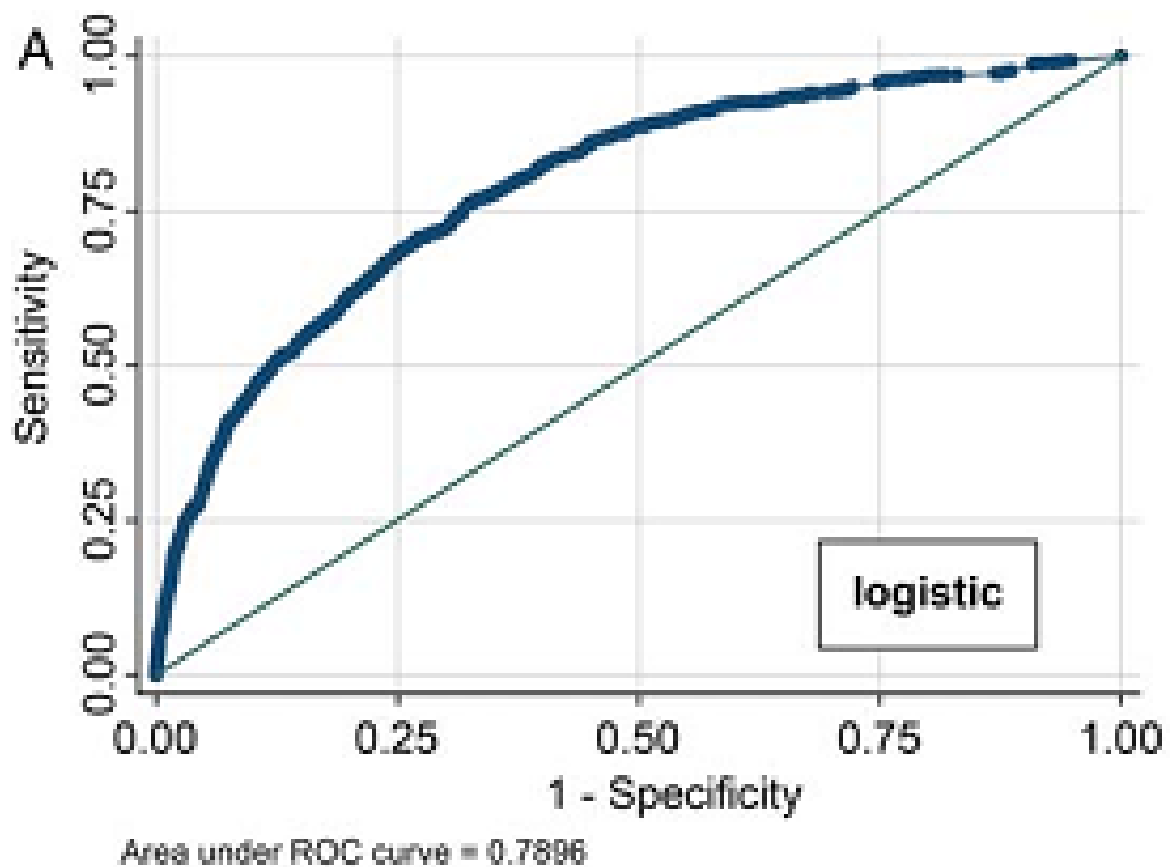


Abbildung 2: ROC-Kurve

1.3. Bypasschirurgie

1.3.1. Grundlagen

Die Blutversorgung des Herzmuskels erfolgt über die rechte und linke Koronararterie, welche direkt proximal der Aortenklappe aus der Aorta abgehen. Die linke Koronararterie verläuft etwa 2 cm als linker Hauptstamm, bis sie sich in folgende Gefäße aufteilt ^[11]:

- Der **Ramus interventricularis anterior (RIVA)** verläuft zur Herzspitze und versorgt die Vorderwand des linken Ventrikels, Teile des rechten Ventrikels, das Kammerseptum und den linksanterioren Faszikel des linken Tawaraschenkels. Außerdem gibt er zwei oder auch mehrere Diagonaläste zur anterioren Wand des linken Ventrikels ab.
- Der **Ramus circumflexus (RCX)** verläuft in der Atrioventrikularfurche und versorgt den anterolateralen, lateralen und posterolateralen Teil des linken Ventrikels sowie den linken Vorhof.

Die rechte Koronararterie gibt einige Äste zur Versorgung des Sinusknotens, des Conus pulmonalis, des rechten Vorhofes sowie des rechten Ventrikels ab, bis sie sich im Bereich des Crux Cordis in folgende Äste aufteilt:

- Der **Ramus interventricularis posterior (RIVP)** versorgt den Großteil der Hinterwand des linken Ventrikels, den hinteren Teil des Kammerseptums, das His-Bündel sowie den posterioren Faszikel des linken Tawaraschenkels.
- Der **Ramus posterolateralis** zieht zur diaphragmalen Hinterwand des linken Ventrikels sowie zum AV-Knoten.

Hierbei handelt es sich um die Normvariante, allerdings ist die Blutgefäßversorgung des Herzen durch Kollateralen verbunden und starken individuellen Schwankungen unterlegen.

Kommt es durch arteriosklerotisch bedingte Veränderungen der Gefäßwände zu Stenosen der Koronararterien, führt dies ab einer Verengung um 70% zu einer

Minderdurchblutung des Gewebes, welches vom betroffenen Ast versorgt wird. Bis zu diesem kritischen Stenosegrad ist die Blutversorgung in Ruhe noch genügend, da sich die Arteriolen maximal dilatieren und es auch über Kollateralen noch zu einem ausreichenden Blutfluss kommt.

Bei körperlicher Belastung allerdings ist der Sauerstoffbedarf des Herzens erhöht, was eine verstärkte Durchblutung durch weitere Dilatation erforderlich macht. Dieser Anforderung können die stenosierten Gefäße nicht mehr gerecht werden, wodurch es im nachfolgenden Gebiet zu einer Minderversorgung kommt.

Dieses Missverhältnis zwischen Sauerstoffangebot und –bedarf des Herzmuskels mit Myokardischämie wird als Koronare Herzkrankheit bezeichnet. Je nach den betroffenen Gefäßen spricht man von einer Ein- Zwei- oder Dreigefäßerkrankung. Wesentliche Risikofaktoren hierfür sind Hypercholesterinämie, Nikotinabusus, arterieller Hypertonus, Alter, Diabetes mellitus, Adipositas, Stress und Bewegungsmangel. Eine Myokardischämie kann asymptomatisch oder symptomatisch sein und sich in diversen Formen äußern ⁽¹¹⁾:

- Stabile Angina pectoris
- Akutes Koronarsyndrom
- Ischämische Herzmuskelschädigung mit Linksherzinsuffizienz
- Herzrhythmusstörungen
- Plötzlicher Herztod

Das Leitsymptom stellt die Angina pectoris dar, die sich durch retrosternalen Druck und Engegefühl mit Ausstrahlung in den Hals und den linken Arm sowie ins Abdomen in Verbindung mit Atemnot präsentiert. Häufig kommt es auch zu Übelkeit und Erbrechen.

Die Diagnostik erfolgt mittels Anamnese und körperlicher Untersuchung, Belastungs-EKG, Myokardszintigraphie, Koronarangiographie und Computertomographie.

Grundsätzlich gibt es für dieses Krankheitsbild medikamentöse, interventionelle und operative Therapiemöglichkeiten, alle mit dem Ziel der Reduktion der Angina-pectoris-Anfälle, der Verbesserung der Lebensqualität, der Verbesserung der Belastbarkeit, der

Reduzierung der kardiovaskulären Komplikationsrate sowie der Reduzierung der Mortalität.

Die medikamentöse Therapie umfasst die Behandlung mit Betablockern, ACE-Hemmern, AT1-Antagonisten, Cholesterinsenkern, Nitraten und Thrombozytenaggregationshemmern. Die interventionelle Therapie besteht in der perkutaten Dilatation des stenosierten Gefäßes mit einem Ballonkatheter und dem Einsetzen eines Stents. Sind diese Therapiemöglichkeiten ausgeschöpft, ist die koronararterielle Bypassoperation die letzte Möglichkeit, Beschwerdenlinderung und Steigerung der Lebensqualität betroffener Patienten und Patientinnen zu erreichen. [14]

Die Voraussetzung für längere Operationen am Herzen ist das Pausieren des Schlagens des Herzen und die externe Übernahme der Kreislaufarbeit, was durch die Verwendung der Herz-Lungen-Maschine und der Kardioprotektion geschehen kann.

1.3.2. Herz-Lungen-Maschine

Der Anschluss der Herz-Lungen-Maschine erfolgt zunächst über einen Katheter, welcher über den rechten Vorhof in die Venae cavae superior und inferior eingeführt wird. Ein zweiter Katheter wird distal des Abganges der Koronararterien in die Aorta eingeführt. Das Blut wird aus den Venae cavae entnommen und mithilfe einer Zentrifugalpumpe über das ableitende Schlauchsystem einem Oxygenator zugeführt, welcher das venöse Blut über eine Membran mit Sauerstoff anreichert. Danach gelangt das Blut über einem Wärmeaustauscher, welcher die Temperatur absenkt, zurück in die Aorta ascendens. Eine Alternative ist der Zugang über die Femoralgefäße. Grundvoraussetzung der Verwendung der HLM ist das Herabsetzen der Gerinnungsfähigkeit des Blutes, da die Kunststoffschläuche eine erhebliche Thrombogenität aufweisen, weswegen dem Patienten oder der Patientin vor Anschluss der Maschine 2-3 mg Heparin pro Kilogramm Körpergewicht verabreicht werden. Die Überwachung der Heparinisierung erfolgt mittels ACT-Test (activated-clot-time). Der Normalwert der ACT beträgt 80-100 Sekunden, während der Zeit an der Maschine beträgt der erwünschte Wert 400-600 Sekunden.

1.3.3. Kardioprotektion

Um in Ruhe am Herzen arbeiten zu können und außerdem eine Luftembolisation zu vermeiden, wird das Herz während der Operation unter Hypothermie in einen kardioplegischen Stillstand versetzt. Das Herz wird von der Zirkulation durch Abklemmen der Aorta über den Abgängen der Koronararterien sowie Abklemmen des rechten Vorhofs getrennt und dann mit 4°C kalter Kardioplegielösung über die Koronararterien und fallweise auch über den Sinus coronarius perfundiert, wodurch die Ischämietoleranz des Herzens erheblich verlängert wird. ⁽¹⁰⁾

1.3.4. Indikation

Die Indikation für die operative Sanierung von Koronarstenosen stellen vor allem Herzgefäßerkrankungen dar, welche aus technischen Gründen nicht mittels PCTA versorgt werden können und bei denen die konservativen Therapiemöglichkeiten ausgeschöpft sind und nicht zu ausreichender Beschwerdebesserung führen. ⁽¹⁴⁾ Weitere Indikationen sind eine Hauptstammstenose, Dreifgefäßerkrankungen mit Bypass-fähigen distalen Koronararterien und Zweifgefäßerkrankungen mit Beteiligung des RIVA beim Linksversorgungstyp außerdem besteht als Notfallsindikation noch der kardiogene Schock. ⁽¹⁰⁾

1.3.5. Prozedere

Der Zugang erfolgt über eine mediane Längssternotomie. Nach Freipräparation des Herzen und der großen Gefäße erfolgt unter Vollheparinisierung der Anschluss an die Herz-Lungen-Maschine und die Perfusion mit der Kardioplegielösung. Eingefäß-Bypass-Operationen am Ramus interventricularis anterior werden häufig auch ohne Herz-Lungen-Maschine am schlagenden Herzen durchgeführt (Off-pump-Chirurgie, MIDCAB: minimally invasive direct coronary artery bypass). Ist dies der Fall, erfolgt der Zugang über eine linkslaterale Minithorakotomie, als Graft wird die linke Arteria

thoracica interna verwendet und das Herz wird beim Nähen der Anastomose mit einer Saugglocke oder einer speziellen Haltevorrichtung, dem sogenannten Oktopus, fixiert. Zur aortokoronaren Bypassoperation gibt es zwei Möglichkeiten:

- Arteria mammaria interna Bypass: Die Arteria thoracica interna, welche im klinischen Gebrauch oftmals als Arteria mammaria (IMA: internal mammary artery) bezeichnet wird, wird von distal aus mit oder ohne umliegendes Bindegewebe bis zum proximalen Ursprung aus der Arteria subclavia abpräpariert. Reicht diese nicht aus, weil mehrere Gefäße betroffen sind oder ist diese zu kaliberschwach und daher nicht als Graft geeignet, wird zusätzlich ein venöser Bypass nötig.
- Aortocoronarer Venenbypass (ACVB): Hierfür wird meist die Vena saphena magna aus dem Bein des Patienten/der Patientin entnommen.

Nachdem die passenden Grafts freipräpariert wurden, erfolgt das Nähen der distalen Anastomosen, distal der stenosierten Stelle des betroffenen Gefäßes. Die Anastomosen werden entweder einzeln oder sequentiell mittels End-zu-Seit oder End-zu-End-Technik mittels 7/0-Prolenefäden angelegt. Anschließend erfolgt bei venösen Bypässen das Anlegen der proximalen Anastomose mittels End-zu-Seit-Technik an die Aorta bei wieder schlagendem Herzen zur Verkürzung der Ischämiezeit. Nach genauester Blutstillung und mehrmaliger Überprüfung der Anastomosen nach dem schrittweisen Abgehen von der Herz-Lungen-Maschine wird die Sternotomie mithilfe von Drahtzerklagen wieder verschlossen und es erfolgt der schrittweise Wundverschluss.⁽¹⁰⁾

1.3.6. Prognose

Aufgrund der stetigen Verbesserung interventioneller Verfahren nimmt die Zahl der durchzuführenden Bypass-Operationen immer weiter ab. Bei den Operationen zeigt der Arteria thoracica interna-Bypass die beste Offenheitsrate mit 80-95% nach 5 Jahren, während der Venenbypass eine Offenheitsrate von 60-80% nach 5 Jahren aufgrund vermehrter Bypassverschlüsse durch arteriosklerotische Veränderungen aufweist. Die intraoperative Letalität beträgt etwa 1-3%, ist aber bei Notfallseingriffen erhöht.

Im ersten Jahr nach der Operation präsentieren sich etwa 80-90% der Patienten beschwerdefrei, nach 5 Jahren sind es noch ca. 70%.

Insgesamt weisen die interventionellen Verfahren eine schlechtere Langzeitprognose auf, es gibt eine höhere Rate an Reinfarkten und erneute Eingriffe werden etwa doppelt so häufig erforderlich. ⁽¹⁰⁾

1.3.7. Risiken

Die Risiken bei der Aortokoronaren Bypassoperation umfassen sowohl das Operationsrisiko an sich als auch das Narkoserisiko.

Im Rahmen der Narkose ist es nötig, folgende Schritte durchzuführen: Anlegen des zentralvenösen Zuganges, Anlegen eines arteriellen Katheters mit invasiver Blutdruckmesssonde, Intubation sowie Legen eines Blasenkatheters. Jeder dieser Schritte bringt eigene Risiken mit sich. So kann es beim Anlegen des zentralvenösen Zuganges und der arteriellen Sonde zu Blutungen kommen, es kann zur Punktion eines falschen Gefäßes kommen und auch die Ablösung eines Plaques mit nachfolgender Embolie oder Infarzierung eines Organs ist ein gefürchtetes Risiko. Folgen der Intubation können Heiserkeit, Stimmbandlähmungen und im Extremfall auch die Perforation der Trachea mit nachfolgender Mediastinitis sein. Auch beim Legen des Blasenkatheters kann es zu Verletzungen im Bereich der Urethra oder der Blase kommen, was zu einer Zystitis, Urethritis oder Pyelonephritis führen kann. Die Narkose selbst birgt durch die dabei verwendeten Medikamente ein gewisses Risiko, so kann es zum Beispiel zu Unverträglichkeiten oder Allergien kommen. Eine gefürchtete Komplikation ist die maligne Hyperthermie, welche nach jeder Narkose auftreten kann. ⁽¹²⁾

Komplikationen als Folge der Myokardischämie umfassen Herzrhythmusstörungen, Myokardinfarkt, plötzlichen Herztod und das Dressler Syndrom oder Postmyokardinfarktsyndrom, welches mit Fieber, Perikarditis und Herzinsuffizienz einhergeht.

Die eingriffsbedingten Risiken umfassen den perioperativen Schlaganfall, Zugangskomplikationen im Sinne von Wundheilungsstörungen am Sternum (in ca. 2-5% der Fälle), Sternumosteomyelitis, Mediastinitis und Instabilität des Sternums,

Wundheilungsstörungen an der unteren Extremität als Folge der Venenentnahme, arteriosklerotische Bypassveränderungen, welche zum Bypassverschluss und zu Restenosen führen können sowie Thrombose und Infektion des Bypasses, Perikarditis und Perikarderguss.

Auch die Verwendung der Herz-Lungen-Maschine birgt einige Risiken, beruhend auf der Notwendigkeit einer Vollheparinisierung des Patienten oder der Patientin. Diese kann zu Blutungen führen (Ulcera, Ösophagusvarizen), zu unspezifischen Entzündungsreaktionen und zu Mikroembolien im Pulmonalkreislauf sowie im ZNS. Beim Abgehen von der Herz-Lungen-Maschine kann es zu hämodynamischer Instabilität kommen, welche im Low cardiac output Syndrom enden kann. ^(11,12,13)

2. Methoden

2.1. Statistische Methoden

2.1.1. ROC-Analysen

2.1.1.1. Grundlagen

ROC Analysen liefern die am besten vergleichbare Beschreibung diagnostischer Genauigkeit,⁽²³⁾ da sie alle Kombinationen von Spezifität und Sensitivität, welche ein diagnostischer Test aufzeigen kann, beurteilen und präsentieren. Im Folgenden werden die sechs Levels, nach welchen diagnostische Effizienz beurteilt werden kann, aufgezeigt und die Grundlagen konventioneller ROC-Analysen und einige Formen von ROC-Kurven, die in Situationen mit mehr als 2 möglichen Ergebnissen verwendet werden können, werden erklärt. Schlüsselfragen, welche zum Thema der Anpassung von ROC-Kurven und zur statistischen Testung aufkommen werden behandelt, um so eine Basis für die Beurteilung von ROC-Studien zu schaffen.

Das NCRP (National Council on Radiation Protection and Measurements) hat ein hierarchisches Modell für diagnostische Effizienz entwickelt, welches einen knappen inhaltlichen Überblick über die Themen, welche in die Evaluation eines diagnostischen Systems entwickelt sind, zeigt. Die sechs Level dieses Modells sind die Folgenden:

- Technische Effizienz: am untersten Level wird ein diagnostischer Test als effektiv gesehen, wenn das Ergebnis im physikalischen Sinne präzise ist, beispielsweise wenn der Test eine oder mehrere physikalische Eigenschaften des menschlichen Körpers so misst, dass es dem Goldstandard entspricht und die Ergebnisse reproduzierbar sind. In der medizinischen Bildgebung beinhalten die Aspekte der technischen Effizienz räumliche und zeitliche Auflösung, Größe und Beschaffenheit und die Kontrastsensitivität.
- Diagnostische Genauigkeit: auf diesem zweiten Level geht es um das Ausmaß, in dem das Ergebnis eines diagnostischen Tests im statistischen Sinne mit dem tatsächlichen Zustand der Gesundheit oder Krankheit eines Patienten übereinstimmt. Nahezu alle praktischen Messungen diagnostischer Genauigkeit quantifizieren die Fähigkeit eines Tests, zwischen zwei Zuständen

der Wahrheit zu unterscheiden, wie zum Beispiel „normal“ und „abnormal“ oder „positiv“ und „negativ“, bezogen auf eine bestimmte Erkrankung. Beispiele zur Messung der diagnostischen Genauigkeit sind Sensitivität und Spezifität und ROC-Kurven. ROC-Kurven bieten hieraus die höchste Vergleichbarkeit, da sie alle Kombinationen von Spezifität und Sensitivität aufzeigen, die ein diagnostischer Test liefern kann.

- Effizienz des diagnostischen Denkens: bei gegebener Prävalenz einer bestimmten Krankheit und bekannter Sensitivität und Spezifität, oder vielmehr vorhandener ROC-Kurve, eines diagnostischen Tests für das Vorhandensein jener Krankheit, kann man leicht den Faktor berechnen, um welchen sich die vorbestehende Wahrscheinlichkeit für das Vorhandensein der Krankheit ändert, wenn das Testergebnis da ist. Das Ausmaß, inwieweit ein Testergebnis die subjektive Einschätzung der Wahrscheinlichkeit einer Krankheit aber beeinflusst, muss empirisch erhoben werden. Dieses Level der Effizienz ist oft schwierig zu quantifizieren, liefert aber eine Verbindung der vorherigen und folgenden Level.
- Therapeutische Effizienz: hierbei handelt es sich um das niedrigste Level, auf welchem die Auswirkung eines diagnostischen Tests auf das Patientenmanagement direkt beurteilt werden kann. Die Grundfrage ist, wie und inwiefern ein bestimmter diagnostischer Test die Behandlung von Patienten verändert, also wie die Therapie mit und ohne bekanntes Testergebnis gewählt wird.
- Therapie des Patienten-Outcomes: hier wird das Ziel der diagnostischen Medizin konfrontiert: ein diagnostischer Test wird auf diesem Level als effizient betrachtet, wenn die Patientengesundheit nachweislich durch die Nutzung dieses Tests verbessert wird. Es ist die Form von Effizienz, die für die meisten Ärzte und Patienten von größtem Interesse ist und ist ein nicht wegdenkbarer Teil einer jeden sinnvollen Kosten-Nutzen-Analyse. Jedoch benötigt die definitive Beurteilung dieses Levels prospektive, randomisierte, kontrollierte klinische Studien, in denen praktische, statistische und ethische Probleme oft auftreten⁽²³⁾.
- Gesellschaftliche Effizienz: jede Kosten-Nutzen-Analyse eines diagnostischen Tests auf Level 5 ist auf die Vorteile und persönlichen Risiken für die Patienten

fokussiert. Die Tatsache jedoch, dass die medizinischen Kosten zunehmend von der Gesellschaft als Ganzes getragen werden, impliziert, dass gesellschaftliche Mittel auch beachtet werden müssen, wenn es zur Kosten-Nutzen-Rechnung kommt. Dies ist die Domäne der gesellschaftlichen Effizienz, wo private und gesellschaftliche Erwägungen vereint werden, um diagnostische Tests im Kontext gesellschaftlicher Bestreben zu betrachten.

Zwar sind die oberen Level der Effizienz eher von direktem Interesse, doch sind die unteren Level leichter verlässlich zu bewerten. Hier ist von Vorteil, dass die höheren Level häufig auf der Basis der unteren geschätzt werden können. Die meisten Studien sind auf die Berechnung der Effizienz im zweiten Level fokussiert, da dies das niedrigste ist, in welche Menschen als Beobachter inkludiert sind und häufig das höchste, auf welchen rigorose statistische Methoden angewandt werden können.

Über viele Jahre wurde die diagnostische Effizienz als eine Art „Schlagdurchschnitt“ berechnet und berichtet: der Prozentsatz diagnostischer Entscheidungen, welche sich als richtig herausstellten. Das Ergebnis dieser Messung allerdings hat die Einschränkung, dass es stark von der Prävalenz einer Krankheit abhängig sein kann: wenn nur 1% der Patienten einer Screening-Gruppe eine Krankheit hat, kann das System 99% Genauigkeit aufweisen, indem es blind alle Patienten als negativ betrachtet. Außerdem kann die prozentuelle Messung keine Aussage über falsch-positive und falsch-negative Fehler treffen, welche aber von Grund auf unterschiedliche klinische Konsequenzen haben⁽²⁴⁾.

Beide Nachteile können behoben werden, indem die diagnostische Leistung im Sinne von Indexpaaren angegeben wird: Die Sensitivität (der Teil der Patienten, der tatsächlich an der Erkrankung leidet und auch als solcher erkannt wird) und die Spezifität (der Teil der Patienten, der gesund ist und als solcher erkannt wird). Diese Indices quantifizieren getrennt die Genauigkeit für tatsächlich positive und tatsächlich negative Patienten. „Falsch-negative“ und „Falsch-positive“ Diagnosen werden nach diesen Indices ausgewiesen und eine Änderung der Prävalenz ändert nichts an deren numerischen Werten, wenn die Entscheidungskriterien dieselben sind. Die Bezeichnungen „Richtig-positive-Fraktion“ und „Richtig-negative-Fraktion“ entsprechen der Sensitivität und der Spezifität. Im Gegensatz dazu repräsentieren die Bezeichnungen „Falsch-negative-Fraktion“ und „Falsch-positive-Fraktion“ die

Wahrscheinlichkeit, mit der tatsächlich positive oder negative Patienten falsch diagnostiziert werden^(24,25).

Somit entspricht die „Falsch-negative-Fraktion“ = $1 - \text{„Richtig-positive-Fraktion“}$ = $1 - \text{Sensitivität}$ und „Falsch-positive-Fraktion“ = $1 - \text{„Richtig-negative-Fraktion“}$ = $1 - \text{Spezifität}$. Aufgrund dieser Wechselbeziehungen muss nur ein einzelnes Paar aufgezeigt werden, meistens werden hier die Indices Sensitivität und Spezifität oder „Richtig-positive-Fraktion“ und „Falsch-negative-Fraktion“ verwendet. Die Verwendung von nur Sensitivität oder „Richtig-positive-Fraktion“ ist unzulässig, da die Performance eines diagnostischen Tests ohne Einbeziehung der negativen Patienten nicht beurteilt werden kann.

Das Paar Sensitivität-Spezifität beschreibt die diagnostische Genauigkeit sinnvoller als die bloße prozentuelle Angabe und wird daher auch überwiegend in medizinischer Literatur genutzt. Allerdings ist auch die bloße Darstellung zweier Zahlen im Sinne von Sensitivität und Spezifität unzureichend, da so zwei Aspekte diagnostischer Effizienz durcheinandergebracht werden, die unabhängig voneinander variieren können: zum Ersten die charakteristische Funktion eines diagnostischen Systems, zwischen tatsächlichen Zuständen von Gesundheit und Krankheit zu unterscheiden, und zum Anderen das Gleichgewicht zwischen der Frequenz falsch-positiver und falsch-negativer Fehler, die derjenige, der die Entscheidung macht, auswählt, um sie in einer klinischen Aufgabe anzuwenden⁽²⁶⁾.

Die Grenzen der Möglichkeit, diagnostische Genauigkeit im Sinne einer bloßen Sensitivität-Spezifität Angabe sind vor allem in Studien offensichtlich, welche darauf abzielen, diagnostische Tests zu vergleichen, da meist ein Test eine höhere Sensitivität aber niedrigere Spezifität hat, als der andere. Anders gesagt kann man mit dem einen Test besser die positiven Patienten finden, der andere ist besser dazu geeignet, die negativen Patienten ausfindig zu machen. Dies macht die Entscheidung zwischen zwei Tests schwierig, da man im Vorhinein nichts über die Präsenz oder Absenz einer Krankheit weiß.

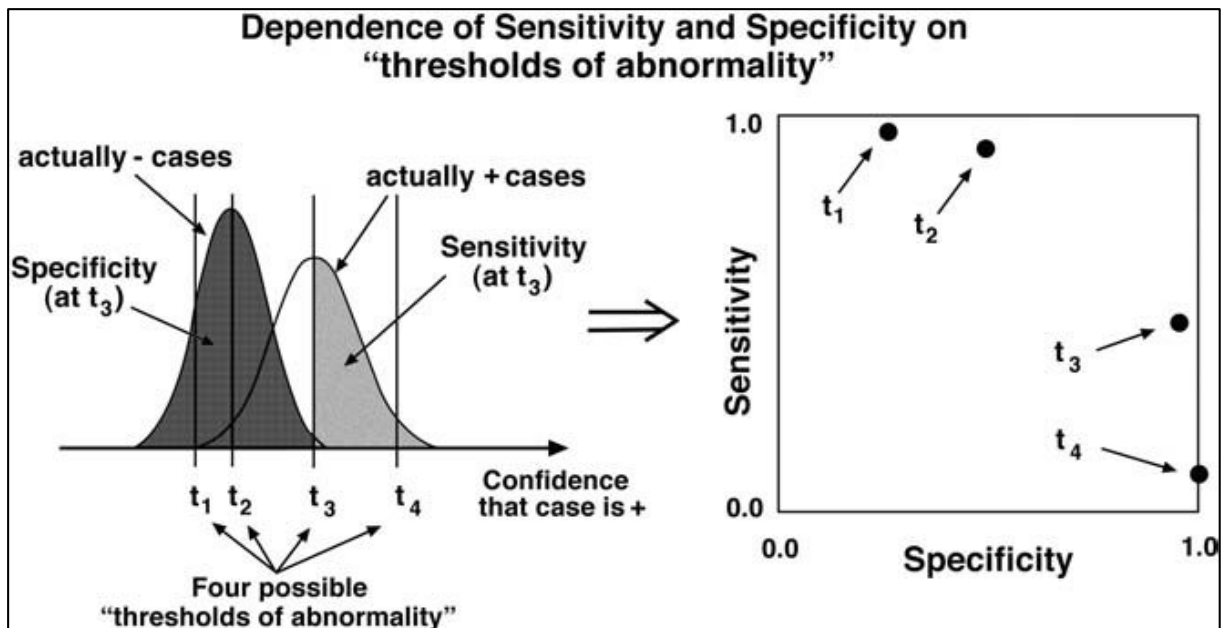


Abbildung 3: Der Effekt des Entscheidungsgrenzwertes bzw. der Abnormitätsschwelle auf Sensitivität und Spezifität. Mit einer Entscheidungsschwelle auf t_3 auf der Entscheidungs-Achse links, ergibt das Verteilungspaar eine Kombination aus Sensitivität (hellgrau) und Spezifität (dunkelgrau), welche rechts im Sensitivität-Spezifität Graphen als t_3 aufgetragen wird. (Gleiche Wirkung bei t_1 , t_2 und t_4) ⁽⁸⁾

2.1.1.2. Das Basiskonzept der ROC-Analyse

Das Problem, welches aufkommt, wenn ein Test höhere Sensitivität, aber geringere Spezifität hat als ein anderer, kann gelöst werden, wenn man erkennt, dass die Sensitivität und die Spezifität praktisch aller diagnostischer Tests verändert werden kann, indem man den Entscheidungsgrenzwert bzw. die Abnormitätsschwelle modifiziert. Wenn man als Beispiel zwei Radiologen mit denselben Fähigkeiten bezüglich der Beurteilung von Mammogrammen annimmt, von denen allerdings einer um vieles strenger bei der Indikationsstellung zur Biopsie ist, so wird dieser strengere natürlich mehr Malignitäten entdecken, gleichzeitig wird er aber auch mehr negative Biopsien veranlassen. Diese umgekehrte Proportionalität ist in Abb. 1 dargestellt. Die zwei Kurven in der linken Darstellung repräsentieren die Verteilung der Überzeugung des Radiologen, dass die Bildgebung das Vorhandensein einer Malignität bei tatsächlich negativen und tatsächlich positiven Patienten anzeigt. Wenn dieser Radiologe die Biopsie nur bei Patienten anordnet, bei denen seine Überzeugung einer positiven Interpretation den Entscheidungsgrenzwert, welcher bei t_3 markiert ist, übersteigt, dann wäre seine Sensitivität der Teil der Fläche der Verteilung der

tatsächlich positiven, der rechts von t_3 liegt, während seine Spezifität der Teil der Fläche der Verteilung der tatsächlich negativen wäre, der links von t_3 liegt. Diese Kombination kann dann als Punkt t_3 im Sensitivität-Spezifität Graphen aufgetragen werden, welcher auf der horizontalen Achse die Spezifität und auf der vertikalen Achse die Sensitivität zeigt. Entsprechend können auch die Entscheidungsgrenzwerte t_1 , t_2 und t_4 so aufgetragen werden.

Wenn man eine viel größere Anzahl möglicher Platzierungen der Entscheidungsgrenzwerte annimmt, kann man viel mehr Punkte im Sensitivität-Spezifität Graphen auftragen, wodurch eine kontinuierliche Kurve, wie in links in Abb. 2 gezeigt, welche sämtliche mögliche Kombinationen aus Sensitivität und Spezifität aufzeigt, entsteht. Aus historischen Gründen wird bei der hier dargestellten Beziehung die Sensitivität meist als $1 - \text{Spezifität}$ bezeichnet, wodurch eine ROC-Kurve, wie rechts in Abb. 2 gezeigt, entsteht. Anfangs verwirrend erscheinen mag, dass die Punkte, also die Kombination falsch-positiver und richtig-positiver Ergebnisse, am linken unteren Ende der ROC-Kurve dadurch entstehen, dass der Entscheidungsgrenzwert am rechten Ende der Entscheidungsachse der linken Seite in Abb. 1 liegen, also wenige falsch-positive, aber auch relativ wenige richtig-positive Ergebnisse produzieren, wogegen die Punkte am rechten oberen Ende der ROC-Kurve auf der Entscheidungsachse links liegen und somit viele richtig positive, aber auch mehrere falsch-positive Ergebnisse produzieren. Aus der näheren Betrachtung lässt sich schließen, dass ROC-Kurven, welche insgesamt höher oder niedriger liegen, als jene in Abb. 2, auf Indexpaaren beruhen, deren Verteilungen mehr oder weniger überlappend dargestellt werden, als jene in Abb. 1. Es ist wichtig, zu erkennen, dass die Überlappung der Verteilungen und somit auch die Höhe der ROC-Kurven nicht nur von den Fähigkeiten des befundenden Arztes, sondern auch von den technischen Aspekten der Bildgebung und der Charakteristik der Studienpopulation wie dem Gleichgewicht früher und später Krankheitsstadien der tatsächlich positiven Patienten abhängig ist.

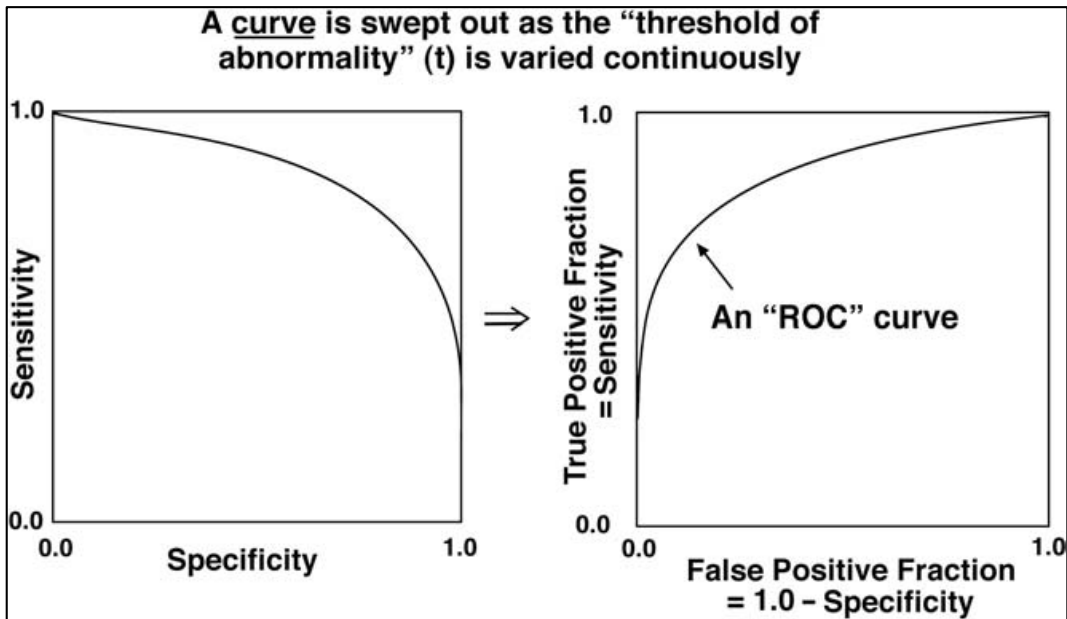


Abbildung 4: Die linke Seite zeigt die kontinuierliche Beziehung zwischen Sensitivität und Spezifität, abgeleitet aus Abb. 3 bei kontinuierlicher Variation des Entscheidungsgrenzwertes. Die rechte Seite, welche eine typische ROC-Kurve repräsentiert, ist ein Spiegelbild, da die Sensitivität gegen $1 - \text{Spezifität}$ aufgetragen wird. ⁽⁸⁾

2.1.1.3. Welche Modalität ist besser?

A dilemma: Which modality is better?

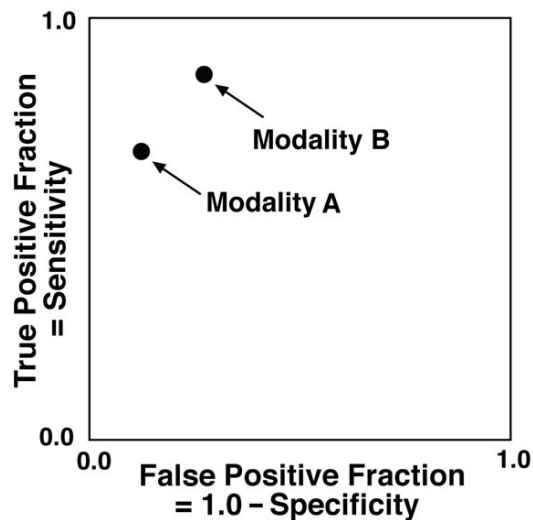


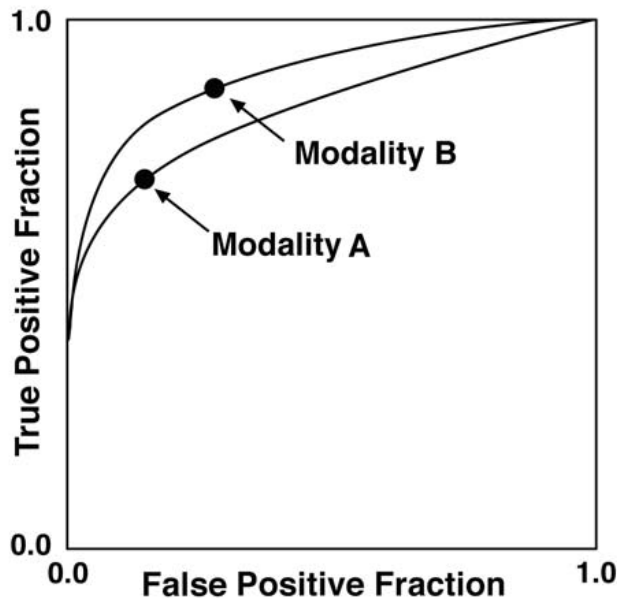
Abbildung 5: Das Problem, die diagnostische Genauigkeit zweier Systeme zu vergleichen, wenn nur die Sensitivität oder die Spezifität bekannt ist. ⁽⁸⁾

Bedenkt man dieses Konzept, kann man das Problem, welches in der Abb. 3 aufgezeigt ist, lösen, nämlich den Fall, dass Test B zwar eine höhere Richtig-positive-Fraktion, aber auch eine höhere Falsch-negative-Fraktion aufweist als Test A.

Vorausgesetzt, dass die ROC-Kurven für Test A und Test B bestimmt wurden und so aussehen, wie in Abb. 4 gezeigt, kann man daraus schließen, dass der Test B eine bessere diagnostische Genauigkeit aufweist, da er bei gleich vielen falsch-positiven Ergebnissen mehr richtig-positive Ergebnisse liefern kann, als Test A., wenn der Entscheidungsgrenzwert für Test B so lange in Richtung des rechten Endes der Entscheidungsachse in Abb. 1 verschoben wird, bis die Falsch-positive-Fraktion von Test B der von Test A entspricht oder wenn der Entscheidungsgrenzwert für Test A solange nach links verschoben wird, bis die Falsch-positive-Fraktion von Test A wiederum der von Test B entspricht. Betrachtet man die ROC-Kurven in Abb.4, so kann Test B auch eine niedrigere Falsch-positive-Fraktion als Test A bei gleicher Richtig-positive-Fraktion erreichen, wenn der Entscheidungsgrenzwert für Test B in Richtung des rechten Endes der Entscheidungsachse in Abb. 1 verschoben wird, bis die Richtig-positive-Fraktion von Test B der von Test A entspricht oder wenn der Entscheidungsgrenzwert für Test A umgekehrt solange in Richtung des linken Endes der Entscheidungsachse in Abb. 1 verschoben wird. Allerdings ist dieser Schluss nicht klar, wenn man nur die beiden Punkte in Abb. 3 zur Verfügung hat, da die Bestimmung der jeweiligen ROC-Kurven auch ein Bild wie in Abb. 5 hervorbringen könnte, welches aufzeigt, dass beide Tests dieselbe diagnostische Genauigkeit liefern und die Punkte übereinstimmen, wenn man nur die Entscheidungsgrenzwerte modifiziert. Außerdem könnte in einem dritten Szenario der Test A überlegen sein, wenn die zugehörige ROC-Kurve an jedem Punkt höher ist, als die von Test B, was passieren kann, wenn die ROC-Kurve A von Beginn an steiler ansteigt, als die ROC-Kurve B. Höhere Kurven zeigen bessere diagnostische Genauigkeit an, und so kann ein diagnostischer Test eindeutig als besser gewertet werden als der andere, wenn nur die ROC-Kurve höher ist.

Der Vergleich zweier diagnostischer Tests ist komplizierter, wenn die ROC-Kurven einander kreuzen, was als nächstes beschrieben wird.

The dilemma is resolved after ROC curves are determined (one possible scenario)



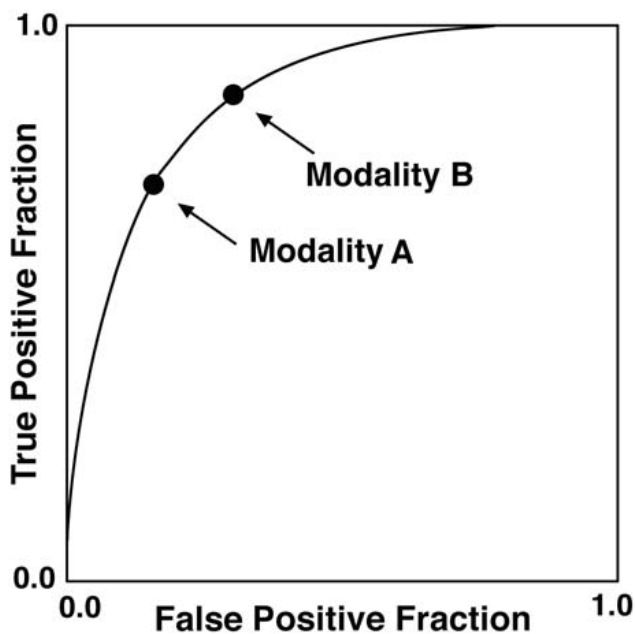
Conclusion:

Modality B is better, because it can achieve:

- higher TPF at same FPF, or
- lower FPF at same TPF

Abbildung 6: Das Problem kann gelöst werden, wenn die ROC-Kurven beider Tests bestimmt werden. ⁽⁸⁾

The dilemma is resolved after ROC curves are determined (another possible scenario)



Conclusion:

Modalities are equivalent, because same combinations of TPF and FPF are available

Abbildung 7: Das Problem kann gelöst werden, wenn die ROC-Kurven beider Tests bestimmt werden. ⁽⁸⁾

2.1.1.4. ROC-bezogene Hinweise diagnostischer Genauigkeit

Die Erfahrung zeigt, dass reale ROC-Kurven durch zumindest zwei Parameter bestimmt werden müssen, die separat die Höhe der Kurve und die Stärke oder Schwäche der Symmetrie rund um die negative Diagonale (-45°) repräsentieren. Somit können ROC-Kurven nicht vollständig durch eine einzige Nummer zusammengefasst werden und Systemvergleiche, die zur Gänze auf einem Einzelzahl-Index beruhen können zu Fehlerhaften Schlussfolgerungen führen. Das Konzept von „besser“ und „schlechter“ ist jedoch gänzlich eindimensional und so setzt der Vergleich von Systemen zum Ranking Einzelzahl-Indices voraus. Die Hauptanforderung in einem solchen Fall ist die Wahl eines Index, der so verlässlich und bedeutsam wie möglich im praktischen Sinne ist. Einige solcher möglichen Indices werden im Folgenden beschrieben.

Wenn zwei ROC-Kurven in einem Vergleich sich nicht überkreuzen, wie in Abb. 4, dann kann die diagnostische Genauigkeit dieser Kurven eindeutig gereiht werden mithilfe der Fläche unter der Kurve im Quadrat, welche als Flächenindex oder als AUC, Area under the Curve, bezeichnet wird. Die AUC kann, wenn die Kurve nach dem Standard binomischen Modell diskutiert wird, auf drei Arten interpretiert werden: als Sensitivität, gemittelt über alle möglichen Spezifitäten (durchschnittliche Kurvenhöhe); als Spezifität, gemittelt über alle möglichen Sensitivitäten (wie weit links die Kurve ist); oder aber als die Wahrscheinlichkeit, alle tatsächlich positiven Fälle in einer „zwei-Alternativen-Entscheidung“ herauszufinden, wobei jeder Versuch einen Beobachter braucht, welcher zwischen zufällig ausgewählten tatsächlich positiven Fällen und zufällig ausgewählten tatsächlich negativen Fällen unterscheidet. Werte der AUC gehen hier von 0.5 für die niedrigste mögliche ROC-Kurve, der positiven Diagonale, auf welcher die Wahrscheinlichkeit einer positiven Auswertung für jeden Fall gleich hoch ist, unabhängig vom tatsächlichen Ergebnis, bis 1.0 für die höchste ROC Kurve, welche am linken und oberen Rand des Quadrats entlangverläuft.

Die übliche AUC fasst die gesamte ROC-Kurve zusammen. Dieser Index kann jedoch missverständlich sein, wenn sich zwei ROC-Kurven kreuzen, da sie dann beide dieselbe AUC aufweisen können, obwohl eine der Kurven in einem bestimmten Bereich der Falsch-positive-Fraktion höher ist, oder, was gefährlicher ist, kann die AUC der Kurve niedriger sein, deren Verlauf im klinisch interessanten Bereich der Falsch-

positive-Fraktion höher ist. Abgezielt auf diese Einschränkung wurden zwei Indices eingeführt, welche Teil-Flächen aufzeigen: einer repräsentiert die AUC innerhalb eines ausgewählten Bereiches der Falsch-positive-Fraktion,⁽²⁷⁾ während der andere die Fläche rechts der ROC Kurve, innerhalb eines ausgewählten Bereiches der Richtig-positive-Fraktion ⁽²⁸⁾.

Der Wert jeder dieser Teilindices hängt ab vom gewählten Bereich der Falsch-positive-Fraktion oder Richtig-positive-Fraktion. Auf diesen muss explizit hingewiesen werden, wenn der Indexwert aufgezeigt wird, und er sollte nach klinischen Überlegungen gewählt werden.

Teilflächenindices basieren auf der Falsch-positive-Fraktion oder Richtig-positive-Fraktion um auf klinisch relevante Sensitivitäten und Spezifitäten abzielen, ohne aber das Gleichgewicht zwischen Sensitivität und Spezifität zu beurteilen, welches in der bestimmten klinischen Situation von Bedeutung ist. Andere ROC-basierte Indices für die diagnostische Genauigkeit können verwendet werden, um eine nähere Vergleichbarkeit zu erzielen. Der einfachste dieser lokalen Indices ist die Richtig-positive-Fraktion bei einer festgelegten Falsch-positive-Fraktion und eine Falsch-positive-Fraktion bei einer festgelegten Richtig-positive-Fraktion, welche Unterschiede von ROC-Kurven bezüglich einzelner Abweichungen der Falsch-positive-Fraktion oder Richtig-positive-Fraktion anzeigen. Trotz der rechnerischen Vorteile und der einfachen Verfügbarkeit sind diese Indices nicht weit verbreitet, da sie bedeutende praktische und theoretische Grenzen haben: Ärzte können sich üblicherweise nicht auf einen einzelnen Wert der Falsch-positive-Fraktion oder der Richtig-positive-Fraktion einigen, anhand welcher zwei diagnostische Tests verglichen werden sollen, und die Entscheidungstheorie impliziert, dass ROC-Kurven auch nicht anhand eines einzelnen Falsch-positive-Fraktion oder Richtig-positive-Fraktion Wertes verglichen werden sollen, sondern stattdessen eher anhand eines Wahrscheinlichkeitsverhältnisses, welches mit einem bestimmten Wert der Krümmung der ROC-Kurve korreliert. Diese Limitation kann übergangen werden, indem man die relativen Vorteile und Kosten von richtig-positiven, falsch-positiven, richtig-negativen und falsch-negativen Entscheidungen abwägt, wenn man erkennt, dass das Gleichgewicht zwischen diesen vier Arten von Entscheidungen sich verändert, wenn der Punkt eines Tests entlang der ROC-Kurve verschoben wird und wenn man dann die ROC-Kurven im Sinne des

maximalen „erwarteten Netto Nutzens“ vergleicht, welcher auf jeder Kurve erreicht werden kann^(24,29).

Jedoch wird dieser in der Theorie attraktive Index aufgrund der Schwierigkeit des Konsenses bezüglich der relativen Nutzen und Kosten der jeweiligen korrekten und inkorrekten Entscheidungen kaum verwendet.

2.1.1.5. Generalisierte ROC-Analyse

Die konventionelle ROC-Analyse beschreibt vollständig alle Kosten-Nutzen-Abwägungen, die ein menschlicher oder automatisierter Entscheider aufgrund der Häufigkeit von richtig-positiven, falsch-positiven, richtig-negativen und falsch-negativen Entscheidungen in jeder Situation, in der es zwei mögliche Entscheidungen gibt, bei einer definierten Patientengruppe erreichen kann. Die Bezeichnungen positiv und negativ werden verwendet, um jeden Zustand zu beschreiben (krank/gesund, Präsenz/Absenz einer Krankheit). In der Praxis können diese beiden Zustände je nach Wunsch definiert werden, solange sich die tatsächliche Zugehörigkeit zu einer der beiden Gruppen mittels eines Goldstandards (z.B. Biopsie) nachweisen lässt, um eine Basis zur Bewertung der Entscheidungsperformance sicherzustellen. Bei der medizinischen Bildgebung sind die beiden Gruppen meist insofern zusammengefasst, als dass jede von ihnen verschiedene Zustände beinhaltet. Bei einer Aufgabe, die das Erkennen der Krankheit X beinhaltet, können Patienten mit der Krankheit X sowohl frühe als auch späte Stadien davon aufweisen, während aber jene Patienten, die nicht an Krankheit X leiden, trotzdem eine Reihe anderer Krankheiten mit ähnlichen Symptomen wie denen der Krankheit X aufweisen. ROC-Kurven, welche aus solchen Situationen hervorgehen, sind abhängig vom Spektrum der Zustände, in denen sich die Patienten einer jeden Gruppe befinden können. Das vielleicht extremste Beispiel zusammengefasster Gruppen in klinische Zwei-Gruppen Klassifikationsaufgaben ist jenes, zwischen normalen und abnormalen Patienten zu unterscheiden. Es ist zwar prinzipiell gültig, aber üblicherweise nicht durchführbar, da viele Auswahlaspekte involviert sind. Das Ausmaß, in welchem eine Zwei-Gruppen Auswertung wirklich relevant ist, hängt zum Teil davon ab, ob die Aufgabe eine Routinearbeit darstellt, zum Teil aber auch davon, welche Rolle die Bildgebung im weiteren Verlauf der Patienten spielen soll, also ob der Radiologe in dem Fall entscheiden soll, welche Patienten zu

Biopsie geschickt werden, oder ob er nur über die Präsenz oder Absenz einer Malignität entscheiden soll. Obwohl diagnostische Entscheidungen, vor allem bei der Entdeckung von Krankheiten, oft als Sequenzen von Zwei-Gruppen Klassifikationsaufgaben dargestellt werden können und somit durch konventionelle ROC-Analysen dargestellt werden können, ist eine Aufgabe, nämlich das Ermitteln der Lokalisation der Malignität neben ihrer bloßen Entdeckung, nicht darstellbar. So kann die konventionelle ROC-Analyse nicht dazu verwendet werden, die Fähigkeit eines diagnostischen Systems, eine Operation auf eine bestimmte Seite, auf welcher der Tumor vorhanden ist, zu lenken, oder aber zwischen Absenz jeglicher Tumore, Präsenz eines bestimmten Tumors und der Präsenz mehrerer Tumore zu unterscheiden, direkt zu quantifizieren. Außerdem kann damit nicht die Klassifikationsleistung in Aufgaben, bei denen Patienten in drei oder mehrere Gruppen gleichzeitig (Entscheidung zwischen: keine Raumforderung, benigne Raumforderung, maligne Raumforderung), anstatt nacheinander (Entscheidung zwischen: keine Raumforderung, Raumforderung → dann Entscheidung zwischen: benigne Raumforderung, maligne Raumforderung) eingeteilt werden müssen, ermittelt werden. Einige generalisierte Formen der ROC-Analyse, welche vorgeschlagen wurden um diesen Anforderungen zu entsprechen, werden im Folgenden kurz vorgestellt, um ihre Unterschiede aufzuzeigen. Obwohl diese Techniken eine Reihe heikler methodischer Probleme aufwerfen, von denen einige noch nicht gelöst werden konnten, verlangt deren zunehmende Verwendung in Evaluierungsstudien Verständnis sowohl ihrer Vorteile als auch ihrer Limitationen.

Lokalisations-ROC-Analysen (LROC) beziehen sich auf Detektionsaufgaben, in welchen jedes Bild entweder eine oder keine Läsion beinhaltet und der Radiologe die genaue Lokalisation spezifizieren soll⁽³⁰⁾.

Die Datensammlung erfolgt in sämtlichen Fällen in Bezug auf bestimmte Konfidenzraten, die die Wahrscheinlichkeit des Vorhandenseins einer Läsion ausdrücken, wie bei konventionellen ROC-Analysen, mit einer zusätzlichen Lokalisationsangabe, welche entweder in einem Raster oder als geometrische Position (Koordinaten) ausgedrückt werden kann. Dann wird eine Kurve auf der Basis dieser Daten in den Konfidenzintervallen dargestellt. Dies geschieht in ähnlicher Weise wie eine konventionelle ROC Kurve aufgetragen wird, der einzige Unterschied besteht darin, dass dem Entscheider die Richtig-positiven-Entscheidungen nur dann

anerkannt werden, wenn auch die angegebene Position nahe der tatsächlichen Lokalisation in einem tatsächlich positiven Bild ist. Tatsächlich-negative-Bilder werden gleich bewertet, wie bei konventionellen ROC-Kurven, sodass die horizontale Achse der generalisierten Kurve jener der konventionellen entspricht (Abb. 6) und somit die Wahrscheinlichkeit, eine Läsion in einem tatsächlich negativen Bild zu entdecken, anzeigt. Die vertikale Achse der LROC jedoch, die aus strengeren Bewertungsregeln für tatsächlich positive Bilder resultiert, repräsentiert die Wahrscheinlichkeit, eine Läsion nicht nur korrekt zu detektieren, sondern darüber hinaus auch korrekt zu lokalisieren. Folglich liegt jeder Punkt auf der LROC-Kurve für Falsch-positive-Fraktion > 0 zumindest leicht niedriger als der Punkt für die selbe Falsch-positive-Fraktion der korrespondierenden ROC-Kurve, die aus derselben Datenmenge errechnet werden kann, indem man die Lokalisationsangabe des Entscheiders ignoriert. Aus Abb. 6 und Abb. 7 kann man erkennen, dass die LROC wie die konventionelle ROC-Kurve in einem Einheitsquadrat aufgetragen wird, aber in der Regel nicht in die rechte obere Ecke des Quadrats läuft, da die Entscheider, welche durch LROC-Analysen bewertet werden, das Äquivalent von Richtig-positive-Fraktion = 1 nicht lediglich dadurch erreichen können, alle Bilder als positiv zu bezeichnen. Festzuhalten ist, dass das Ergebnis einer jeden LROC-Studie davon abhängt, wie hoch die vom Datenanalysten erlaubte Rate an Lokalisationsfehlern ist. Diese Toleranz bestimmt nämlich, ob die Einstufung oberhalb der Entscheidungsgrenzwertes eines tatsächlich positiven Bildes als richtig-positiv oder falsch-negativ gewertet wird. Die unkomplizierte theoretische Beziehung zwischen einer LROC-Kurve und ihrer korrespondierenden konventionellen ROC-Kurve für jede festgelegte Lokalisationsfehler-Toleranz zeigt an, dass Vergleiche, basierend auf LROC und konventioneller ROC-Analyse, Bildgebungssysteme auf die gleiche Art einstufen werden. Gleichzeitig ist aber die Versicherung gegeben, dass konventionelle ROC-Analysen keine irreführenden Ergebnisse in Vergleichsstudien produzieren, obwohl sie die Lokalisationsgenauigkeit nicht bewerten. Andere Überlegungen jedoch deuten darauf hin, dass die Lokalisationsdaten, welche in LROC-Studien gesammelt werden, zu einer größeren statistischen Aussagekraft führen.

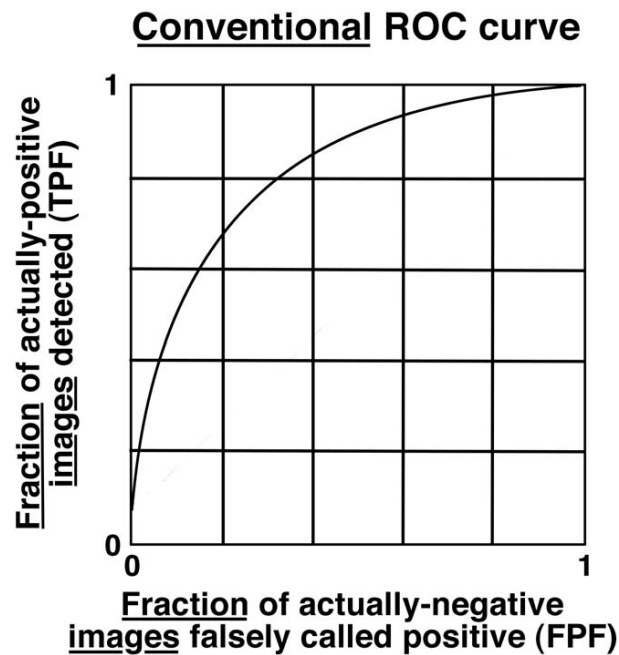


Abbildung 8: Eine konventionelle ROC-Kurve. Die Aufgabe des Entscheiders ist es, zwischen tatsächlich positiven Fällen und tatsächlich negativen Fällen zu unterscheiden. Es wird keine Lokalisation der Läsion verlangt⁽³¹⁾.

ROC-Analysen mit freien Antworten (FROC)^(32,33) können als Verallgemeinerung der LROC-Methodik verstanden werden, die die Möglichkeit von mehr als nur einer Läsion pro Bild unterbringen, indem sie multiple Berichte pro Bild erlauben. Die Datensammlung erfolgt in Bezug auf Konfidenzintervalle betreffend die mögliche Präsenz einer Läsion in jeder Lokalisation, die der Entscheider für erwähnenswert befindet. Indem man die vielen verschiedenen möglichen Positionen des Entscheidungsgrenzwertes, welcher positive von negativen Berichten unterscheidet, bedenkt, kann man aus den Konfidenzraten und Lokalisationsdaten eine Kurve auftragen, die die Beziehung zwischen der Fraktion der detektierten Läsionen und der Anzahl der falsch-positiven Berichte pro Bild aufzeigt. Die horizontale Achse der FROC Kurve ist nicht auf eine Fraktion mit einem maximalen Wert von 1.0 normiert, stattdessen muss sie sich für eine willkürlich große Anzahl von falsch positiven Berichten pro Bild ausdehnen, da die Anzahl möglicher Lokalisationen für falsch positive Berichte unbekannt ist. Wie auch bei LROC Studien können die Lokalisationsdaten, welche in einer FROC Studie gesammelt werden, zu einer größeren statistischen Aussagekraft als die ROC-Analyse führen, jedoch hängt das

Studienergebnis wieder von der vom Datenanalysten erlaubten Menge von Lokalisationsfehlern ab.

Im Bereich der Entwicklung generalisierter ROC-Methoden, die die Entscheidungsleistung in Klassifikationsaufgaben mit mehr als zwei möglichen Gruppen quantifizieren sollen wurden weniger Fortschritte gemacht.

2.1.1.6. Kurvenanpassung

Die Kurvenanpassung beinhaltet immer drei Schritte: 1. Auswählen einer Gruppe von Anpassungsfunktionen mit einstellbaren Parametern, welche die Daten ausreichend genau wiedergeben kann. 2. Aussuchen eines Maßstabes, welcher die jeweilige Anpassungsgüte quantifizieren kann. 3. Berechnung der Werte der Parameter der Anpassungsfunktion welche die beste Anpassung nach dem ausgewählten Maßstab aufweisen.

Konventionelle ROC-Kurven werden üblicherweise angepasst, indem man das Binome Modell anwendet, welches voraussetzt, dass die wahre Form der Kurve eine jener ist, welche aufgrund eines Paares normalverteilter, entscheidungsvariabler Verteilungen entstehen können ^(39,40,41).

Die Erfahrung zeigt, dass das Standard Binome Modell exzellente Anpassung für statistisch verlässliche ROC-Kurven in einem sehr breit gefächerten Bereich möglicher Anwendungen aufweist. Dieses Modell kann jedoch Kurven mit unpassender Form hervorbringen, wenn es auf Basis einer schwachen Datensammlung angewandt wird (z.B. geringe Fallzahl) ⁽⁴²⁾.

Deshalb wurden einige Modelle zur Anpassung konventioneller ROC-Kurven eingeführt. Aus diesen scheint das korrekte Binome Modell das verlässlichste zu sein, außerdem hat es den Vorteil, dass die Anpassungen weitgehend ident zu denen des Standard Binomen Modells sind, wenn dieses eine Anpassung mit passender Form hervorbringt ^(31,43,44,45).

2.1.1.7. Testung der statistischen Signifikanz von Abweichungen

Abweichungen zwischen geschätzten ROC-Kurven müssen statistischen Signifikanztestungen unterzogen werden um herauszufinden, ob diese Abweichungen zufälligen Schwankungen zugeschrieben werden können oder ob sie tatsächliche Abweichungen darstellen. Eine große Anzahl statistischer Teste wurde entwickelt, um die Signifikanz von Abweichungen zu beurteilen. Im Folgenden werden zwei fundamentale Fragestellungen kurz besprochen, um eine Basis für die Auswahl eines passenden statistischen Tests zu schaffen, wenn es dazu kommt, die Gültigkeit einer ROC-Studie zu bewerten.

Eine Unterscheidung zwischen momentan verfügbaren statistischen Tests ist die Art, wie „Unterschied“ quantifiziert wird: in Bezug auf die gesamte ROC-Kurve oder in Bezug auf einen bestimmten globalen, regionalen oder lokalen Index. Die statistische Praxis stellt die Anforderung, dass eine klinisch relevante Definition von Unterschied gewählt werden muss, bevor jegliche Signifikanzberechnung durchgeführt wird. Die berichtete Signifikanz wird nämlich fälschlicherweise höher oder niedriger, wenn man zuerst die Ergebnisse verschiedener statistischer Tests berechnet und erst dann das am meisten oder am wenigsten signifikante Ergebnis wählt.

Die andere wichtige Unterscheidung zwischen statistischen Tests für Abweichungen zwischen ROC-Kurven betrifft die Ursprünge statistischer Schwankungen, welche der Test berücksichtigt. Ein „p-Wert“ sollte die Wahrscheinlichkeit repräsentieren, einen Unterschied zu finden, der größer ist, als der im Experiment gezeigte Unterschied, wenn aber in Wirklichkeit kein Unterschied in der Population, von welcher die Daten gewonnen wurden, besteht. Diese Wahrscheinlichkeit kann als eine Proportion von Ergebnissen, welche durch mehrmalige Wiederholung des Experiments unter denselben Bedingungen überprüft werden, interpretiert werden. So muss sich die Frage stellen, unter welchen Bedingungen das Experiment wiederholt werden muss. Beispielsweise bei der Wiederholung eines ROC-Experimentes mit unabhängig ausgewählten Radiologen und denselben Fällen, mit unabhängig ausgewählten Radiologen und unabhängig ausgewählten Fällen, mit denselben Radiologen und denselben Fällen oder mit denselben Radiologen und unabhängig ausgewählten Fällen. Die Variationen der Ergebnisse eines mehrmalig wiederholten ROC-Experiments in jedem dieser Szenarien können stark unterschiedlich sein: die meisten

Unterschiede können erwartet werden, wenn sowohl Radiologen als auch Fälle bei jeder Wiederholung zufällig ausgewählt werden, die wenigsten dann, wenn die Radiologen und die Fälle bei jeder Wiederholung konstant bleiben ⁽⁴⁶⁾.

Somit muss ein statistischer Test zur Überprüfung der Abweichung gewählt werden, welcher die Ursache der Unterschiede miteinbezieht, um zu einem sinnvollen Schluss zu gelangen. So beziehen zum Beispiel einige statistische Tests nur die Variation des Auslesers mit ein, während andere nur die Variation der Fallbeispiele berücksichtigen. Mit beiden kann man jeweils nur einen Teil der in dem Experiment vorkommenden Szenarien auswerten, womit der Schluss naheliegt, dass beide Methoden berücksichtigt werden müssen, wenn man die statistische Signifikanz einer ROC-Studie überprüfen möchte, da wissenschaftlicher Fortschritt von der Möglichkeit abhängt, Experimente in verschiedenen Labors zu replizieren, was automatisch zu einer Variation sowohl der Fälle als auch der Ausleser führt. Dies ist der Bereich der Multi-case, Multi-reader (MRMC) ROC-Analyse, welche während der letzten Jahre stetig weiterentwickelt wurde und heute die Methode der Wahl für die meisten ROC-Vergleichsstudien darstellt ⁽⁴⁷⁾.

Was man immer beachten sollte, ist, dass ROC-Vergleichsstudien manchmal statistisch signifikante Unterschiede aus zwei von Grund auf verschiedenen Gründen nicht aufzeigen können: zum Ersten, weil es tatsächlich nur wenig oder keinen Unterschied zweier Systeme in der gewählten Population gab oder zum Zweiten, weil die im Experiment verwendete Fallzahl zu gering war, um eine adäquate statistische Aussagekraft aufzuweisen. Somit kann das Versagen, einen statistisch signifikanten Unterschied zu demonstrieren nicht gleichgesetzt werden mit dem Beweis, dass kein Unterschied existiert. Stattdessen muss zusätzlich ein Maßstab angegeben werden, in welchem Bereich der tatsächliche Unterschied liegen sollte, unter Einbeziehung der Daten des Experiments. Konfidenzintervalle sind besonders hilfreich, wenn es darum geht, p-Werte in Situationen, in denen kein statistisch signifikanter Unterschied gefunden werden konnte, zu interpretieren⁽⁴⁸⁾.

2.1.2. VLAD und Funnel Plot

Eine der Methoden, um eine Anpassung an das individuelle Risiko einzelner Patienten und Patientinnen vorzunehmen, ist das Variable-Live-Adjusted Display (VLAD). Hier wird für jeden Patienten und jede Patientin das erwartete Risiko einer speziellen Outcome-Variable (z.B. das Versterben des Patienten oder der Patientin als in zeitlicher Abfolge zu definierende Mortalität; eine ebenfalls zu definierende eingetretene Komplikation als Morbidität), angepasst an Alter, Geschlecht sowie Komorbiditäten, berechnet. Dann wird das erwartete Risiko vom eintretenden Outcome (für Präsenz, 0 für Abwesenheit) abgezogen und das Ergebnis wird in einem Diagramm eingetragen. Eine Aufwärtskurve bedeutet, dass die Nummer der Fälle (z.B. eingetretene Todesfälle) geringer war, als die erwartete, während eine Abwärtskurve vom Gegenteil zeugt. Im Folgenden werden an der Stelle, an der der Kurvenverlauf abflacht, Schwellenwerte festgesetzt (siehe nachfolgende Abbildung 9 in englischer Fachsprache):

VLAD with control limits & flags

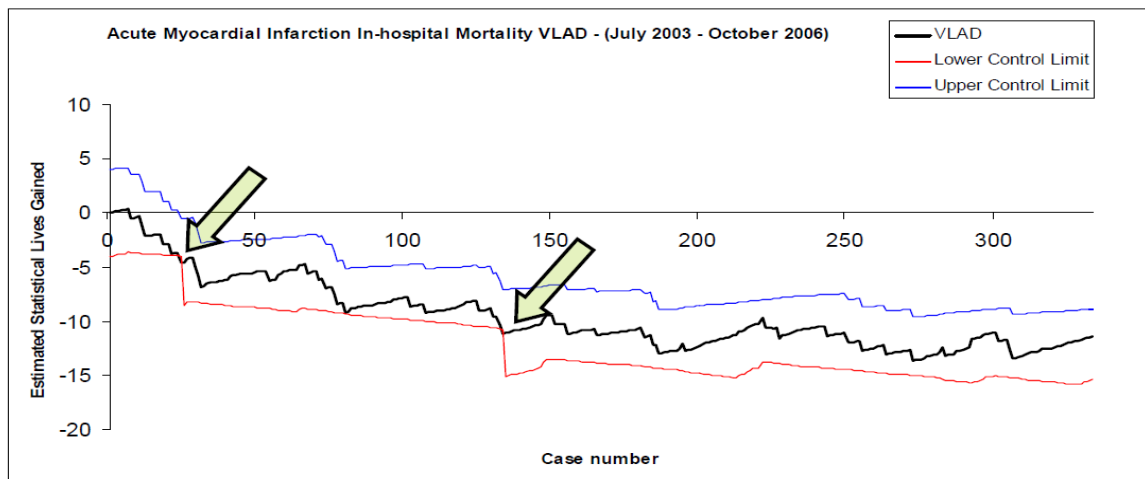


Abbildung 9: Festlegung der Schwellenwerte in einer Beispielkurve

Nun besteht Qualitätsverbesserung aber nicht darin, Daten zu sammeln, sondern eben darin, sicherzustellen, dass Abweichungen erkannt werden und zu Änderungen im Ablauf führen. Hierfür sind Grenzwerte nötig, welche Normalwerte von Variationen, welchen nachgegangen werden muss, trennen.

Das VLAD beruht nun also auf der „Anzahl geretteter Leben“. Wie schon gesagt, haben unterschiedlicher Herzoperationen auch verschiedene vorausgesagte Mortalitäten. Operateure und Operateurinnen oder Institutionen werden je nach Schwere und Ausgang des Eingriffes nach einem kumulativen Bonus-Malus-System bewertet: Verstirbt ein Patient oder eine Patientin, erhält der Operateur oder die Operateurin einen Malus in Höhe $1 - (\text{vorhergesagter Sterbewahrscheinlichkeit } [\%]/100)$. Das heißt, je schwieriger die Ausgangslage ist, und damit je höher der EuroSCORE berechnet ist, desto geringer ist der Malus. Stirbt aber ein Patient oder eine Patientin mit geringem Risiko, fällt der Malus entsprechend höher aus. Umgekehrt erhält der Operateur für jeden überlebenden Patienten oder Patientin einen Bonus in Höhe der vorhergesagter Sterbewahrscheinlichkeit $[\%]/100$. Somit ist der Bonus umso höher, je risikoreicher die Operation und je höher der EuroSCORE ist.

Über einen längeren Zeitraum hinweg sollte sich das Ergebnis bei gleich verteiltem Patientengut um die Nullachse einpendeln: Wenn ein Operateur oder eine Operateurin 100 Patienten und Patientinnen mit einer vorausgesagten Mortalität von 7% operiert, dann sollten 7 Patienten und Patientinnen versterben. Sind es weniger, steigt die Kurve, sind es mehr, sinkt sie entsprechend schnell.

Bei einem homogenen Patientenmix und einem aussagekräftigen Prognosemodell kann das VLAD eine gute Übersicht über die Ergebnisqualität geben. Allerdings wird in der Regel kein Operateur oder Operateurin und keine Institution ein so homogenes Patientengut aufweisen, wie es für die Interpretation des VLAD nötig ist. Eine Häufung besonders risikoreicher oder risikoarmer Patienten und Patientinnen über einen gewissen Zeitraum hinweg erlauben die Interpretation eines VLAD nur noch in eingeschränktem Maß. So lässt ein Abweichen der Kurve nach oben oder nach unten keine Aussage über die Ergebnisqualität mehr zu.

Nun braucht man, um die Aussage des VLAD in Abhängigkeit vom Patientenmix richtig einschätzen zu können eine Maßzahl der Güte. Hierzu wurde der Funnel Plot untersucht und eingeführt.

Abhängig von der Verwendung des jeweiligen EuroSCOREs wird die beobachtete, d.h. tatsächliche Mortalität in Bezug zur vorausgewiesenen, d.h. erwarteten Mortalität gesetzt, und daraus die logistische Performance Ratio berechnet. Diese sollte im Idealfall < 1 sein. (siehe nachfolgende Abbildung 10):

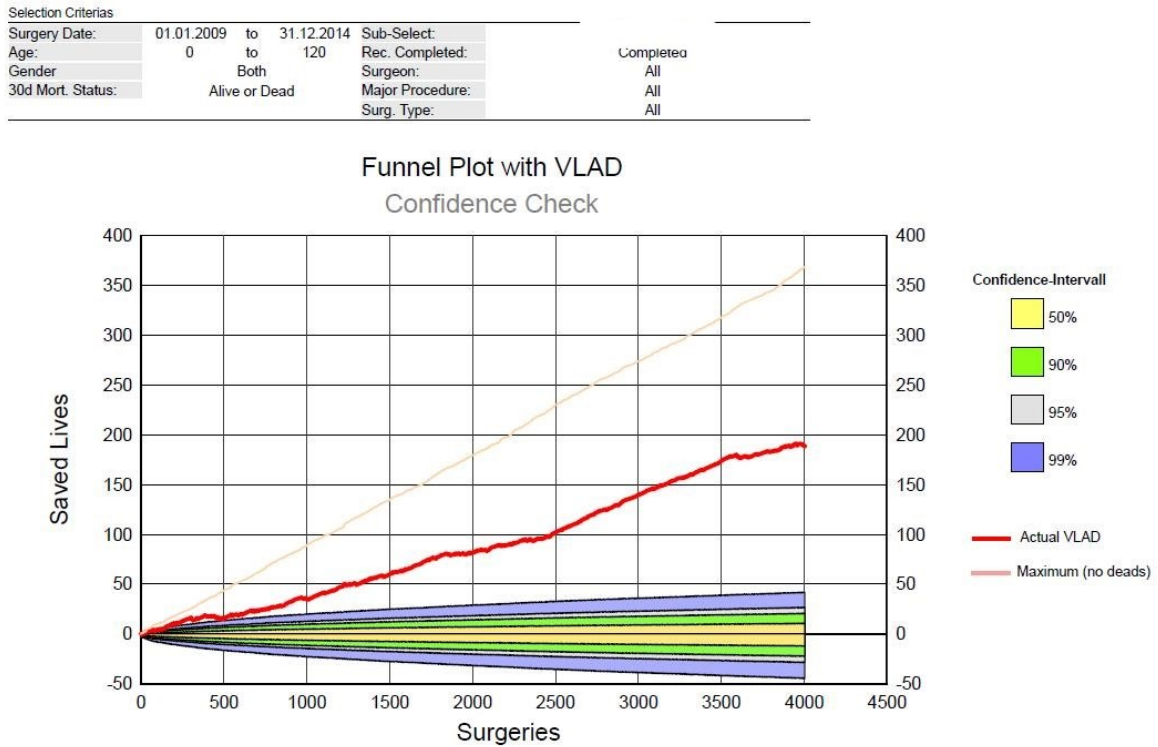


Abbildung 10: Funnel Plot und VLAD

Report shows LOGISTIC EuroSCORE

Selection Criterias 05.05.2015 Seite 1 von 1

Surgery Date:	01.01.2009	to	31.12.2014	Sub-Select:	
Age:	0	to	120	Rec. Completed:	All
Gender			Both	Surgeon:	All
30d Mort. Status:			All	Major Procedure:	All
				Surg. Type:	All

Overall Statistics

	Total	male	female
Total number of surgeries	4.850	3.239	1.611
30d Status = Censored			0
30d Status = Unknown	87	61	26
30d Status = Alive	4.503	3.021	1.482
30d Status = Dead	260	157	103
Total known 30d Status	4.763	3.178	1.585
Deads	260	157	103
Observed Mortality (%)	5,46	4,94	6,50
Average Pred. Mort.	9,48	8,16	12,13
Log. Performance Ratio	0,58	0,61	0,54
Average age (years)	68,11	66,72	70,92
Standard-Deviation	11,51	11,30	11,43

Abbildung 11: Statistik im Funnel Plot

2.1.3. Youden-Index

Beim Youden-Index J handelt es sich um einen Wert, mithilfe dessen man die maximale potentielle Effektivität eines diagnostischen Tests ermitteln kann. An jedem Punkt einer ROC-Kurve kann man den Wert der Sensitivität und den Wert 1-Spezifität ablesen. Verläuft die ROC-Kurve allerdings nicht, wie sie es bei einem idealen Test tun würde, entlang der linken und oberen Begrenzung, so kann man den Punkt, an dem der Test die maximale Effektivität aufweist nicht einfach ablesen und es bedarf weiterer Berechnungen. Der Youden-Index ergibt sich an jedem beliebigen Punkt einer ROC-Kurve durch die Formel $J = \text{Sensitivität} + \text{Spezifität} - 1$. Der Punkt, an dem J den maximalen Wert der jeweiligen ROC-Kurve erreicht, wird Cut-off-Wert genannt. An diesem Punkt hat der diagnostische Test seine maximale Effizienz, die optimale Kombination aus Sensitivität und Spezifität.

Graphisch betrachtet handelt es sich um jenen Punkt der Kurve, welcher die größte vertikale Distanz zur Diagonalen aufweist⁽⁹⁵⁾.

2.1.4. Datenreview, verwendete statistische Methoden, Zulassung

Die Daten wurden, nachdem die jeweiligen Operateure sie eingetragen hatten, aus dem Cardiac bzw. dem QS II des Universitätsklinikum Graz entnommen. Die Qualität derselben wird hausintern standardmäßig überprüft, bevor sie durch Herrn Dr. Roland Schistek in Salzburg, dem Beauftragten der Österreichischen Gesellschaft für Herz-Thoraxchirurgie, und in weiterer Folge durch die Arbeitsgruppe des ÖBIG/GÖG in einem externen Audit vidiert, überprüft, verifiziert und schließlich zur Verwendung freigegeben werden.

Wie bereits beschrieben, wurde in der vorliegenden Studie die Performance des logistischen EuroSCORE und des EuroSCORE II anhand von Diskriminationsstärke und Kalibrierung beschrieben. Dann wurde die risikobereinigte Mortalität errechnet, indem man die untersuchte durch die berechnete Mortalität dividiert. Die Genauigkeit der Risikostratifizierungsprogramme wurde durch die Erstellung von ROC-Analysen veranschaulicht und durch die Berechnung der AUC und deren 95% Konfidenzintervall angegeben. Die Diskriminationskraft wurde anhand des Youden-Index berechnet,

dadurch konnte der Cut-off Wert sowie die Sensitivität und Spezifität der Scores errechnet werden.

Bei den Ergebnissen wurden quantitative Variablen anhand von Mittelwerten +/- Standardabweichung sowie Median, Minimum und Maximum präsentiert, während die kategorialen Variablen mittels Prozentwerten und absoluten Häufigkeiten angegeben wurden.

Sämtliche Berechnungen wurden mit dem Statistikprogramm SPSS durchgeführt, der Youden-Index wurde zusätzlich in Excel berechnet.

Die Studie wurde von der Ethikkommission der Medizinischen Universität Graz in allen Punkten geprüft und mit der *EK-Nummer 28-061 ex 15/16* unter dem Studientitel *„Cardiosurgical Risk Assessment: How does EuroScore II alter clinical results compared to EuroSCORE representing an appropriate Method for quality management in cardiac surgery – a retrospective study“* approbiert (siehe Ende: Appendix 1 – Ethikvotum).

3. Ergebnisse

3.1. Demographische Daten

Das gesamte zur Studie herangezogene Patientenkollektiv bestand aus n=1542 Patienten, welche einer koronararteriellen Bypassoperation unterzogen wurden, mit einem Patientenalter zwischen 34 und 89 Jahren. Die Gesamtanzahl der Operationen betrug 1542.

Name	Beschreibung
Patienten ID	N= 1542 Patienten
OP-Jahr	2012-2016
Geschlecht	1= männlich
	2= weiblich
Status	1= Elektiv
	2= Dringend
	3= Notfall
	4= Lebensbedrohend
30-Tages-Status	1= lebend
	2= tot
Alter	34-89 Jahre
Add. EuroSCORE	Additiver EuroSCORE
Log. EuroSCORE	Logistischer EuroSCORE
EuroSCORE II	EuroSCORE II
Chronische Lungenerkrankung/ COPD	0= Nein
	1= Schwach
	2= Moderat
	3= Schwer
Periphere vaskuläre Verschlusskrankung/ PAVK	0= Nein
	1= Ja
Rankin Skala >= 4	0= Nein
	1= Ja
Kardiogener Schock	0= Nein
	1= Ja
Angina Pectoris Typ	1= Stabil
	2= Instabil
Auswurfraction	In %
	0= keiner
	1= <6 Stunden
	2= >=6 aber <24 Stunden
Myokardinfarkt Wann	3= 1-7 Tage
	4= 8-21 Tage
	5= 22-89 Tage
	6= >=90 Tage

Tabelle 7: Patientengut

Von den insgesamt 1542 Patienten und Patientinnen waren 1231 (79,8%) männlich. An einer chronischen Lungenerkrankung litten insgesamt 557 (36,1%), periphere vaskuläre Verschlusskrankungen konnten bei 352 (22,8%) nachgewiesen werden und 90 (5,8%) hatten einen kardiogenen Schock. Unter instabiler Angina Pectoris litten 431 (28%), die restlichen 72% wiesen entweder eine stabile Angina Pectoris oder gar keine auf. Die durchschnittliche linksventrikuläre Auswurffraktion betrug $53\% \pm 12\%$. Insgesamt war bei 574 Patienten (37,2%) ein Myokardinfarkt innerhalb der letzten 90 Tage vor der Operation aufgetreten.

		MW	SD	Med	Min.	Max.	N	%
Alter		67	9,8	68	34	89		
Geschlecht	Männlich						1231	79,80%
	Weiblich						311	20,20%
COPD	Nein						985	63,90%
	Schwach						210	13,60%
	Moderat						279	18,10%
	Schwer						68	4,40%
PAVK	Nein						1190	77,20%
	ja						352	22,80%
Kardiogener Schock	Nein						1452	94,20%
	Ja						90	5,80%
Angina Pectoris Typ	Stabil						1111	72,00%
	instabil						431	28,00%
Auswurffraktion (%)		53	12	60	1	87		
Myokardinfarkt Wann	keiner						866	56,20%
	<6h						62	4,00%
	>=6h; <24h						149	9,70%
	1-7d						165	10,70%
	8-21d						93	6,00%
	22-89d						105	6,80%
	>=90d						102	6,60%

Tabelle 8: Risikofaktoren

Von 2012 bis 2015 wurden jährlich durchschnittlich 346 Patienten und Patientinnen operiert, wobei im Jahr 2014 die höchste Zahl an Patienten und Patientinnen (N=391) zu zählen war. Das Jahr 2016 wurde hier nicht in die Berechnung eingeschlossen, da nur die Daten bis Juni vorhanden waren. Mehr als die Hälfte der Operationen hatte elektiven Status (61,9%), als dringend wurden 27,6% durchgeführt und als Notfall

galten 9,1% der Operationen. 1,5% aller Eingriffe wurden als „lebensbedrohend“ bezeichnet. Im Beobachtungszeitraum von 30 Tagen sind 59 Patienten und Patientinnen verstorben, das entspricht 3,8%. Diese Statistiken sind über die beobachteten Jahre hinweg sehr konstant. Das durchschnittliche Alter der Patienten und Patientinnen zum Zeitpunkt der Eingriffe betrug 66,96 Lebensjahre, wobei minimales und maximales Alter, in dieser Reihenfolge, 34 und 89 Lebensjahre betragen (siehe nachfolgende Tabellen 6, 7 und 8):

		N	%
OP-Jahr	2012	342	22,20%
	2013	323	20,90%
	2014	391	25,40%
	2015	327	21,20%
	2016	159	10,30%
Status	Elektiv	954	61,90%
	Dringend	425	27,60%
	Notfall	140	9,10%
	Lebensbedrohend	23	1,50%
30-Tages-Status	Lebend	1483	96,20%
	Tot	59	3,80%

Tabelle 9: Status und 30-Tages-Status

	MW	SD	Med.	Min.	Max.
Alter	67	9,8	68	34	89
Add. EuroSCORE	5,11	3,53	5	0	21
Log. EuroSCORE	7,33	11,01	3,5	0,9	93,3
EuroSCORE II	4,79	9	2	0,5	90,4

Tabelle 10: Alter und Scores der Patienten und Patientinnen

		OP-Jahr									
		2012		2013		2014		2015		2016	
		N	%	N	%	N	%	N	%	N	%
Status	Elektiv	232	67,8%	204	63,2%	227	58,1%	211	64,5%	80	50,3%
	Dringend	71	20,8%	90	27,9%	122	31,2%	84	25,7%	58	36,5%
	Notfall	37	10,8%	26	8,0%	31	7,9%	28	8,6%	18	11,3%
	Lebensbedr.	2	0,6%	3	0,9%	11	2,8%	4	1,2%	3	1,9%
30-Tages-	lebend	327	95,6%	313	96,9%	376	96,2%	313	95,7%	154	96,9%
Status	tot	15	4,4%	10	3,1%	15	3,8%	14	4,3%	5	3,1%

Tabelle 11: Status und 30-Tages-Status nach OP-Jahr

Der mittlere additive bzw. mittlere logarithmische EuroSCORE für die angegebenen Kenngrößen betrug 5,11 und 7,33% der mittlere EuroSCORE II betrug 4,79%.

Der höchste mittlere additive EuroSCORE betrug im Studienjahr 2016 mit $5,68 \pm 3,27\%$ (min. 0,0 vs. max. 19,0), während im Studienjahr 2013 der niedrigste mittlere additive EuroSCORE mit $4,78 \pm 3,36\%$ (min 0,0 vs. max. 20,0) zu finden war. Der mittlere logistische EuroSCORE zeigte ein ähnliches Verhalten, der höchste mittlere Wert war im Studienjahr 2016 mit $8,27\% \pm 11,1\%$ (min. 0,9% vs. max. 79,1%) zu finden und der niedrigste Wert im Studienjahr 2013 mit $6,39\% \pm 9,65\%$ (min. 0,9% vs. max. 88,6%). Diese Mittelwerte müssen jedoch nicht unbedingt in Zusammenhang mit dem Lebensalter der operierten Patienten und Patientinnen stehen, das höchste mittlere Alter ist im Studienjahr 2014 mit $67,96 \pm 9,84\%$ Jahren zu finden, das niedrigste im Jahr 2012 mit $66,15 \pm 10,09\%$ Jahren. Der EuroSCORE II hatte den höchsten Mittelwert im Studienjahr 2014 und betrug $5,58\% \pm 10,36\%$ (min. 0,6% vs. max. 90,4%). Den niedrigsten Mittelwert des EuroSCORE II fanden wir im Studienjahr 2013 mit $4,25\% \pm 8,4\%$ (min. 0,5% vs. max. 78,8%; siehe nachfolgende Tabelle 9):

		MW	SD	Med.	Min.	Max.
Alter	2012	66,15	10,09	67	36	86
	2013	66,17	9,84	68	34	88
	2014	67,96	9,84	69	36	89
	2015	67,02	9,65	68	42	88
	2016	67,75	8,94	68	43	84
Add. EuroSCORE	2012	5,12	3,76	5	0	21
	2013	4,78	3,36	4	0	20
	2014	5,28	3,61	5	0	21
	2015	4,93	3,41	4	0	18
	2016	5,68	3,27	5	0	19
Log. EuroSCORE	2012	7,65	11,79	3,5	0,9	91,7
	2013	6,39	9,65	3,2	0,9	88,6
	2014	7,84	11,89	3,7	0,9	93,3
	2015	6,83	10,24	3,5	0,9	78,7
	2016	8,27	11,1	4,2	0,9	79,1
EuroSCORE II	2012	4,71	8,4	2,1	0,5	74,9
	2013	4,25	8,4	1,8	0,5	78,8
	2014	5,58	10,36	2,1	0,6	90,4
	2015	4,43	9,27	1,7	0,5	77,6
	2016	4,9	7,03	2,6	0,59	47,1

Tabelle 12: Alter und Scores nach OP-Jahr

3.2. ROC-Analyse: Logistischer EuroSCORE vs. EuroSCORE II für den Zeitraum von 2012 bis 2016

Die Kurven der ROC-Analysen für den Einsatz von logistischem EuroSCORE und EuroSCORE II wurden als ein Prädiktor für die 30-Tagen-Mortalität gerechnet. Für den logistischen EuroSCORE bzw. EuroSCORE II wurden AUC - Werte von 0,778 (95% KI 0,719-0,837) und 0,789 (95% KI 0,732-0,846) beobachtet. Für den logistischen EuroSCORE mit einem Cut-off-Niveau von 5,95% (basierend auf dem Youden Index) ergab sich eine Sensitivität von 72,9% und eine Spezifität von 69,5%. Mit einem Cut-off-Niveau von 1,95% (ebenfalls auf dem Youden Index basierend) ergab sich für den EuroSCORE II eine Sensitivität von 91,5% und eine Spezifität von 49,2% (siehe nachfolgende Tabelle und Abbildungen):

30-Tages-Status	Gültige Anzahl (listenweise)
Positiv	59
Negativ	1483

Tabelle 13: Anzahl der Überlebenden und Toten nach 30 Tagen

	AUC (95% KI)	p-Wert	Cut-off	Sensitivität	Spezifität
Add. EuroSCORE	0,780 (0,721-0,838)	<0,001	0,695	69,50%	73,00%
Log. EuroSCORE	0,778 (0,719-0,837)	<0,001	5,95	72,90%	69,50%
EuroSCORE II	0,789 (0,732-0,846)	<0,001	1,95	91,50%	49,20%

Tabelle 14: Die AUC-Werte zeigen statistische Signifikanz für den Zeitraum von 2012-2016

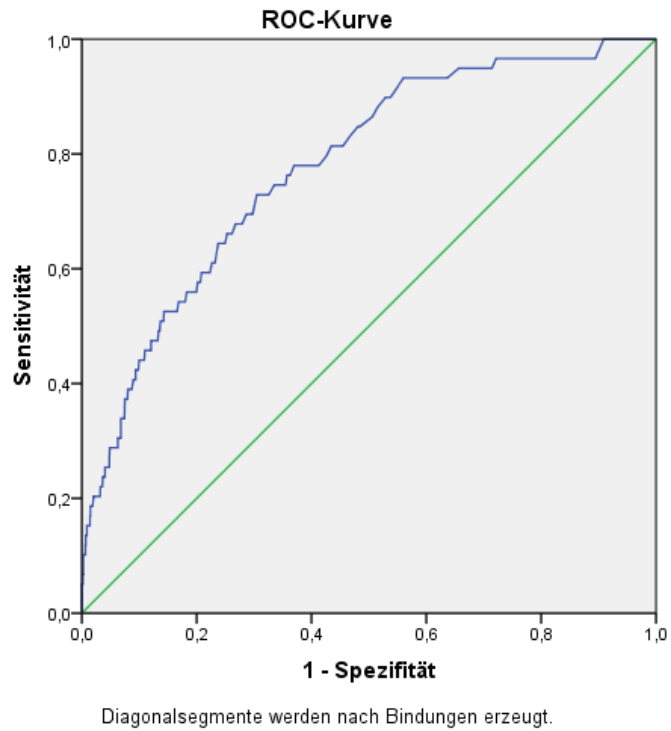


Abbildung 12: ROC-Kurve logistischer EuroSCORE für 2012-2016

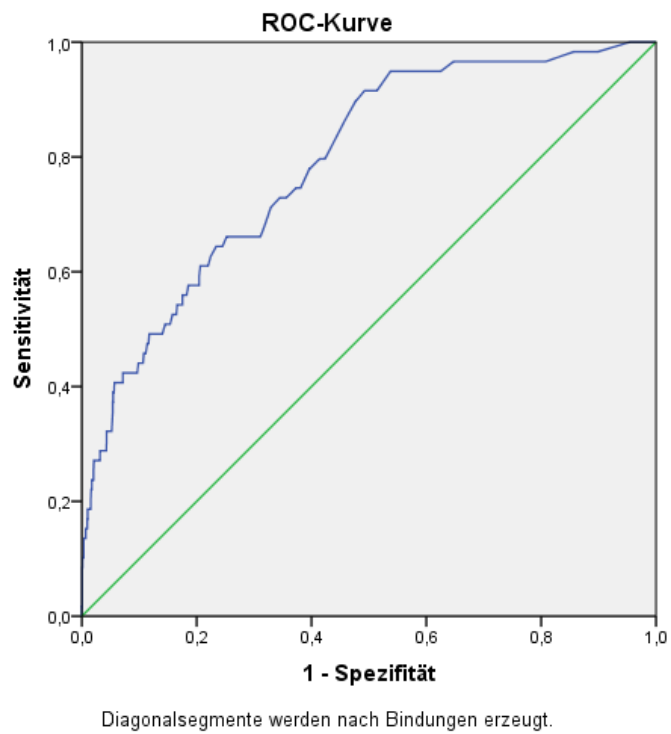


Abbildung 13: ROC-Kurve EuroSCORE II für 2012-2016

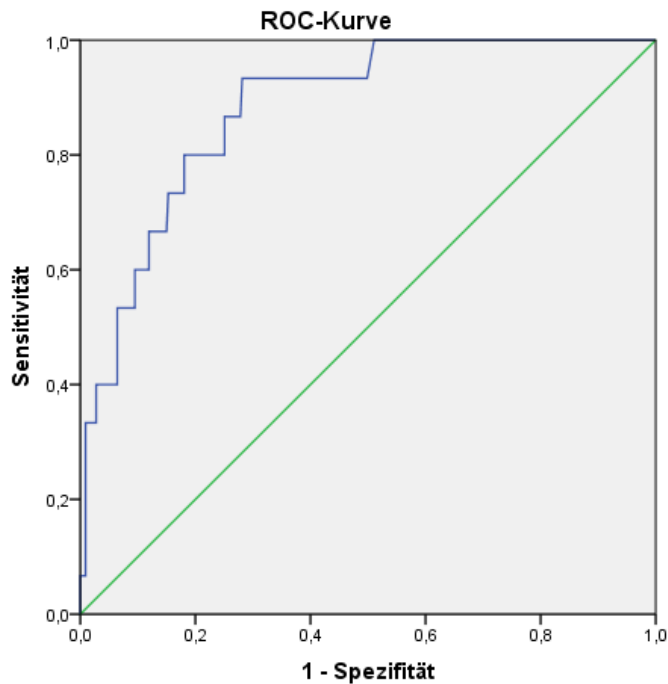
Im Weiteren finden sich die Aufschlüsselungen der Ergebnisse nach den Einzeljahren:

	OP-Jahr									
	2012		2013		2014		2015		2016	
	N	%	N	%	N	%	N	%	N	%
30-Tages- Lebend	327	96	313	97	376	96	313	96	154	97
Status Tot	15	4,4	10	3,1	15	3,8	14	4,3	5	3,1

Tabelle 15: 30-Tages-Status nach OP-Jahr

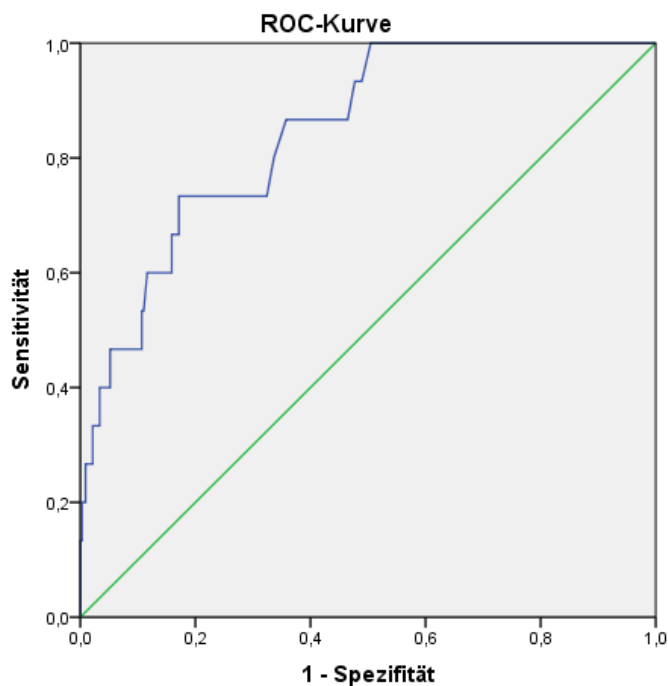
Jahr	Score	AUC (95%KI)	p-Wert
2012	Log. EuroSCORE	0,882 (0,810-0,953)	<0,001
	EuroSCORE II	0,846 (0,757-0,934)	<0,001
2013	Log. EuroSCORE	0,701 (0,543-0,859)	0,03
	EuroSCORE II	0,692 (0,541-0,844)	0,039
2014	Log. EuroSCORE	0,744 (0,636-0,852)	0,001
	EuroSCORE II	0,788 (0,691-0,884)	<0,001
2015	Log. EuroSCORE	0,783 (0,645-0,921)	<0,001
	EuroSCORE II	0,798 (0,660-0,936)	<0,001
2016	Log. EuroSCORE	0,720 (0,469-0,971)	0,094
	EuroSCORE II	0,844 (0,707-0,920)	0,009

Tabelle 16: Scores und AUC mit KI nach OP-Jahr



Diagonalsegmente werden nach Bindungen erzeugt.

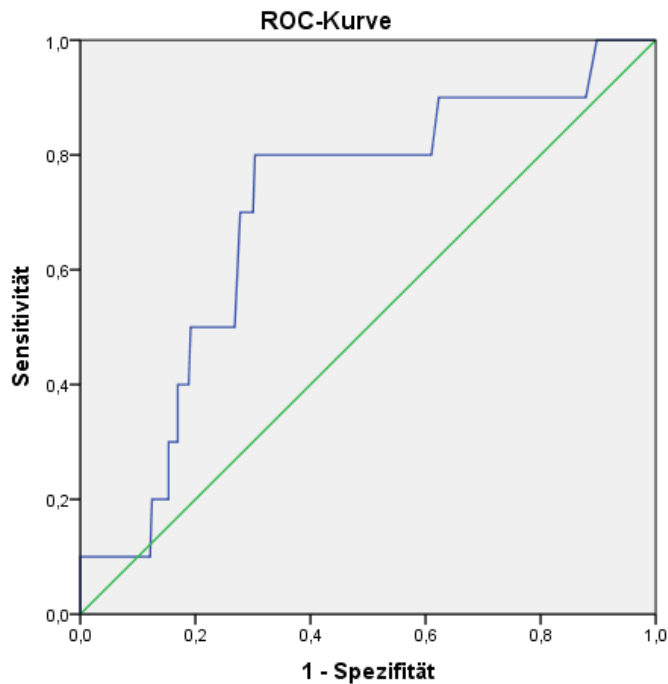
Abbildung 14: logistischer EuroSCORE im Beobachtungsjahr 2012



Diagonalsegmente werden nach Bindungen erzeugt.

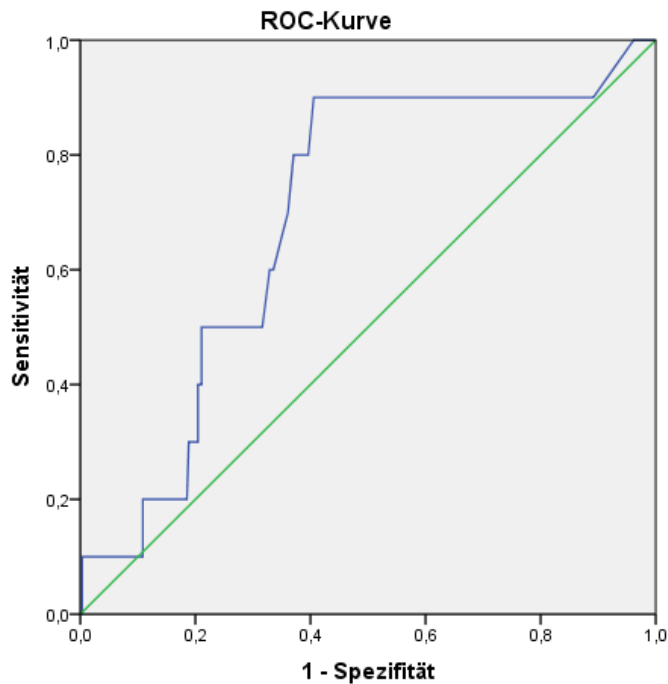
Abbildung 15: EuroSCORE II im Beobachtungsjahr 2012

Die AUC (95% KI)-Werte für den logistischen EuroSCORE und EuroSCORE II waren im Jahr 2012 0,882 (0,810-0,953) und 0,846 (0,757-0,934).



Diagonalsegmente werden nach Bindungen erzeugt.

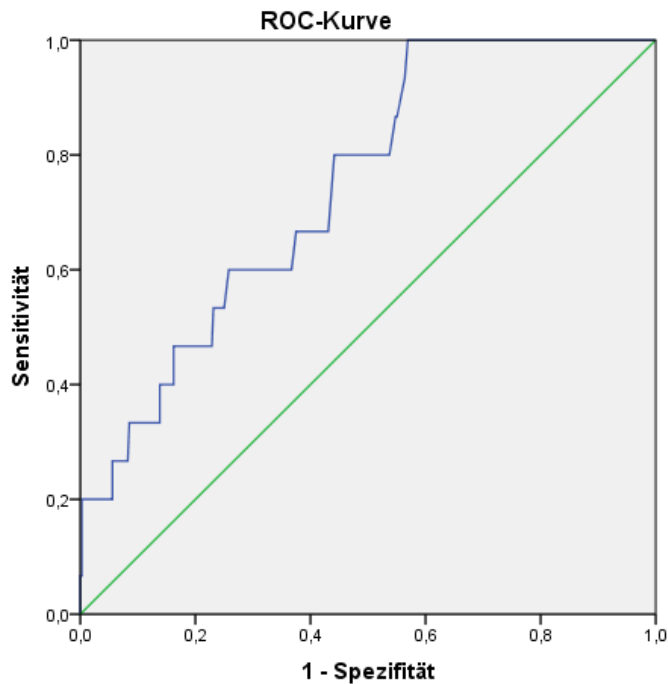
Abbildung 16: logistischer EuroSCORE im Beobachtungsjahr 2013



Diagonalsegmente werden nach Bindungen erzeugt.

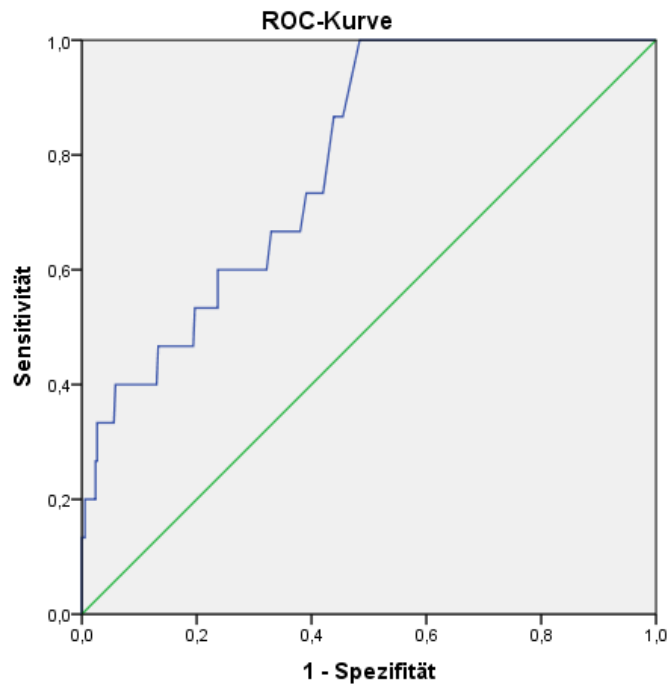
Abbildung 17: logistischer EuroSCORE im Beobachtungsjahr 2013

Die AUC (95% KI)-Werte für den logistischen EuroSCORE und EuroSCORE II waren im Jahr 2013 0,701 (0,543-0,859) und 0,692 (0,541-0,844).



Diagonalsegmente werden nach Bindungen erzeugt.

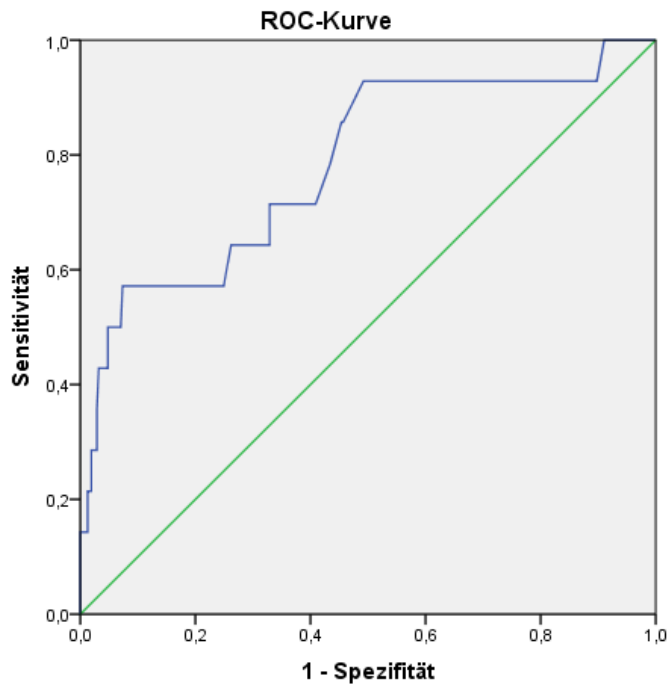
Abbildung 18: logistischer EuroSCORE im Beobachtungsjahr 2014



Diagonalsegmente werden nach Bindungen erzeugt.

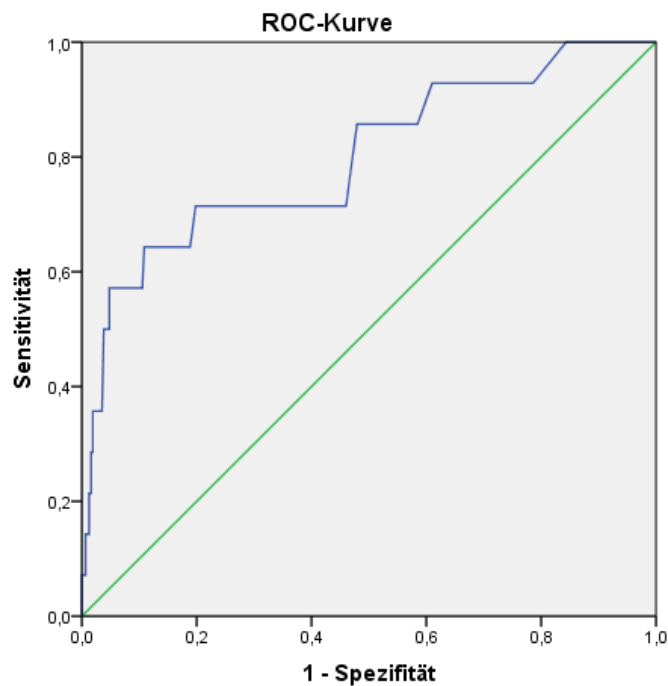
Abbildung 19: EuroSCORE II im Beobachtungsjahr 2014

Die AUC (95% KI)-Werte für den logistischen EuroSCORE und EuroSCORE II waren im Jahr 2014 0,744 (0,636-0,852) und 0,788 (0,691-0,884).



Diagonalsegmente werden nach Bindungen erzeugt.

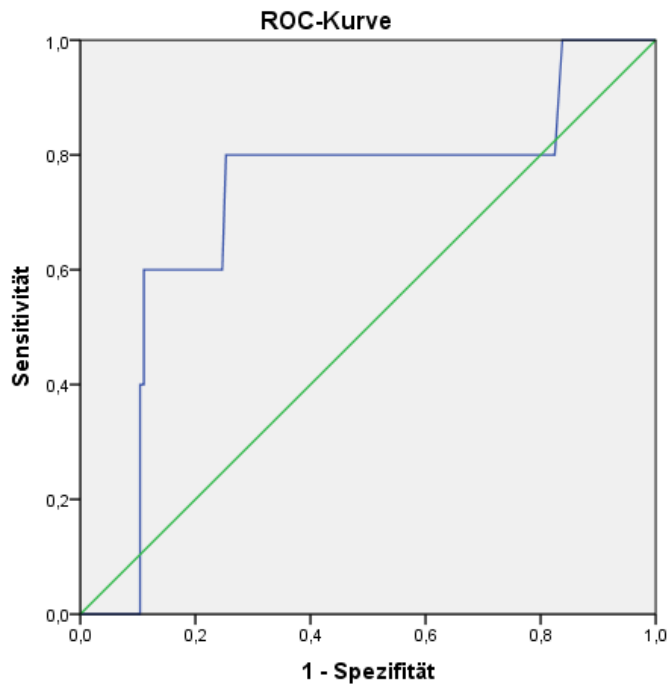
Abbildung 20: logistischer EuroSCORE im Beobachtungsjahr 2015



Diagonalsegmente werden nach Bindungen erzeugt.

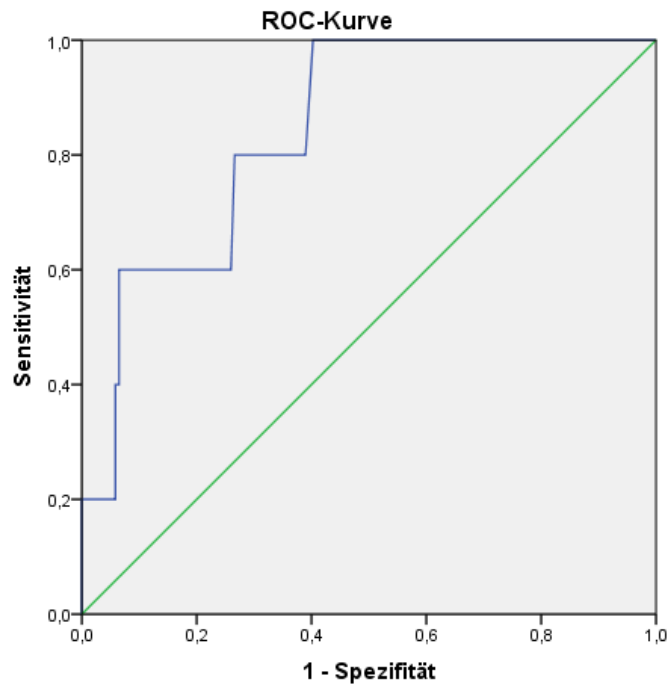
Abbildung 21: EuroSCORE II im Beobachtungsjahr 2015

Die AUC (95% KI)-Werte für den logistischen EuroSCORE und EuroSCORE II waren im Jahr 2015 0,783 (0,645-0,921) und 0,798 (0,660-0,936).



Diagonalsegmente werden nach Bindungen erzeugt.

Abbildung 22: logistischer EuroSCORE im Beobachtungsjahr 2016



Diagonalsegmente werden nach Bindungen erzeugt.

Abbildung 23: EuroSCORE II im Beobachtungsjahr 2016

Die AUC (95% KI)-Werte für den logistischen EuroSCORE und EuroSCORE II waren im Jahr 2016 0,720 (0,469-0,971) und 0,844 (0,707-0,920).

3.3. Funnelplot und VLAD-Analysen

Gesamtübersicht

Operationen

	Summe	Männlich	Weiblich
Summe	1411	1122	289
Lebend	1345	1076	269
Tot	56	41	15
Unbekannt	33	20	13

Tabelle 17: Gesamtübersicht über die Ergebnisse

3.3.1. Funnelplot und VLAD für den logistischen EuroSCORE (2012-2016)

Die Tabelle zeigt die Auswertung des Funnelplots für den logistischen EuroSCORE für den Zeitraum 2012-2015. Darin resultiert die Performance Ratio von 0.56 als außerordentlich positives Ergebnis.

Ergebnisse

OP-Anzahl	1401	1117	284
Verstorbene	56	41	15
Verstorbene (% von Ges.)	4,00%	3,67%	5,28%
Durchschn. Score	7,10%	6,37%	8,55%
Ergebnis	0,56	0,55	0,62

Tabelle 18: Performance Ratio des logistischen EuroSCORE

Die daraus resultierende Abbildung für den VLAD bei logistischem EuroSCORE für den Beobachtungszeitraum 2012-2015 zeigt die aktuelle rote VLAD Kurve im hochpositiven Bereich.

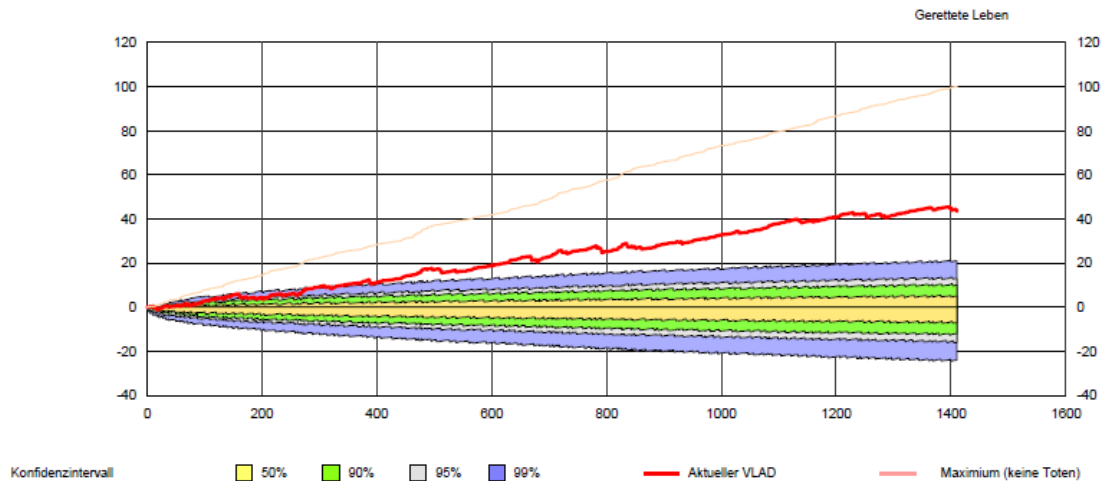


Abbildung 24: Funnelplot und VLAD für den logistischen EuroSCORE

3.3.2. Funnelplot und VLAD für den EuroSCORE II (2012-2016)

Die Tabelle zeigt die Auswertung des Funnelplots für den EuroSCORE II für den Zeitraum 2012-2015. Darin resultiert die Performance Ratio von 0.86 als außerordentlich positives Ergebnis.

Ergebnisse

OP-Anzahl	1401	1117	284
Verstorbene	56	41	15
Verstorbene (% von Ges.)	4,00%	3,67%	5,28%
Durchschn. Score	4,64%	4,42%	5,50%
Ergebnis	0,86	0,83	0,96

Tabelle 19: Performance Ratio des EuroSCORE II

Die Abbildung für den VLAD bei EuroSCORE II für den Zeitraum 2012-2015 zeigt die aktuelle rote VLAD Kurve ebenfalls im hochpositiven Bereich, jedoch flacher verlaufend

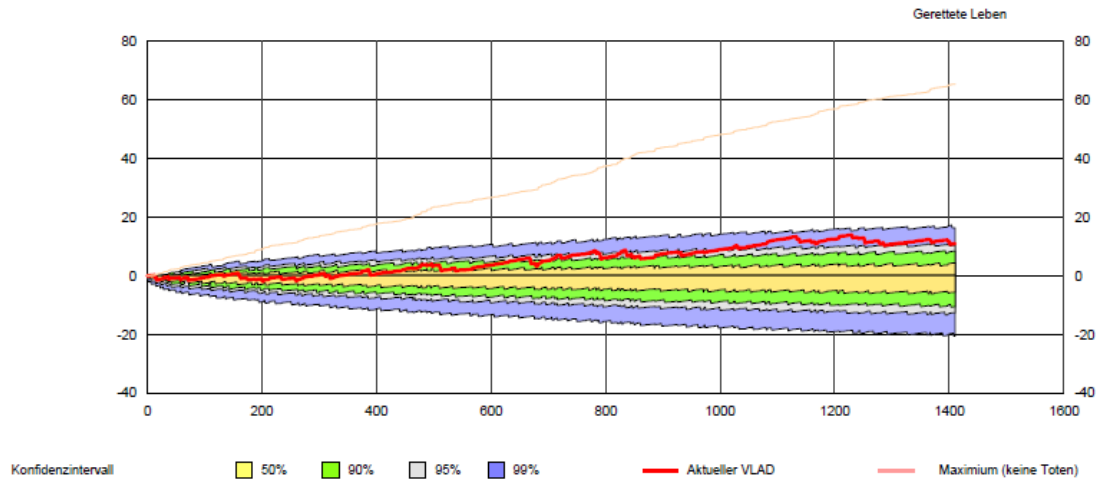


Abbildung 25: Funnelplot und VLAD für den EuroSCORE II

4. Diskussion

4.1. Allgemeines

Obwohl im Prinzip alle in einer Klinik tätigen Berufsgruppen gleichberechtigt sind, an der Qualitätsentwicklung mitzuwirken, kann Qualitätsmanagement nicht als Form der „Basisdemokratie“, sondern vielmehr als Führungsinstrument und Form der Klinikleitung angesehen werden. Qualitätsmanagement kann nur dann erfolgreich durchgeführt werden, wenn alle Mitarbeiter und Mitarbeiterinnen die Theorie des Systems akzeptieren und in ihren beruflichen Alltag integrieren. Zwar ist das Qualitätsmanagement als Führungsinstrument anzusehen, doch kann es nicht als rein von oben kommende Anordnung existieren, jeder einzelne Mitarbeiter und jede Mitarbeiterin sollte sich Gedanken zur Verbesserung der Qualität und zur Umsetzung machen und diese auch präsentieren können.

Besonders deutlich wird dies zur Geltung kommen, wenn Systeme zum Qualitätsmanagement nicht mehr aus dem Team selbst kommen, sondern von externen Beratern und Beraterinnen aufgebaut werden. Hier müssen in definierten zeitlichen Abständen konkrete Qualitätsziele formuliert werden, deren Erreichung an festgelegten Kriterien gemessen werden sollten. Wichtig ist, dass keine allgemeinen Aussagen wie z. B. „unsere Qualität muss besser werden“ getroffen werden. Es müssen empirisch messbare Kriterien gegeben sein, anhand derer man eine Verbesserung oder auch Verschlechterung eindeutig sehen kann. Am Ende einer Periode wird nun eine Bewertung erstellt, welche die Erreichung – oder ggf. auch Nichterreichung – der Ziele kritisch beurteilt. Da es nicht um die Beschönigung der Realität, sondern um eine möglichst realitätsnahe Betrachtung des Klinikalltages geht, können hier Analysen der Misserfolge bzw. des Nichterreichens gesteckter Ziele ebenso wichtig sein, wie die Veranschaulichung des Erreichens von Zielen.

4.2. Scoring

Die Grundlage dafür, dass Systeme zur Berechnung des Operationsrisikos auch zu Zwecken des Qualitätsmanagements herangezogen werden können, ist die

strukturierte Erfassung aller Daten, welche die Operation selbst, sowie die prä- und postoperative Periode betreffen. Hierfür gibt es Systeme, wie zum Beispiel QS II, welche darauf ausgelegt sind, mit minimalem Aufwand alle relevanten Daten eingeben zu können. Hierbei beruht die Datenstruktur auf weltweit normierten, praxistauglichen Records, was für eine gute Vergleichbarkeit sorgt, auch mit ausländischen Zentren mit anderen Systemen zur Datenerfassung. Auch wird die Gefahr, zu viele oder zu wenige Daten zu erfassen, durch die Verwendung des Standards minimiert.

Zur Vorhersage der Behandlungsergebnisse nach herzchirurgischen Eingriffen sind weltweit verschiedene Risikomodelle weit verbreitet. Die meisten davon integrieren die operative Mortalität als Maß für die Qualität der Eingriffe. Die Modelle zur Berechnung sollten möglichst realitätsgetreu sein, einfach zu handhaben, eindeutige Definitionen beinhalten, wenige Risikofaktoren erfordern und vor allem dem tatsächlichen Ausgang der Operation entsprechen.

Die Risikoscores sollen allen an der Entscheidung über das Stattfinden und die Art der Operation beteiligten Parteien möglichst objektive Kriterien bieten, anhand derer sie dann entscheiden können. Vor allem heutzutage, wo Begriffe wie „evidenzbasierte Medizin“ und der „mündige Patient“, der gut informiert mitentscheiden möchte, nicht mehr wegzudenken sind, können solche Modelle Hilfe durch die Abschätzung des Risikos bieten⁽⁴⁹⁾.

Natürlich dienen die Modelle auch der Qualitätssicherung, worauf der Fokus dieser Arbeit liegt. Optimale Vergleichbarkeit erhält man nur bei einem gut kalibrierten Score, während schlechte Scores dazu führen, die Ergebnisqualität einer herzchirurgischen Einheit oder einzelner Operateure und Operateurinnen zu über- oder unterschätzen.

Zwei solcher Risikostratifizierungsmodelle sind der STS (Risikoscore der Society of Thoracic Surgeons) und der EuroSCORE^(50,51,52,53).

Die ursprüngliche Version des logistischen EuroSCOREs wurde zwischen 1995 und 1999 aus einer europäischen Datenbank von mehr als 19 000 herzchirurgischen Patienten entwickelt. In den Jahren nach seiner Veröffentlichung wurde der logistische EuroSCORE weitgehend in Kliniken in Europa, Nordamerika und Asien implementiert um Risiken zu beurteilen und verbesserte operative Standards für kardiochirurgische Patienten zu etablieren. Im Laufe der beiden folgenden Jahrzehnte hatten Fortschritte in der herzchirurgischen Praxis, aber auch Veränderungen des Patientengutes bezüglich Risikofaktoren und Komorbiditäten zur Folge, dass der logistische

EuroSCORE das Risiko der Sterblichkeit bei Patienten und Patientinnen mit geringem vorausgesagten Risiko in Wirklichkeit überschätzte, wobei das Risiko für Patienten mit hohem Risiko infolgedessen massiv unterschätzt wurde^(50,54,51).

Außerdem wurde mit der Zeit klar, dass im ursprünglichen Modell einige Risikofaktoren, wie zum Beispiel die klinische Funktion des Organs Leber nicht berücksichtigte. Ebenso zeigte sich, dass die Kreatinin-Clearance anstelle des Serumkreatinins ein genaueres Maß für die Beurteilung der Nierenfunktion darstellte. Es konnte auch nachgewiesen werden, dass der logistische EuroSCORE das perioperative Mortalitätsrisiko bei Patienten und Patientinnen mit prothetischem Aortenklappenersatz erheblich überschätzt. Aufgrund dieser Einschränkungen und Nachteile wurde im Jahr 2012 der logistische EuroSCORE aktualisiert und in den EuroSCORE II überarbeitet, um die Berechnung des Risikos an die heute vorhandenen Techniken und Probleme anzupassen. Noch gibt es kaum Beweise, dass der EuroSCORE II möglicherweise dem logistischen EuroSCORE und dem STS Score beim Schätzen des Sterblichkeitsrisikos unterzuordnen wäre ⁽⁵⁵⁻⁷⁷⁾.

Die wohl wichtigste Arbeit diesbezüglich, welche als Metaanalyse angelegt wurde und die Risk-Prädiktionsmodelle des Euroscore II mit anderen vergleicht, wurde von Sullivan et al im Jahr 2016 im American Journal of Cardiology publiziert. Hier wurden drei gängige multivariable Prädiktionsmodelle verwendet, allen voran der Euroscore II. Für alle Publikationen, welche die Inklusionskriterien erfüllt hatten, wurden paarweise Vergleiche zwischen den verwendeten Risk-Modellen durchgeführt. Für die statistische Diskrimination wurden alle AUC-Werte berechnet und mit einem fünf prozentigen Konfidenzintervall in der Studie verglichen. Weiters wurde auch noch Metaanalysen der AUC-Schätzungen für drei Kernmodelle durchgeführt. Als Ergebnis zeigte sich, dass in 33 paarvergleichenden Publikationen unter den genannten drei Risk-Modellen der Euroscore II als signifikant hochpraktikabel zur Risk-Stratifikation vor herzchirurgischen Eingriffen zu bezeichnen war. Als Erklärung diente die Tatsache, dass die schlechter abschneidenden Scores (wie der STS oder der ACEF-Score) prinzipiell entwickelt worden waren, um jene Mortalitäten vorauszusagen, welche bei Patienten und Patientinnen bei rein elektiven herzchirurgischen Eingriffen auftreten. Somit könnten sie nie das Gesamtkollektiv aller herzchirurgischen Patienten und Patientinnen beschreiben. Eine weitere Begründung war, dass der Euroscore II über eine multiklinische Multicenter-Studie berechnet wurde und wird ⁽⁷⁸⁾.

Im Jahr 2014 wurde in *The Thoracic and Cardiovascular Surgeon* von Velicki et al die klinische Performance des EuroSCORE II mit der des Vorgängermodells verglichen. Die Studie inkludierte 1247 aufeinanderfolgende Herzpatienten und –patientinnen in einem Zeitraum von 14 Monaten ab Anfang 2012 mit dem Fokus auf die postoperative Mortalität. Sensitivität und Spezifität der Modelle wurden anhand des Kontingenzkoeffizienten C, der ein Maß für den statistischen Zusammenhang ist, verglichen. Hier konnte die Diskriminationskraft des EuroSCORE II mit einem C-Wert von 0,743 berechnet werden, während derselbe für den logistischen EuroSCORE 0,759 und für den additiven EuroSCORE 0,754 betrug. In dieser Studie konnte der EuroSCORE II zwar als praktikables Risikostratifizierungsmodell validiert werden, jedoch konnte keine Überlegenheit über den logistischen EuroSCORE gezeigt werden⁽⁷⁹⁾.

Qadir et al veröffentlichten im Februar 2014 eine retrospektive Studie, in welcher die Ergebnisse des EuroSCORE II mit denen des logistischen EuroSCORE und des STS-Scores verglichen wurden. Hier wurden die Daten von 2004 Patienten und Patientinnen, welcher einer isolierten Bypassoperation unterzogen wurden, ausgewertet und mit den Vorhersagen aller drei Modelle verglichen. Die vom EuroSCORE II vorhergesagte Mortalität betrug $3.72\% \pm 5.11\%$, die des logistischen EuroSCORE dagegen ergab $6.41\% \pm 10.06\%$ und die des additiven EuroSCORE betrug $4.35\% \pm 3.65\%$. Die tatsächliche Mortalität innerhalb der Kohorte betrug 3,8%. Die Diskriminationskraft des additiven EuroSCORE war hier besser als die des EuroSCORE II und auch die Kalibrierung des additiven EuroSCORE war hier mit einem C-Wert von 0,866 besser als die des EuroSCORE II mit 0,836. Auch der STS-Score zeigte hier in einer Untergruppe von 380 Patienten und Patientinnen mit einer vorausgesagten Mortalität von $2.30\% \pm 4.16\%$ gegen eine vorausgesagte Mortalität des EuroSCORE II von $4.27\% \pm 5.22\%$ bei einem beobachteten Wert von 2,89% eine bessere Aussagekraft. Hier konnte für den EuroSCORE II eine AUC von 0,759 und für den STS-Score eine AUC von 0,898 berechnet werden, womit in der Gruppe von Patienten und Patientinnen in Pakistan keine Überlegenheit des EuroSCORE II gegenüber den Vorgängermodellen gezeigt werden konnte⁽⁸⁰⁾.

Der EuroSCORE wird entweder als additiver oder als summarischer Score angewendet. Der additive EuroSCORE wird durch Addieren der einzelnen Risikoscores in einer Summe von 17 Risikovariablen berechnet, während der

logistische EuroSCORE nach dem logistischen Regressionsmodell dargestellt wird, und dadurch in eine Beschränkungen des additiven Modells transzendiert. Dem EuroSCORE II wurden neue Variablen hinzugefügt: die Kreatinin-Clearance und die Funktion der Leber. Dabei wurden die Definitionen und Typen der Variablen neu festgelegt und die Gewichtung einiger dieser Variablen wurde auch angepasst. Die Faktoren der Patientenzusammensetzung und der möglichen Abweichung der Patientenprobe von der Normalverteilung, welche tatsächlich verwendet wurde (die auch der normalverteilten Population der kardiochirurgischen Patienten und Patientinnen entspricht), könnte als eine wichtige Quelle für Fehler dienen.

Darüber hinaus ist es wahrscheinlich, dass nicht alle Risikovariablen auch wirklich als Risikofaktor für die 30-Tage Mortalität verstanden werden können. Die Charakterisierung eines Patienten oder einer Patientin mit einer nur limitierten Anzahl an Risikovariablen bietet kein komplettes klinisches Bild. Daher könnten beide - als potenzielle Fehlerquellen angesehen - dazu führen, die 30-Tage Mortalität zu überschätzen oder zu unterschätzen ^(50,51,81-86,55-61,68-76,87-89).

4.3. ROC-Analysen

In der vorliegenden Studie zeigte sich eine 30-Tages-Mortalität von 3,8%, wobei die Berechnungen durch den Logistischen EuroSCORE bzw. den EuroSCORE II einen Mittelwert von 7,33% bzw. 4,79% ergaben.

Wenn man diese Mittelwerte mit der tatsächlichen Mortalität vergleicht, zeigt sich, dass die Ergebnisse der EuroSCORE II näher an der Realität liegt, als es bei denen des logistischen EuroSCORE der Fall ist. Das Verhältnis von beobachteter und erwarteter Mortalität betrug beim logistischen EuroSCORE 0,52, während es beim EuroSCORE II 0,79 betrug.

Um die Trennschärfe der Scores zu beurteilen, wurden ROC-Kurven erstellt. Die ROC-Analysen beinhalten die Berechnung der Sensitivität und Spezifität, wobei der Cut-off Wert, an welchem die beiden Werte abgelesen werden, durch den Youden-Index definiert wird. Dieser Index ist ein Maß zur Beurteilung der Güte eines diagnostischen Tests und ist graphisch betrachtet jener Punkt der ROC-Kurve, welcher die größte

vertikale Distanz zur diagonalen Linie aufweist. Damit ist er der Punkt, an welchem die größtmögliche Spezifität bei größtmöglicher Sensitivität besteht ⁽⁹⁰⁾.

Des Weiteren kann aus der ROC-Analyse die Fläche unter der Kurve abgelesen werden. Im Falle unserer Berechnung gäbe ein Wert von 0,8 an, dass in 80% der Fälle jene Patienten und Patientinnen, die verstorben sind, eine höhere Risikovorhersage hatten, als solche, die überlebt haben. Alle AUC-Werte größer als 0,7 deuten auf statistische Signifikanz hin, eine ROC-Kurve nahe der Diagonalen, also mit einer AUC von etwa 0,5, deutet auf einen Zufallsprozess hin.

Für den logistischen EuroSCORE konnte ein AUC-Wert von 0,778 bei einem Konfidenzintervall von 0,719 bis 0,837 errechnet werden, während der AUC-Wert für den EuroSCORE II 0,789 bei einem Konfidenzintervall von 0,732 bis 0,846 betrug und somit nur geringfügig besser war. Die Werte waren aufgrund ihrer p-Werte von $>0,001$ statistisch signifikant.

Der Youden-Index ergab sowohl für den logistischen EuroSCORE, als auch für den EuroSCORE II 0,42, womit die beste Kombination aus Sensitivität und Spezifität bei beiden Systemen gleich ist.

Im Juni 2012 veröffentlichten Biancari et al in den *Annals of Thoracic Surgery* eine Studie zur Validierung des EuroSCORE II in einer Gruppe von 1027 Bypasspatienten und –patientinnen. Die tatsächlichen Ergebnisse wurden anhand von ROC-Analysen mit den Vorhersagen der Scores verglichen. Auch hier konnte die Überlegenheit des EuroSCORE II mit einer AUC von 0,852 gegenüber dem logistischen EuroSCORE mit einer AUC von 0,838 gezeigt werden. Das Verhältnis von erwarteter zu beobachteter Mortalität betrug beim logistischen EuroSCORE 1,8, während es beim EuroSCORE II mit 1,2 deutlich besser war. In unserer Studie betrugen die Werte zwar 0,52 bzw. 0,79 für den logistischen bzw. den EuroSCORE II, jedoch konnten auch wir hier ein besseres Verhältnis beim EuroSCORE II feststellen. Zusammenfassend ergab die Studie von Biancari et al, dass die Vorhersagekraft der operativen Mortalität des EuroSCORE II besser ist, als die des Vorgängers ⁽⁹¹⁾.

Chalmers et al publizierten im Juli 2012 eine Studie im *European Journal of Cardio-Thoracic Surgery*, in der sie den EuroSCORE II mit dem bestehenden logistischen EuroSCORE verglichen. Dies geschah basierend auf einer Kohorte von 5576 Patienten und Patientinnen, an welchen zwischen Jänner 2006 und März 2010 eine isolierte koronararterielle Bypassoperation, ein Aortenklappenersatz, ein

Mitralklappenersatz oder eine Kombination der Operationen durchgeführt wurde. Die Auswertung wurde sowohl für die gesamte Kohorte, als auch für jede einzelne Gruppe durchgeführt. Innerhalb der Bypassgruppe konnte nur eine statistisch nicht signifikante Verbesserung des C-Wertes von 0,76 auf 0,77 des EuroSCORE II im Vergleich zum logistischen EuroSCORE gezeigt werden. Durch den Hosmer-Lemeshow Test allerdings konnte hier eine bessere Kalibrierung des logistischen EuroSCORE nachgewiesen werden ⁽⁹²⁾.

Wie von Kalender et al im September 2014 im Polish Journal of Thoracic and Cardiovascular Surgery beschrieben, konnten auch wir anhand der ROC Analysen zeigen, dass der EuroSCORE II verglichen mit dem logistischen EuroSCORE eine bessere AUC aufweist, dass aber beide Scores eine gute diskriminative Power und Kalibrierung aufweisen, wenn es um die Risikoberechnung bei Bypasspatienten und –patientinnen geht.

In dieser Studie wurden die Daten von 105 Bypassoperationen an 80-89-jährigen Hochrisikopatienten und –patientinnen zwischen Jänner 2000 und Jänner 2010 mit den Ergebnissen des logistischen bzw. des EuroSCORE II verglichen. Beide Scores zeigten eine gute Diskriminationskraft, wobei die AUC des EuroSCORE II mit 0,772 (0,673-0,872 95% KI) besser war, als die des logistischen EuroSCORE⁽⁹³⁾.

Im März 2013 wurde im European Journal of Cardio-Thoracic Surgery eine Arbeit von Kirmani et al publiziert, in welcher die EuroSCORE Modelle mit dem STS-SCORE verglichen wurden. Die Daten von 14432 Operationen im Zeitraum von Februar 2001 bis März 2010 wurden herangezogen, die Risikoberechnungsmodelle wurden angewandt und die Ergebnisse wurden mittels ROC-Kurven präsentiert. Die durchgeführten Operationen beinhalteten Klappenersatzoperationen, Bypassoperationen sowie Kombinationen derselben. Es handelte sich um 9477 Bypassoperationen, 2964 Klappenoperationen und 1726 Kombinationen.

In der gesamten Kohorte betrug die Mortalität im Krankenhaus 3,5% (95% KI: 3,3–3,8). Die vorhergesagte Mortalität des EuroSCORE II betrug 2,53% (95% KI: 2,48-2,59), die AUC betrug 0,818. Beim STS-Score wurde die Mortalität mit 2,39% (95% KI: 2,34-2,45) berechnet und die AUC war 0,805. Somit wurde in dieser Studie für beide Modelle eine ähnliche Diskriminationskraft in der Britischen Population nachgewiesen ⁽⁹⁴⁾.

4.4. Funnel-Plots und VLADS

Wie bereits unter 2.1.2 beschrieben, bietet die VLAD-Analyse die Möglichkeit, die Überprüfung der Leistungsqualität an das individuelle Risiko der Patienten und Patientinnen anzupassen. Das erwartete Risiko der 30-Tages-Mortalität wird angepasst an Alter, Geschlecht und Komorbidität berechnet, das erwartete Risiko wird vom eintretenden Ergebnis abgezogen und dann werden die Ergebnisse in einem Diagramm dargestellt. Das Prinzip ist die „Anzahl geretteter Leben“.

Um die Aussage des VLAD in Abhängigkeit vom Patientenmix einschätzen zu können, brauchen wir eine Maßzahl der Güte und damit den Funnel Plot. Man erhält eine Grafik sowie die logistische Performance-Ratio, welche im Idealfall <1 beträgt. Im vorliegenden Fall ergab sich für den logistischen EuroSCORE eine Performance Ratio von 0,56, für den EuroSCORE II betrug der Wert 0,86. Beide Werte deuten auf ein außerordentlich positives Ergebnis hin, jedoch ist die Performance Ratio des EuroSCORE II deutlich besser.

4.5. Zusammenfassung

Zusammenfassend lässt sich nun sagen, dass es auf jeden Fall aufgrund der Veränderung der herzchirurgischen Möglichkeiten sowie der Komorbiditäten der Patienten Zeit für die Überarbeitung des EuroSCOREs war.

Es ist einerseits wichtig, akkurate Vorhersagen zur Verfügung zu haben, um die Patienten und Patientinnen so umfassend und genau wie möglich über die Vor- und Nachteile eines geplanten Eingriffes informieren zu können und so gemeinsam zu einer Entscheidung über das weitere Vorgehen zu kommen. Andererseits braucht man aber – heute mehr als jemals zuvor – objektive Parameter, anhand welcher die Qualität eines Ergebnisses sowie der Operateure und Operateurinnen im Nachhinein möglichst gut beurteilt werden kann. Nur so kann erkannt werden, ob die erbrachte Leistung der Ärzte und Ärztinnen auch der erforderlichen entspricht oder gar über dem Durchschnitt liegt.

Durch diese Studie konnte gezeigt werden, dass zumindest im vorliegenden Patientenmix von CABG-Patienten und Patientinnen der EuroSCORE II mit einer besseren klinischen Performance als der logistische EuroSCORE überzeugen kann und dass dieser gewisse Unzulänglichkeiten seines Vorgängers ausgleichen kann.

Es ist wichtig, zu bedenken, dass man sich nie vollständig auf die Vorhersage eines Risikostratifizierungsprogrammes verlassen können wird, da es sich immer noch um ein computergesteuertes Programm handelt, welches natürlich Einflüsse, wie die Verfassung des operierenden Arztes oder der operierenden Ärztin oder besondere Vorkommnisse vor, während oder nach der Operation nicht miteinbeziehen kann. Aber je besser solche Modelle kalibriert und an verschiedene Patienten- und Risikogruppen angepasst sind, desto eher bekommt man eine akkurate Aussage über den möglichen Ausgang der Operation. Es ist von großer Bedeutung für alle Beteiligten, zumindest unabhängig von solchen Einflüssen, das Risiko einer Operation und die Wahrscheinlichkeit für einen positiven Ausgang derselben möglichst genau abschätzen zu können. Sowohl die behandelnden Ärzte und Ärztinnen, als auch die betroffenen Patienten und Patientinnen sollen, mit dem Wissen, dass es sich immer noch „nur“ um Statistiken handelt, mithilfe der Scores weitestgehend aussagekräftige Entscheidungshilfen zur Verfügung haben.

Auch zum Zwecke der Qualitätssicherung und –verbesserung kann man mit einem gut kalibrierten, aussagekräftigen System viel eher eine Feststellung darüber treffen, inwieweit die erbrachte Leistung einer Abteilung verbesserungswürdig ist. Man sollte sich auch vor Augen führen, dass bei einer ausreichend großen Fallzahl außergewöhnliche Ereignisse, welche den Operations- und Therapieverlauf möglicherweise positiv oder negativ beeinflussen können, in den Hintergrund rücken. Gerade wenn es um die Qualitätssicherung innerhalb einer Abteilung geht, erfolgt die retrospektive Auswertung der Leistung der Operateure und Operateurinnen eher anhand größerer Datenmengen, um so ein möglichst realitätsnahes Bild des Qualitätsverlaufes auch über längere Perioden hinweg zu erhalten. Subjektive Faktoren und sonstige Umwelteinflüsse werden nie in ein System zur Risikostratifizierung miteinbezogen werden können. Im Rahmen des Möglichen jedoch ist eine stetige Anpassung an relevante Risikofaktoren und verbesserte Techniken der Chirurgie sinnvoll und notwendig.

Literaturverzeichnis

1. Aktuerk, Dincer; McNulty, David; Ray, Daniel; Begaj, Irena; Howell, Neil; Freemantle, Nick; Pagano, Domenico (2016): National administrative data produces an accurate and stable risk prediction model for short-term and 1-year mortality following cardiac surgery. In *International journal of cardiology* 203, pp. 196–203. DOI: 10.1016/j.ijcard.2015.10.086.
2. Collas, Valerie M.; van de Heyning, Caroline M.; Paelinck, Bernard P.; Rodrigus, Inez E.; Vrints, Christiaan J.; Bosmans, Johan M. (2016): Validation of transcatheter aortic valve implantation risk scores in relation to early and mid-term survival: a single-centre study. In *Interactive cardiovascular and thoracic surgery* 22 (3), pp. 273–279. DOI: 10.1093/icvts/ivv340.
3. Csordas, Adam; Nietlispach, Fabian; Schuetz, Philipp; Huber, Andreas; Muller, Beat; Maisano, Francesco et al. (2015): Midregional Proadrenomedullin Improves Risk Stratification beyond Surgical Risk Scores in Patients Undergoing Transcatheter Aortic Valve Replacement. In *PloS one* 10 (12), e0143761. DOI: 10.1371/journal.pone.0143761.
4. Garcia-Valentin, Antonio; Mestres, Carlos A.; Bernabeu, Eduardo; Bahamonde, Jose A.; Martin, Ivan; Rueda, Cristina et al. (2016): Validation and quality measurements for EuroSCORE and EuroSCORE II in the Spanish cardiac surgical population: a prospective, multicentre study. In *European journal of cardio-thoracic surgery : official journal of the European Association for Cardio-thoracic Surgery* 49 (2), pp. 399–405. DOI: 10.1093/ejcts/ezv090.
5. Guida, Pietro; Mastro, Florinda; Scrascia, Giuseppe; Whitlock, Richard; Paparella, Domenico (2014): Performance of the European System for Cardiac Operative Risk Evaluation II: a meta-analysis of 22 studies involving 145,592 cardiac surgery procedures. In *The Journal of thoracic and cardiovascular surgery* 148 (6), 3049-57.e1. DOI: 10.1016/j.jtcvs.2014.07.039.
6. Janikowski, Kamil; Morawiec, Robert; Jegier, Bogdan; Jaszewski, Ryszard; Lelonek, Malgorzata (2016): EuroSCORE II does not show better accuracy nor predictive power in comparison to original EuroSCORE: a single-centre study. In *Kardiologia polska* 74 (5), pp. 469–475. DOI: 10.5603/KP.a2015.0215.

7. Kalender, Mehmet; Adademir, Taylan; Tasar, Mehmet; Ecevit, Ata Niyazi; Karaca, Okay Guven; Salihi, Salih et al. (2014): Validation of EuroSCORE II risk model for coronary artery bypass surgery in high-risk patients. In *Kardiochirurgia i torakochirurgia polska = Polish journal of cardio-thoracic surgery* 11 (3), pp. 252–256. DOI: 10.5114/kitp.2014.45672.
8. Metz, Charles E. (2006): Receiver operating characteristic analysis: a tool for the quantitative evaluation of observer performance and imaging systems. In *Journal of the American College of Radiology : JACR* 3 (6), pp. 413–422. DOI: 10.1016/j.jacr.2006.02.021.
9. Moscarelli, Marco; Bianchi, Giacomo; Margaryan, Rafik; Cerillo, Alfredo; Farneti, Pierandrea; Murzi, Michele; Solinas, Marco (2015): Accuracy of EuroSCORE II in patients undergoing minimally invasive mitral valve surgery. In *Interactive cardiovascular and thoracic surgery* 21 (6), pp. 748–753. DOI: 10.1093/icvts/ivv265.
10. Müller, M. *Chirurgie für Studium und Praxis, Unter Berücksichtigung des Gegenstandskataloges und der mündlichen Examina in den Ärztlichen Prüfungen*, 12. Auflage. Breisach: Medizinische Verlags- und Informationsdienste (2014/15).
11. Smith P.K. et al. (2011): ACCF/AHA Guideline for Coronary Artery Bypass Graft Surgery: executive summary: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines *Circulation*: 124(23):2610-42.
12. Striebel, HW. et al.: *Anästhesie, Intensivmedizin, Notfallmedizin: für Studium und Ausbildung; mit 61 Tabellen 6. aktualisierte und erw. Aufl.* Stuttgart: Schattauer (2005).
13. Tscheliessnigg K, Uranüs S, Pierer G. et al: *Lehrbuch der allgemeinen und speziellen Chirurgie: 55 Tabellen. 3. Aufl.* Wien, München: Maudrich (2005).
14. *Tübinger Curriculum. Herz und Gefäße, Ein handlungsorientierter Leitfaden für Medizinstudenten*, Stuttgart: Thieme (2006).
15. Qualitätsmanagement- Total Quality Management Statistical Process Control: *Definiton: Qualitätsmanagement.* Available from: http://quality.kenline.de/seiten_d/qm_definition.htm (5.1.2017)
16. Duckett J.S, Coory M, Sketcher- Baker K (2007): Identifying variations in quality of care in Queensland hospitals. In *Medical Journal of Australia* 187(10), pp. 571-575.

Available from:
https://www.mja.com.au/system/files/issues/187_10_191107/duc10459_fm.pdf

17. Kleine, P., Ennker, J. (2008). Qualitätsmanagement in der operativen Medizin. Leitfaden für Ärzte und Pflegekräfte. Wiesbaden: Steinkopff Verlag.

18. Roques F, Michel P, Goldstone AR, Nashef SA (2004): The logistic Euro SCORE. In European Heart Journal . 24(9):882-3.

19. Institut für Qualität und Patientensicherheit : Risikoadjustierung in der Herzchirurgie. Kapitel 3, pp. 26-29. Available from: <http://info.bqs-online.de/outcome/MKL/Risiko.pdf> (5.1.2017).

20. Borde D, Gandhe U, Hargave N, Khullar V, Pandey K (2013): The application of European system for cardiac operative risk evaluation II (EuroSCORE II) and Society of Thoracic Surgeons (STS) risk-score for risk stratification in Indian patients undergoing cardiac surgery. In Annals of Cardiac Anaesthesia 16(3), pp. 163-166. Available from: <http://medind.nic.in/aad/t13/i3/aadt13i3p163.html>

21. Euroscore: *What is Euroscore?* Available from: http://www.euroscore.org/what_is_euroscore.htm (5.1.2017)

22. Metz CE. Quality of the observed image. ICRU News 1989;2:20-3.

23. Fryback DG, Thornbury JR. The efficacy of diagnostic imaging. Med Decis Making 1991;11:88-94.

24. Metz CE. Basic principles of ROC analysis. Semin Nucl Med 1978;8: 283-98.

25. Metz CE. ROC methodology in radiologic imaging. Invest Radiol 1986; 21:720-33.

26. Swets JA. ROC analysis applied to the evaluation of medical imaging techniques. Invest Radiol 1979;14:109-21.

27. McLish DK. Analyzing a portion of the ROC curve. Med Decis Making 1989; 9:190-5.

28. Jiang Y, Metz CE, Nishikawa RM. A receiver operating characteristic partial area index for highly sensitive diagnostic tests. Radiology 1996; 201:745-50.

29. Halpern EJ, Alpert M, Krieger AM, Metz CE, Maidment AD. Comparisons of ROC curves on the basis of optimal operating points. Acad Radiol 1996;3:245-53.

30. Starr SJ, Metz CE, Lusted LB, Goodenough DJ. Visual detection and localization of radiographic images. *Radiology* 1975;116:533-8.
31. Swensson RG. Unified measurement of observer performance in detecting and localizing target objects on images. *Med Phys* 1996;23:1709-25.
32. Egan JP, Greenberg GZ, Schulman AI. Operating characteristics, signal detection, and the method of free response. *J Acoust Soc Am* 1961;33: 993-1007.
33. Bunch PC, Hamilton JF, Sanderson GK, Simmons AH. A free response approach to the measurement and characterization of radiographic observer performance. *J Appl Photog Eng* 1978;4:166-72.
34. Berbaum KS, Franken EA, Dorfman DD, et al. Satisfaction of search in diagnostic radiology. *Invest Radiol* 1990;25:133-40.
35. Chakraborty DP, Winter LHL. Free-response methodology: alternate analysis and a new observer-performance experiment. *Radiology* 1990; 174:873-81.
36. Chakraborty DP. Statistical power in observer performance studies: a comparison of the ROC and free-response methods in tasks involving localization. *Acad Radiol* 2002;9:147-56.
37. Edwards DC, Metz CE, Kupinski MA. Ideal observers and optimal ROC hypersurfaces in N-class classification. *IEEE Trans Med Imaging* 2004; 23:891-5.
38. Edwards DC, Metz CE. Review of several proposed three-class classification decision rules and their relation to the ideal observer decision rule. *Proc SPIE Med Imag* 2005;5749:128-37.
39. Dorfman DD, Alf E. Maximum likelihood estimation of parameters of signal detection theory and determination of confidence intervals—rating method data. *J Math Psych* 1969;6:487-96.
40. Swets JA. Form of empirical ROCs in discrimination and diagnostic tasks: implications for theory and measurement of performance. *Psychol Bull* 1986;99:181-98.
41. Hanley JA. The robustness of the “binormal” assumptions used in fitting ROC curves. *Med Decis Making* 1988;8:197-203.
42. Metz CE. Some practical issues of experimental design and data analysis in radiological ROC studies. *Invest Radiol* 1989;24:234-45.

43. Dorfman DD, Berbaum KS, Metz CE, Lenth RV, Hanley JA, Dagga HA. Proper ROC analysis: the bigamma model. *Acad Radiol* 1997;4:138-49.
44. Pan X, Metz CE. The "proper" binormal model: parametric ROC curve estimation with degenerate data. *Acad Radiol* 1997;4:380-9.
45. Metz CE, Pan X. "Proper" binormal ROC curves: theory and maximumlikelihood estimation. *J Math Psych* 1997;43:1-33.
46. Roe CA, Metz CE. Variance-component modeling in the analysis of receiver operating characteristic index estimates. *Acad Radiol* 1997;4: 587-600.
47. Wagner RF, Beiden SV, Campbell G, Metz CE, Sacks WM. Assessment of medical imaging and computer-assist systems: lessons from recent experience. *Acad Radiol* 2002;8:1264-77.
48. Metz CE. Quantification of failure to demonstrate statistical significance: the usefulness of confidence intervals. *Invest Radiol* 1993;28:59-63. 422 *Journal of the American College of Radiology/ Vol. 3 No. 6 June 2006*
49. (Head SJ, Bogers AJ, Serruys PW, Takkenberg JJ, Kappetein AP. A crucial factor in shared decision making: the team approach. *Lancet*. May 28 2011;377(9780):1836.)
50. Roques F, Nashef SA, Michel P, Gauducheau E, Vincentiis C de, Baudet E et al. Risk factors and outcome in European cardiac surgery: analysis of the EuroSCORE multinational database of 19030 patients. *Eur J Cardiothorac Surg* 1999; 15(6):816-22; discussion 822-3.
51. Roques F, Nashef SA, Michel P, Pinna Pintor P, David M, Baudet E. Does EuroSCORE work in individual European countries? *Eur J Cardiothorac Surg* 2000; 18(1):27–30.
52. Nashef SA, Roques F, Michel P, Gauducheau E, Lemeshow S, Salamon R. European system for cardiac operative risk evaluation (EuroSCORE). *Eur J Cardiothorac Surg* 1999; 16(1):9–13.
53. Nashef SA, Roques F, Michel P, Cortina J, Faichney A, Gams E et al. Coronary surgery in Europe: comparison of the national subsets of the European system for cardiac operative risk evaluation database. *Eur J Cardiothorac Surg* 2000; 17(4):396–9.)

54. Nashef SAM, Roques F, Hammill BG, Peterson ED, Michel P, Grover FL et al. Validation of European System for Cardiac Operative Risk Evaluation (EuroSCORE) in North American cardiac surgery. *Eur J Cardiothorac Surg* 2002; 22(1):101–5.
55. Wang C, Tang Y-F, Zhang J-J, Bai Y-F, Yu Y-C, Zhang G-X et al. Comparison of four risk scores for in-hospital mortality in patients undergoing heart valve surgery: A multicenter study in a Chinese population. *Heart Lung* 2016; 45(5):423–8.
56. Wang TKM, Li AY, Ramanathan T, Stewart RAH, Gamble G, White HD. Comparison of four risk scores for contemporary isolated coronary artery bypass grafting. *Heart Lung Circ* 2014; 23(5):469–74.
57. 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. *J Thorac Cardiovasc Surg* 2015; 149(2):443–8.
58. Wang TKM, Choi DH-M, Haydock D, Gamble G, Stewart R, Ruygrok P. Comparison of Risk Scores for Prediction of Complications following Aortic Valve Replacement. *Heart Lung Circ* 2015; 24(6):595–601.
59. Rosenhek R, lung B, Tornos P, Antunes MJ, Prendergast BD, Otto CM et al. ESC Working Group on Valvular Heart Disease Position Paper: assessing the risk of interventions in patients with valvular heart disease. *Eur Heart J* 2012; 33(7):822-8, 828a, 828b.
60. Rabbani MS, Qadir I, Ahmed Y, Gul M, Sharif H. Heart valve surgery: EuroSCORE vs. EuroSCORE II vs. Society of Thoracic Surgeons score. *Heart Int* 2014; 9(2):53–8.
61. Paparella D, Guida P, Di Eusanio G, Caparrotti S, Gregorini R, Cassese M et al. Risk stratification for in-hospital mortality after cardiac surgery: external validation of EuroSCORE II in a prospective regional registry. *Eur J Cardiothorac Surg* 2014; 46(5):840–8.
62. Nashef SAM. Risk scores and how to evaluate them. *Eur J Cardiothorac Surg* 2016; 50(3):519.
63. Kieser TM, Rose MS, Head SJ. Comparison of logistic EuroSCORE and EuroSCORE II in predicting operative mortality of 1125 total arterial operations. *Eur J Cardiothorac Surg* 2016; 50(3):509–18.

64. Kalender M, Adademir T, Tasar M, Ecevit AN, Karaca OG, Salihi S et al. Validation of EuroSCORE II risk model for coronary artery bypass surgery in high-risk patients. *Kardiochir Torakochirurgia Pol* 2014; 11(3):252–6.
65. Holinski S, Jessen S, Neumann K, Konertz W. Predictive Power and Implication of EuroSCORE, EuroSCORE II and STS Score for Isolated Repeated Aortic Valve Replacement. *Ann Thorac Cardiovasc Surg* 2015; 21(3):242–6.
66. Guida P, Mastro F, Scrascia G, Whitlock R, Paparella D. Performance of the European System for Cardiac Operative Risk Evaluation II: a meta-analysis of 22 studies involving 145,592 cardiac surgery procedures. *J Thorac Cardiovasc Surg* 2014; 148(6):3049-57.e1.
67. Garcia-Valentin A, Mestres CA, Bernabeu E, Bahamonde JA, Martin I, Rueda C et al. Validation and quality measurements for EuroSCORE and EuroSCORE II in the Spanish cardiac surgical population: a prospective, multicentre study. *Eur J Cardiothorac Surg* 2016; 49(2):399–405.
68. Collas VM, van de Heyning CM, Paelinck BP, Rodrigus IE, Vrints CJ, Bosmans JM. Validation of transcatheter aortic valve implantation risk scores in relation to early and mid-term survival: a single-centre study. *Interact Cardiovasc Thorac Surg* 2016; 22(3):273–9.
69. Cockburn J, Dooley M, Trivedi U, Belder A de, Hildick-Smith D. Which surgical risk score is best for predicting outcome in patients undergoing trans-catheter aortic valve implantation: comparison of additive EuroScore, logistic EuroScore, EuroScore II, and the society of thoracic surgery global scoring systems. *J Cardiovasc Surg (Torino)* 2016.
70. Chan V, Ahrari A, Ruel M, Elmistekawy E, Hynes M, Mesana TG. Perioperative deaths after mitral valve operations may be overestimated by contemporary risk models. *Ann Thorac Surg* 2014; 98(2):605-10; discussion 610.
71. Borde D, Gandhe U, Hargave N, Pandey K, Khullar V. The application of European system for cardiac operative risk evaluation II (EuroSCORE II) and Society of Thoracic Surgeons (STS) risk-score for risk stratification in Indian patients undergoing cardiac surgery. *Ann Card Anaesth* 2013; 16(3):163–6.
72. Barili F, Pacini D, Rosato F, Roberto M, Battisti A, Grossi C et al. In-hospital mortality risk assessment in elective and non-elective cardiac surgery: a comparison

between EuroSCORE II and age, creatinine, ejection fraction score. *Eur J Cardiothorac Surg* 2014; 46(1):44–8.

73. Barili F, Pacini D, Rosato F, Parolari A. The role of surgical procedures on discriminative performance of the updated euroSCORE II. *J Thorac Cardiovasc Surg* 2013; 146(4):986–7.

74. Barili F, Pacini D, Grossi C, Di Bartolomeo R, Alamanni F, Parolari A. Reliability of new scores in predicting perioperative mortality after mitral valve surgery. *J Thorac Cardiovasc Surg* 2014; 147(3):1008–12.

75. Barili F, Pacini D, D'Ovidio M, Dang NC, Alamanni F, Di Bartolomeo R et al. The Impact of EuroSCORE II Risk Factors on Prediction of Long-Term Mortality. *Ann Thorac Surg* 2016; 102(4):1296–303.

76. Barili F, Pacini D, Capo A, Rasovic O, Grossi C, Alamanni F et al. Does EuroSCORE II perform better than its original versions? A multicentre validation study. *Eur Heart J* 2013; 34(1):22–9.

77. Ad N, Holmes SD, Patel J, Pritchard G, Shuman DJ, Halpin L. Comparison of EuroSCORE II, Original EuroSCORE, and The Society of Thoracic Surgeons Risk Score in Cardiac Surgery Patients. *Ann Thorac Surg* 2016; 102(2):573–9.

78. 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. *Am J Cardiol* 2016; 118(10):1574–82

79. Velicki L, Cemerlic-Adjic N, Pavlovic K, Mihajlovic BB, Bankovic D, Mihajlovic B, Fabri M. Clinical performance of the EuroSCORE II compared with the previous EuroSCORE iterations. *Thorac Cardiovasc Surg*. 2014 Jun;62(4):288-97

80. Qadir I, Alamzaib SM, Ahmad M, Perveen S, Sharif H. EuroSCORE vs. EuroSCORE II vs. Society of Thoracic Surgeons risk algorithm. *Asian Cardiovasc Thorac Ann*. 2014 Feb;22(2):165-71.

81. Bagnall NM, Faiz O, Darzi A, Athanasiou T. What is the utility of preoperative frailty assessment for risk stratification in cardiac surgery? *Interact Cardiovasc Thorac Surg* 2013; 17(2):398–402.

82. Barili F, Di Gregorio O, Capo A, Ardemagni E, Rosato F, Argenziano M et al. Aortic valve replacement: reliability of EuroSCORE in predicting early outcomes. *Int J Cardiol* 2010; 144(2):343–5.
83. Barili F, Pacini D, Capo A, Ardemagni E, Pellicciari G, Zanobini M et al. Reliability of new scores in predicting perioperative mortality after isolated aortic valve surgery: a comparison with the society of thoracic surgeons score and logistic EuroSCORE. *Ann Thorac Surg* 2013; 95(5):1539–44.
84. Choong CK, Sergeant P, Nashef SAM, Smith JA, Bridgewater B. The EuroSCORE risk stratification system in the current era: how accurate is it and what should be done if it is inaccurate? *Eur J Cardiothorac Surg* 2009; 35(1):59–61.
85. Roques F, Nashef SA, Michel P. Risk factors for early mortality after valve surgery in Europe in the 1990s: lessons from the EuroSCORE pilot program. *J Heart Valve Dis* 2001; 10(5):572-7; discussion 577-8.
86. Wang C, Li X, Lu F-L, Xu J-B, Tang H, Han L et al. Comparison of six risk scores for in-hospital mortality in Chinese patients undergoing heart valve surgery. *Heart Lung Circ* 2013; 22(8):612–7.
87. Poullis M, Fabri B, Pullan M, Chalmers J. Sampling time error in EuroSCORE II. *Interact Cardiovasc Thorac Surg* 2012; 14(5):640–1.
88. Poullis M, Pullan M, Chalmers J, Mediratta N. The validity of the original EuroSCORE and EuroSCORE II in patients over the age of seventy. *Interact Cardiovasc Thorac Surg* 2015; 20(2):172–7.
89. O'Boyle F, Mediratta N, Fabri B, Pullan M, Chalmers J, McShane J et al. Long-term survival after coronary artery bypass surgery stratified by EuroSCORE. *Eur J Cardiothorac Surg* 2012; 42(1):101-6; discussion 106-7
90. Helfried Moosbrugger, Augustin Kelava: *Testtheorie und Fragebogenkonstruktion*. Springer, 2011, ISBN 978-3-642-20072-4, S. 186–187
91. Biancari F, Vasques F, Mikkola R, Martin M, Lahtinen J, Heikkinen J. Validation of EuroSCORE II in patients undergoing coronary artery bypass surgery. *Ann Thorac Surg*. 2012 Jun;93(6):1930-5.

92. Chalmers J1, Pullan M, Fabri B, McShane J, Shaw M, Mediratta N, Poullis M. Validation of EuroSCORE II in a modern cohort of patients undergoing cardiac surgery. *Eur J Cardiothorac Surg*. 2013 Apr;43(4):688-94
93. Mehmet Kalender, Taylan Adademir, Mehmet Tasar, Ata Niyazi Ecevit, Okay Guven Karaca, Salih Salihi, Fuat Buyukbayrak, Mehmet Ozkokeli. Validation of EuroSCORE II risk model for coronary artery bypass surgery in high-risk patients. *Kardiochir Torakochirurgia Pol*. 2014 Sep; 11(3): 252–256.
94. Kirmani BH, Mazhar K, Fabri BM, Pullan DM. Comparison of the EuroSCORE II and Society of Thoracic Surgeons 2008 risk tools. *Eur J Cardiothorac Surg*. 2013 Dec;44(6):999-1005.
94. Marcus D. Ruopp, Neil J. Perkins, Brian W. Whitcomb, and Enrique F. Schisterman. Youden Index and Optimal Cut-Point Estimated from Observations Affected by a Lower Limit of Detection. *Biom J*. 2008 Jun; 50(3): 419–430.

Appendix 1 – Ethikvotum

Ethikkommission



Medizinische Universität Graz

Auenbruggerplatz 2, A-8036 Graz

ethikkommission@medunigraz.at

Tel.: +43 / 316 / 385-13920, Fax: -14348

VOTUM gültig bis 22.01.2017

EK-Nummer: 28-061 ex 15/16
Studientitel: Cardiosurgical Risk Assessment: How does EuroSCORE II alter clinical results compared to EuroSCORE representing an appropriate Method for quality management in cardiac surgery - a retrospective study
Prüfer: Ao. Univ.-Prof. Dr. med. univ. Igor Knez
Medizinische Universität Graz
Sponsor: Klinische Abteilung für Herzchirurgie
Ansprechpartner: Ao. Univ.-Prof. Dr. med. univ. Igor Knez, 8036 Graz, Auenbruggerplatz 29
CRO: -
Antragsteller: Klinische Abteilung für Herzchirurgie
Ansprechpartner: Ailisa Strugger

Die o.a. Studie wurde von der Ethikkommission erstmals im 'expedited Review' am 21.10.2015 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:

Dokumente eingegangen am 12.10.2015, begutachtet im 'expedited Review' am 21.10.2015

✓ Cover Letter Ailisa Strugger Anschreiben 1 01	25.06.2015
✓ Antragsformular ECS	12.10.2015
Originalprotokoll Prof_retrop (1) 1.0	07.10.2015
Sonstiges: Konzeptformular fertig 01	05.06.2015
✓ Sonstiges: Antrag 1.1	10.09.2015

Dokumente eingegangen am 03.12.2015 (in der nächsten Begutachtung mitbegutachtet)

✓ Cover Letter	17.11.2015
✓ Originalprotokoll 1.1	17.11.2015
✓ Sonstiges: Konzeptformular 1.1	17.11.2015

Dokumente eingegangen am 14.01.2016, begutachtet im 'expedited Review' am 22.01.2016

✓ Antragsformular ECS unterschrieben	07.12.2015
--------------------------------------	------------

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:

EK-Nummer: 28-061 ex 15/16

Votum (22.01.2016)

Seite 1 von 2

Medizinische Universität Graz, Auenbruggerplatz 2, A-8036 Graz, www.medunigraz.at

Rechtstext: Juristische Person öffentlichen Rechts gem. Universitätsgesetz 2002, Inkubation: Mitgliedschaft der Universität und www.medunigraz.at, CVR/Nr.: 310 0484, UID: ATU 378 171 78, Bankverbindung: Bank Austria Creditanstalt (BLZ 12000) Konto-Nr.: 500 048 400 04, Raiffeisen Landesbank Steiermark (BLZ 20000) Konto-Nr.: 490 10.

Abbildung 26: Ethikvotum 1

Appendix 1 – Ethikvotum

- Abweichungen vom Protokoll aus Sicherheitsgründen oder Protokolländerungen
- 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, 22. Jänner 2016


Univ.-Prof. Dr. Josef Haas
Vorsitzender


Univ.-Prof. Dr. Hermann Toplak
Stv. Vorsitzender

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

Abbildung 27: Ethikvotum 2

Appendix 2 – Benutzeroberfläche EuroSCORE II

Patient related factors		
Age ¹ (years)	<input type="text" value="0"/>	<input type="text" value="0"/>
Gender	<input type="text" value="select"/>	<input type="text" value="0"/>
Renal impairment ² <i>See calculator below for creatinine clearance</i>	<input type="text" value="normal (CC >85ml/min)"/>	<input type="text" value="0"/>
Extracardiac arteriopathy ³	<input type="text" value="no"/>	<input type="text" value="0"/>
Poor mobility ⁴	<input type="text" value="no"/>	<input type="text" value="0"/>
Previous cardiac surgery	<input type="text" value="no"/>	<input type="text" value="0"/>
Chronic lung disease ⁵	<input type="text" value="no"/>	<input type="text" value="0"/>
Active endocarditis ⁶	<input type="text" value="no"/>	<input type="text" value="0"/>
Critical preoperative state ⁷	<input type="text" value="no"/>	<input type="text" value="0"/>
Diabetes on insulin	<input type="text" value="no"/>	<input type="text" value="0"/>
EuroSCORE II <input type="text" value="EuroSCORE II"/>		<input type="text" value="0"/>
Cardiac related factors		
NYHA	<input type="text" value="select"/>	<input type="text" value="0"/>
CCS class 4 angina ⁸	<input type="text" value="no"/>	<input type="text" value="0"/>
LV function	<input type="text" value="select"/>	<input type="text" value="0"/>
Recent MI ⁹	<input type="text" value="no"/>	<input type="text" value="0"/>
Pulmonary hypertension ¹⁰	<input type="text" value="no"/>	<input type="text" value="0"/>
Operation related factors		
Urgency ¹¹	<input type="text" value="elective"/>	<input type="text" value="0"/>
Weight of the intervention ¹²	<input type="text" value="isolated CABG"/>	<input type="text" value="0"/>
Surgery on thoracic aorta	<input type="text" value="no"/>	<input type="text" value="0"/>

Abbildung 28: Maske im Onlinekalkulator für den EuroSCORE II

Appendix 3 – Funnelplots nach den Beobachtungsjahren

Operationen

	Summe	männlich	weiblich
Summe	349	263	86
lebend	331	253	78
tot	15	9	6
unbekannt	3	1	2

Ergebnisse

OP-Anzahl	346	262	84
Verstorbene	15	9	6
Verstorbene (% von Ges)	4,34%	3,44%	7,14%
Durchschn. Score	7,32%	7,00%	8,32%
Ergebnis	0,59	0,49	0,86

Abbildung 29: Funnelplot logistischer EuroSCORE 2012

Operationen

	Summe	männlich	weiblich
Summe	349	263	86
lebend	331	253	78
tot	15	9	6
unbekannt	3	1	2

Ergebnisse

OP-Anzahl	346	262	84
Verstorbene	15	9	6
Verstorbene (% von Ges)	4,34%	3,44%	7,14%
Durchschn. Score	4,45%	4,32%	4,85%
Ergebnis	0,97	0,80	1,47

Abbildung 30: Funnelplot EuroSCORE II 2012

Appendix 3 – Funnelplots nach den Beobachtungsjahren

Operationen

	Summe	männlich	weiblich
Summe	335	265	70
lebend	324	255	69
tot	11	10	1

Ergebnisse

OP-Anzahl	335	265	70
Verstorbene	11	10	1
Verstorbene (% von Ges)	3,28%	3,77%	1,43%
Durchschn. Score	6,63%	6,40%	7,53%
Ergebnis	0,49	0,59	0,19

Abbildung 31: Funnelplot logistischer EuroSCORE 2013

Operationen

	Summe	männlich	weiblich
Summe	335	265	70
lebend	324	255	69
tot	11	10	1

Ergebnisse

OP-Anzahl	335	265	70
Verstorbene	11	10	1
Verstorbene (% von Ges)	3,28%	3,77%	1,43%
Durchschn. Score	4,44%	4,26%	5,13%
Ergebnis	0,74	0,89	0,28

Abbildung 32: Funnelplot EuroSCORE II 2013

Appendix 3 – Funnelplots nach den Beobachtungsjahren

Operationen

	Summe	männlich	weiblich
Summe	397	314	83
lebend	375	302	73
tot	15	8	7
unbekannt	7	4	3

Ergebnisse

OP-Anzahl	390	310	80
Verstorbene	15	8	7
Verstorbene (% von Ges)	3,85%	2,58%	8,75%
Durchschn. Score	7,63%	6,91%	10,43%
Ergebnis	0,50	0,37	0,84

Abbildung 33: Funnelplot logistischer EuroSCORE 2014

Operationen

	Summe	männlich	weiblich
Summe	397	314	83
lebend	375	302	73
tot	15	8	7
unbekannt	7	4	3

Ergebnisse

OP-Anzahl	390	310	80
Verstorbene	15	8	7
Verstorbene (% von Ges)	3,85%	2,58%	8,75%
Durchschn. Score	5,35%	4,82%	7,41%
Ergebnis	0,72	0,54	1,18

Abbildung 34: Funnelplot EuroSCORE II 2014

Appendix 3 – Funnelplots nach den Beobachtungsjahren

Operationen

	Summe	männlich	weiblich
Summe	330	280	50
lebend	315	266	49
tot	15	14	1

Ergebnisse

OP-Anzahl	330	280	50
Verstorbene	15	14	1
Verstorbene (% von Ges	4,55%	5,00%	2,00%
Durchschn. Score	6,70%	6,58%	7,37%
Ergebnis	0,68	0,76	0,27

Abbildung 35: Funnelplot logistischer EuroSCORE 2015

Operationen

	Summe	männlich	weiblich
Summe	330	280	50
lebend	315	266	49
tot	15	14	1

Ergebnisse

OP-Anzahl	330	280	50
Verstorbene	15	14	1
Verstorbene (% von Ges	4,55%	5,00%	2,00%
Durchschn. Score	4,21%	4,24%	4,03%
Ergebnis	1,08	1,18	0,50

Abbildung 36: Funnelplot EuroSCORE II 2015

Appendix 4

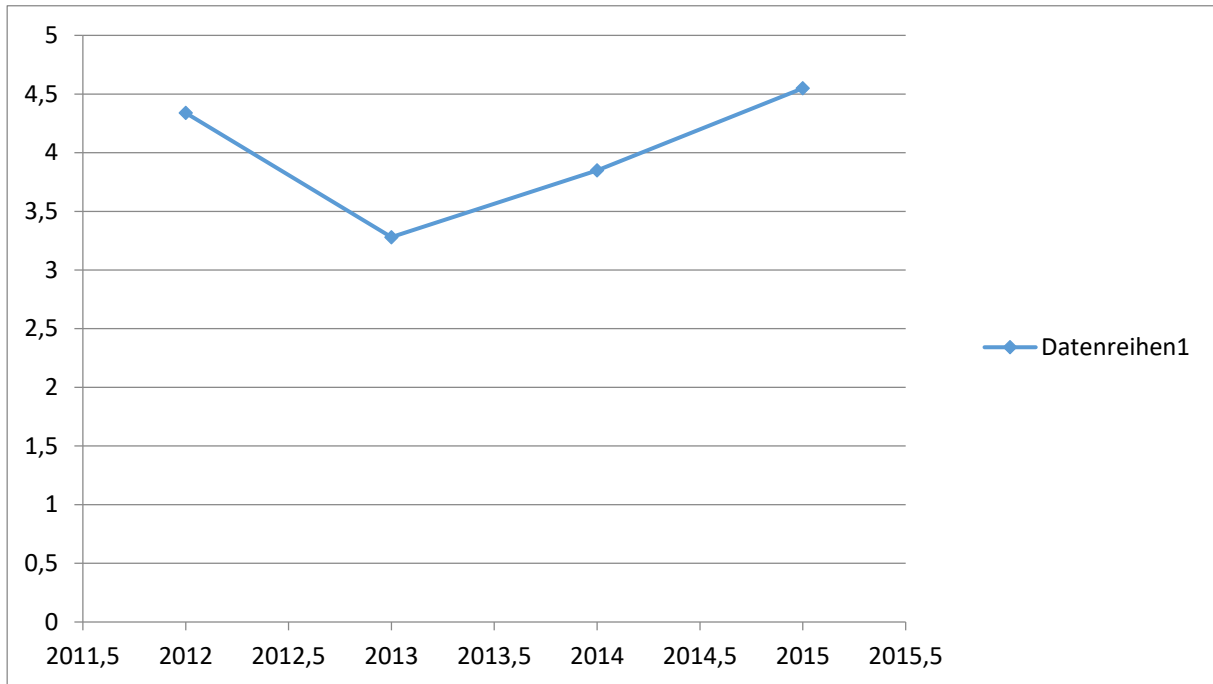


Abbildung 37: CABG 2012-2015 relative Mortalität in %

Appendix 4

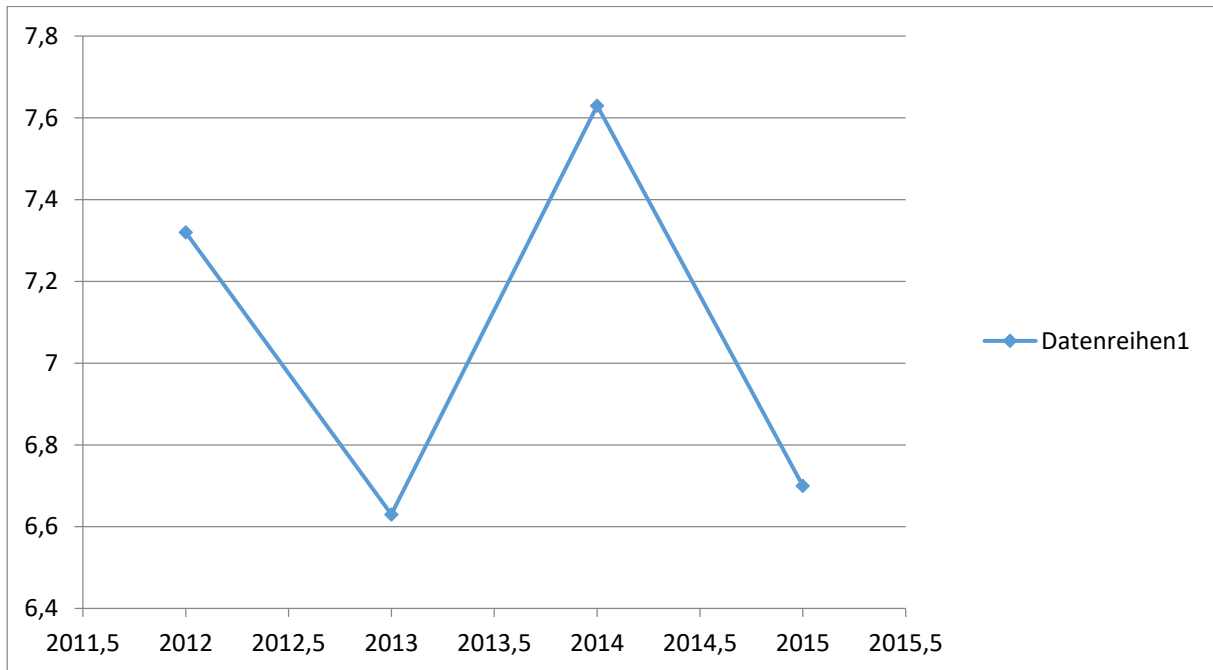


Abbildung 38: CABG 2012-2015 durchschnittlicher logistischer EuroSCORE in %

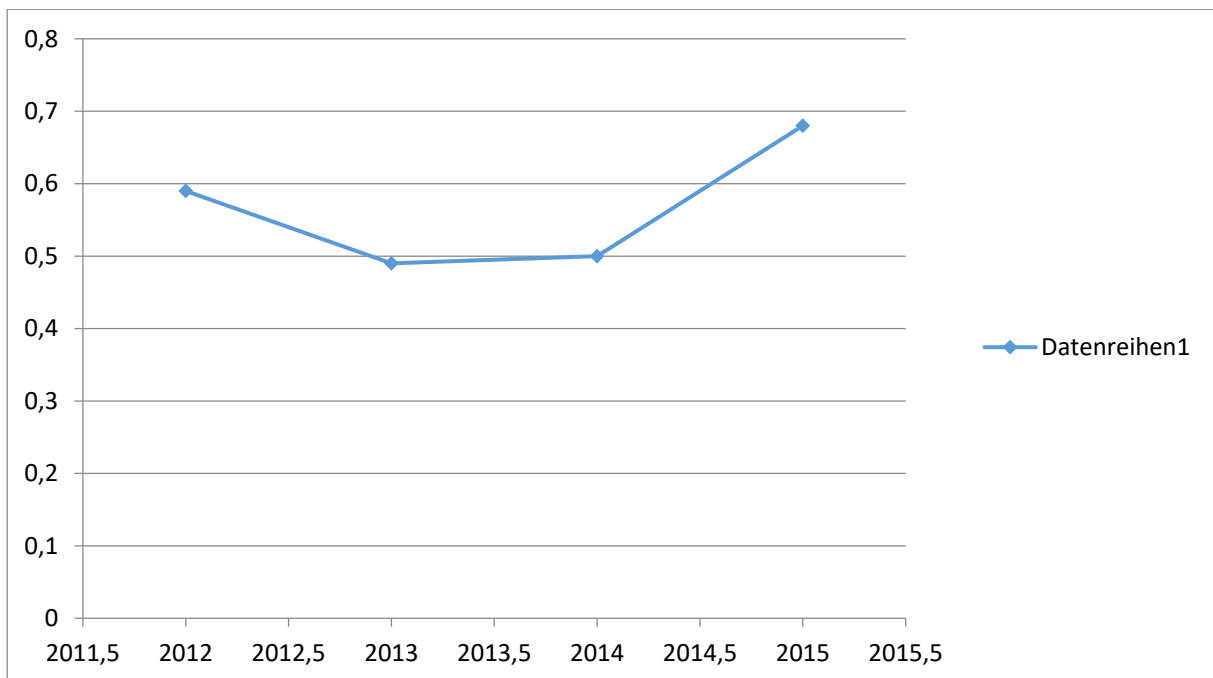


Abbildung 39: CABG 2012-2015 Performance Ratio logistischer EuroSCORE

Appendix 4

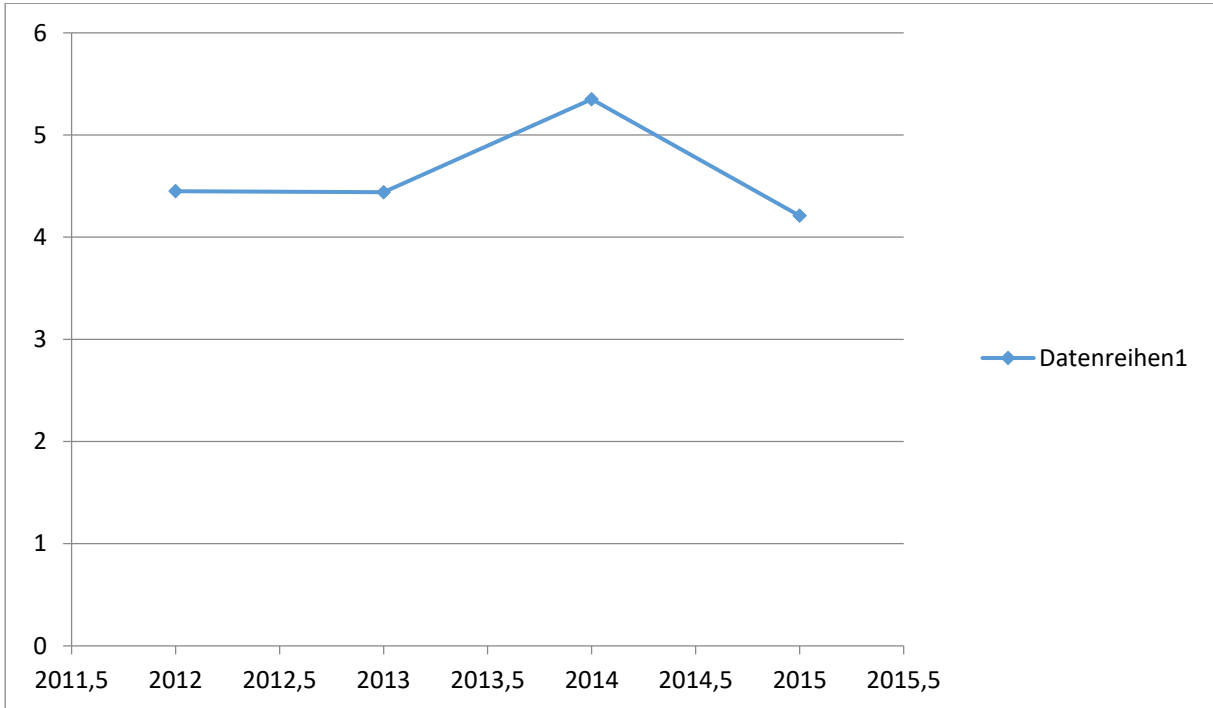


Abbildung 40: CABG 2012-2015 durchschnittlicher EuroSCORE II in %

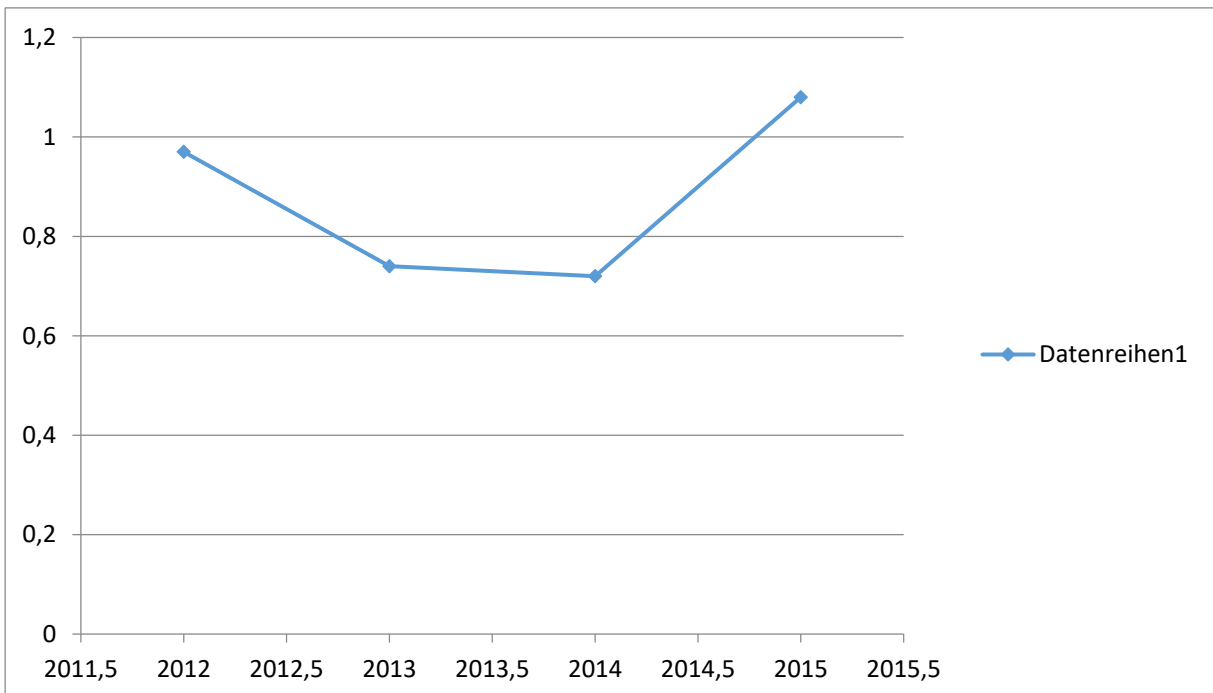


Abbildung 41: CABG 2012-2015 Performance Ratio EuroSCORE II