

Aus der chirurgischen Klinik
der Medizinischen Fakultät Mannheim
(Direktor: Prof. Dr. med. Christoph Reißfelder)

„Validation of a novel score for urgent bariatric and metabolic surgery“

Inauguraldissertation
zur Erlangung des medizinischen Doktorgrades
der
Medizinischen Fakultät Mannheim
der Ruprecht-Karls-Universität
zu
Heidelberg

vorgelegt von
Ramie Kamal Mohammed Alaeddin

aus
Leipzig

2025

Dekan: Prof. Dr. med. Sergij Goerdt
Referentin: Frau Priv.-Doz. Dr. med. Susanne Blank

“Corpulence is not only a disease itself, but the harbinger of others”
Hippocrates
Griechischer Arzt (460 BC-370 BC)

INHALTSVERZEICHNIS

Abkürzungsverzeichnis	1
1 Einleitung	2
1.1 Gütekriterien diagnostischer Verfahren	3
1.1.1 Objektivität	3
1.1.2 Reliabilität	3
1.1.3 Validität	4
1.2 Definition, Klassifikation und Epidemiologie von Adipositas	4
1.3 Komorbiditäten bei Adipositas	5
1.4 Therapiemöglichkeiten	7
1.5 Scores in der Bariatrie	8
2 Ziel dieser Arbeit	11
3 Material und Methoden	12
3.1 Patientenkollektiv	14
3.2 Die Rater	14
3.3 Statistische Methoden	14
3.4 Dokumentation	16
4 Ergebnisse	17
4.1 Häufigkeitsverteilung der Scoreergebnisse für die Kategorie "age"	18
4.2 Häufigkeitsverteilung der Scoreergebnisse für die Kategorie "gender"	19
4.3 Häufigkeitsverteilung der Scoreergebnisse für die Kategorie "BMI"	20
4.4 Häufigkeitsverteilung der Scoreergebnisse für die Kategorie "functional status"	21
4.5 Häufigkeitsverteilung der Scoreergebnisse für die Kategorie "physical status"	22
4.6 Häufigkeitsverteilung der Scoreergebnisse für die Kategorie "pulmonary status"	23
4.7 Häufigkeitsverteilung der Scoreergebnisse für die Kategorie "diabetes status"	24
4.8 Ergebnisse des Gesamtscore	25
4.9 Varianzhomogenität	26
4.10 Interklassenkorrelationskoeffizient	26
4.11 Korrelationsanalyse nach Pearson	29
5 Diskussion	31
5.1 Limitierungen und Ausblick	34
6 Zusammenfassung	35
7 Literaturverzeichnis	37
8 Abbildungsverzeichnis	42
9 Tabellenverzeichnis	43
10 Lebenslauf	45
11 Danksagung	47
12 Appendix	48

ABKÜRZUNGSVERZEICHNIS

**	<i>Correlation is significant at the 0.01 level</i>
AHA	<i>American Heart Association</i>
APACHE II	<i>Acute physiology and chronic health evaluation</i>
APGAR	<i>Appearance, puls, grimace, activity, respiration</i>
ASA	<i>American Society of anaesthesiologists</i>
BMI	<i>Body Mass Index</i>
COPD	<i>chronic obstructive pulmonary disease</i>
Ggf.	gegebenenfalls
ICC	Interclasscorrelation-coefficient
IRS	<i>Internal Rating Scale</i>
KOSC	<i>Kings Obesity Staging Criteria</i>
NAFL	nichtalkoholische Fettleber
NASH	nichtalkoholische Steatosis hepatis
OHV	<i>obesity hypoventilation syndrome</i>
OSAS	obstruktives Schlafapnoesyndrome
RYGB	<i>Roux-en-Y Gastric bypass</i>

1 EINLEITUNG

Scores (englisch, zu deutsch Punktsystem oder Skala) in der Medizin haben sich in den letzten Jahren zu etablierten Messsystemen entwickelt. Sie unterstützen medizinisches Personal in der Triage, Diagnostik, Entscheidungsfindung zur Therapie, zur Klassifikation einer Erkrankung, zur Vorhersage eines oder mehrerer Outcomes, Erfassung der Lebensqualität und im Rahmen der Qualitätskontrolle. Anwendung finden Scores vor allem in der Notfall- und Intensivmedizin, da hier kritische kranke Patienten behandelt werden und Diagnostik- sowie Therapieentscheidungen oft an Standardisierungen gekoppelt sind (Oprita et al., 2014). Durch diese standardisierten Bewertungsinstrumente und Abläufe wird die Qualität und Behandlung der Patienten maßgeblich gesteigert, da sie eine konsistente und evidenzbasierte Entscheidungsfindung ermöglichen.

Etabliert haben sich primär Punkte-Scores aus physiologischen Parametern aufgrund ihrer objektiven Interpretationsmöglichkeit und leichten Verständlichkeit (Schuster and Dick, 1994). Beispiele hierfür sind die Glasgow-Coma-Scale (Teasdale and Jennett, 1974) oder der APACHE-II (Knaus et al., 1985) sowie als krankheitsspezifische Scores der Injury Severity Score (Baker et al., 1974). Verlaufsscores wie der APGAR-Score sind in der pädiatrischen Intensivmedizin verpflichtender Standard (Apgar, 1953).

Nach Bouch und Thompson (2008) ist der ideale Score durch folgende Kriterien definiert (Bouch and Thompson, 2008):

1. Basiert auf leicht/routinemäßig erfassbaren Werten
2. Gut kalibriert
3. Ein hohes Maß an Unterscheidungskraft
4. Anwendbar auf alle Patientenpopulationen
5. Kann international verwendet werden
6. Die Fähigkeit, den funktionellen Zustand oder die Lebensqualität nach einer Behandlung vorherzusagen (Bouch and Thompson, 2008)

Wie auch in physikalischen Messverfahren nehmen systematische und zufällige Fehler und Störeffekte Einfluss auf die Ergebnisse der Scores. Umso wichtiger ist es objektive, reliable und valide Scoringsysteme einzusetzen um Störvariablen kontrollieren zu können und möglichst genaue Ergebnisse zu erzielen.

Im Folgenden werden die Gütekriterien von Messverfahren erläutert. Darüber hinaus wird der aktuelle Stand der Adipositaspandemie sowie der damit verbundenen

Komorbiditäten, Therapiemöglichkeiten und Scores in der Bariatrie umfassend dargestellt.

1.1 Gütekriterien diagnostischer Verfahren

Die allgemeinen Gütekriterien, welche in der quantitativen Forschung Anwendung finden und auch für diagnostische Messverfahren gelten sind Validität, Reliabilität und Objektivität. Diese Hauptgütekriterien wissenschaftlicher Mess- und Beurteilungsverfahren sichern die Qualität eines Scores. Zu den Nebengütekriterien gehören Fairness, Ökonomie, Nützlichkeit, Unverfälschbarkeit, Zumutbarkeit und Normierung. Nebengütekriterien sind nicht zwingend notwendig (Krampen, 2022).

1.1.1 Objektivität

Die Objektivität lässt sich in drei Teilgebiete unterteilen. Die Durchführungsobjektivität gibt an, ob ein Testergebnis von der durchführenden Person unabhängig ist. Die Auswertungsobjektivität gibt an, ob ein Test unabhängig von einer Person ausgewertet werden kann (standardisierte Auswertung und Abläufe). Die Interpretationsobjektivität zeigt ob ein Testergebnis auf Basis von standardisierten Normen unabhängig vom Anwender interpretiert werden kann. Objektivität wird nicht statistisch berechnet (Bortz, 2016). Holle 1995 beschreibt die Objektivität nicht als eigenes Gütekriterium und ordnet sie der Reliabilität zu, Bortz und Döring 2016 beschreiben sie im Gegenzug als eigenständiges Gütekriterium. In der Literatur hat sich die Objektivität als eigenständiges Gütekriterium etabliert. Objektivität bildet die Basis eines jeden Testinstruments.

1.1.2 Reliabilität

Reliabilität bezeichnet die Eigenschaft einer Messmethode, reproduzierbare, also identische oder nahezu ähnliche Messergebnisse zu liefern, wenn die Messung unter den exakt gleichen Bedingungen durchgeführt wird (Holle, 1995). Die Reliabilität zeigt somit die Messgenauigkeit eines Messinstruments. Zum Nachweis der Reliabilität muss mindestens ein Reliabilitätskoeffizient statistisch geprüft werden (Bortz, 2016). Zur Testung der Stabilität eines Verfahrens wird die Testwiederholungsreliabilität geprüft. Die Paralleltestreliabilität und die Interraterreliabilität prüfen die Äquivalenz eines Instruments. Bei der Paralleltestreliabilität geht es darum nachzuweisen ob mit einem ähnlichen Konstrukt ähnlich zuverlässige Ergebnisse gemessen werden können. Die Interraterreliabilität zeigt ob ähnliche Ergebnisse zwischen mehreren Ratern bei einem Test zu erzielen sind. Die Konsistenz der Reliabilität wird über die

Testhalbierungsreliabilität oder die interne Konsistenz (Zusammenhang der Fragen) geprüft. Die interne Konsistenz zeigt inwieweit die einzelnen Items (Fragen) eines Kriteriums zusammenhängen (Bortz, 2016).

1.1.3 Validität

Die Validität, auch Gültigkeit, zeigt ob das Messinstrument in der Lage ist, das zu messen was es messen soll. Bei der Bewertung der Validität ist es erforderlich, dass ein objektiver Wert vorhanden ist, der idealerweise messbar ist. In den Naturwissenschaften, wo physikalische und chemische Messgrößen vorliegen, ist dies in der Regel unproblematisch (Heesen, 2005).

Bei der Testentwicklung muss mindestens eine von drei Validierungsmethoden geprüft werden. Unterformen der Validität sind die Inhalts-, die Kriteriums- und die Konstruktvalidität. Die Inhaltsvalidität auch Augenscheinvalidität genannt gibt an, ob das Messinstrument die inhaltlichen Relevanzaspekte vollständig und sinnvoll abbildet. Diese Validierung erfolgt argumentativ und kann nicht statistisch berechnet und dargestellt werden. Die Konstruktvalidität zeigt ob die Messungen des Messinstruments mit anderen bereits validierten Konstrukten (andere Messinstrumente oder Skalen) korreliert oder eben auch nicht. Zur Absicherung bzw. um die Trennschärfe einer Messkonstruktes einzuordnen, werden auch Konstrukte herangezogen die vermeidlich nicht korrelieren sollten (also eine Kontrastierung, um sicher zu gehen, dass etwas anderes als das ggf. ähnliche Konstrukt gemessen wird). Dies kann beispielsweise mit Regressions- und einer Faktorenanalysen durchgeführt werden. Die Kriteriumsvalidität gibt an, ob die Testwerte mit Außenkriterien positiv korrelieren. Zur statistischen Überprüfung werden hierfür Korrelationsanalysen verwendet. Die Validität wird als das maßgeblichste Gütekriterium im Bereich der Testtheorie gewertet (Bortz, 2016).

1.2 Definition, Klassifikation und Epidemiologie von Adipositas

Adipositas bezeichnet eine chronische Erkrankung durch pathologische Akkumulation des Körperfettgewebes begleitet durch hohe Anzahl an Komorbiditäten welche mit einer reduzierten Lebenserwartung assoziiert ist (Brix et al., 2023; Burki, 2021). Klassifiziert wird die Adipositas durch den BMI. Um den BMI zu berechnen wird das Körpergewicht einer Person in Kilogramm durch das Quadrat der Körpergröße (gemessen in Metern) dividiert.

Die BMI-Kategorien werden nach der Klassifikation der Weltgesundheitsorganisation wie folgt definiert:

Tab. 1 Gewichtsklassifikation nach (WHO et al., 2000)

Untergewicht	BMI < 18,5 kg/m ²
Normalgewicht	BMI 18,5 – < 25,0 kg/m ²
Übergewicht/ Pre- Adipositas	BMI 25 – < 29,9 kg/m ²
Adipositas Grad 1	BMI 30 – < 34,9 kg/m ²
Adipositas Grad 2	BMI 35 – < 39,9 kg/m ²
Adipositas Grad 3	BMI ≥ 40,0 kg/m ²

Bereits heute sind weltweit ca. 38 % der Weltbevölkerung übergewichtig oder adipös ($\text{BMI} > 25 \text{ kg/m}^2$). Laut dem World Obesity Atlas 2023 wird diese Zahl bis zum Jahr 2035 auf 51 % steigen. Es wird erwartet das, 18 % dieser Population einen BMI größer 30 kg/m² , 6 % werden einen BMI > 35 kg/m², und 2 % einen BMI > 40 kg/m² haben (WOA, 2023).

In Deutschland zeigt sich laut RKI das insgesamt 53,5 % der Bevölkerung (46,6 % der Frauen und 60,5 % der Männer) von Übergewicht (einschließlich Adipositas) betroffen sind. Bei 19,0 % der Erwachsenen liegt eine Adipositas vor (Schienkiewitz et al., 2022). Hieraus lässt sich ableiten, dass sich das Gesundheitswesen auf eine stetige Zunahme der stark übergewichtigen und multimorbid Patienten vorbereiten muss (Geiling, 2010; Manning, 2020). Dies verdeutlicht die Notwendigkeit lebenserhaltende bariatrische Operationen betroffenen Patienten zu ermöglichen und hierdurch eine adäquate Gewichtsreduktion und gesündere Lebensumstände zu ermöglichen.

1.3 Komorbiditäten bei Adipositas

Adipositas ist aufgrund von einer Vielzahl an schweren Begleiterkrankungen mit einem erhöhten Risiko für Morbidität und Mortalität vergesellschaftet (Di Giacinto et al., 2021). Das Mortalitätsrisiko ist bei Patienten mit Adipositas Grad III im Vergleich zu normalgewichtigen Patienten um 173 % erhöht (Hauner, 2014).

Adipositas gilt laut der AHA als Risikofaktor für kardiovaskuläre Erkrankungen wie arterielle Hypertonie, koronare Herzkrankheit, Herzinsuffizienz und plötzlichem Herzstillstand (Powell-Wiley et al., 2021). Ebenfalls findet sich eine enge Korrelation zwischen einem erhöhten BMI und einer früheren Herz-Kreislauf-Morbidität und

Mortalität (Khan et al., 2018). Es wurde gezeigt das alle fünf Einheiten eines BMI-Anstieges zu einem 16 % höherem Risiko führt einen plötzlichen Herzstillstand zu erleiden (Aune et al., 2018).

Bereits die *Framingham Heart Study* belegte ein erhöhtes Risiko für koronare Herzkrankheit, Schlaganfall, Herzinsuffizienz und kardiovaskulären Tod bei Adipositas unabhängig von den üblichen Risikofaktoren wie Alter, Geschlecht, Nikotinabusus, erhöhtem Cholesterin und Bluthochdruck (Hubert et al., 1983).

Ein weiterer Zusammenhang zeigt sich zwischen Adipositas und einer erhöhten Inzidenz respiratorischer Erkrankungen wie Asthma (1,5–1,9-fach) und obstruktive Schlafapnoe (40–90 %) (Brock, 2021). Spezifisch für Adipositas ist das Obesity Hypoventilationssyndrom (OHS). Ursache für dieses sind übermäßige zervikale, thorakale, abdominale Fettgewebsablagerungen. Die Prävalenz des OHS ist in den letzten Jahrzehnten weltweit gestiegen und betrifft schätzungsweise 10–20 % der Patienten mit schwerer Adipositas. Bei Patienten mit einem BMI über 50 kg/m² steigt die Prävalenz auf bis zu 50 % (Athayde et al., 2018).

Die enge Beziehung zwischen Adipositas und Diabetes Typ II haben den Neologismus „diabesity“ entstehen lassen. Es bedeutet, dass die meisten Patienten mit einem Typ-2 Diabetes auch schwer übergewichtig sind (Sims et al., 1973). Eine Studie vor 20 Jahren konnten bereits ein 3,5–4,6-fach höheres Risiko einer Type-2 Diabetes Erkrankung bei Patienten mit einem BMI größer 29,9 kg/m² im Vergleich zu normalgewichtigen Patienten nachweisen (Field et al., 2001).

Heute sind ca. 80–90 % der Diabetes Typ-2 Erkrankten übergewichtig (Hauner, 2014; Hossain et al., 2007).

Im Hinblick auf die Leber führt die Adipositas oft zu einer nicht-alkoholischen Fettlebererkrankung (NAFLD). Eine nicht-alkoholische Steatohepatitis (NASH), Leberzirrhose und das hepatozelluläre Karzinom können Folgen sein (Polyzos et al., 2019). Mehrere Studien konnten eine Linearität zwischen steigendem BMI und der Prävalenz für NAFLD nachweisen (Chang et al., 2016; Eguchi et al., 2012).

Ein erhöhtes Risiko findet sich auch für diverse Neoplasien wie das Mamma-, Gallblasen-, Ösophagus oder Nierenzellkarzinom (Pati et al., 2023). Der Bewegungsapparat wird durch Arthrosen multipler Gelenke oder Wirbelsäulensyndrome negativ beeinflusst. Polyzystisches Ovar-Syndrom und reduzierte Fertilität sind Auswirkungen auf das Reproduktionssystem (Brix et al., 2023).

Für nahezu all diese Komorbiditäten gilt jedoch das eine Gewichtsreduktion und somit Senkung des BMI und vor allem der Abbau der abdominalen Fettmasse teilweise zu einer Heilung bzw. reduzierten Ausprägung der Begleiterkrankung führt (Hauner, 2014).

1.4 Therapiemöglichkeiten

Wesentliche Ziele einer Adipositastherapie sind primär die Gewichtsreduktion durch Abbau der Körperfettmasse, die Stabilisierung des reduzierten Gewichts und die Heilung oder Milderung der Symptome der Begleiterkrankungen (Müller and Blüher, 2023). Eine Reduktion der Energiezufuhr durch Anpassung der Ernährung auf niedrigkalorisch ausgewogene Lebensmittel am Beispiel der mediterranen Diät oder im Rahmen einer *Low-carb* oder *Low-Fat* Diät, sportliche Betätigung im Rahmen der Möglichkeiten als auch entsprechende Aufklärung und Verhaltensanpassung bilden den Grundstein dieser Therapien (Clodi et al., 2023; Müller and Blüher, 2023). Sollte der gewünschte Gewichtsverlust trotzdem ausbleiben, stehen medikamentöse sowie chirurgische Therapieoptionen zur Verfügung. Diese werden nachfolgend beschrieben.

Die Indikation zur medikamentösen Therapie der Adipositas ist ab einem BMI von 30 kg/m² gegeben. Liegen Nebenerkrankungen vor so sinkt die Indikationsschwelle auf einen BMI von 27 kg/m² (Müller and Blüher, 2023).

Zur Gewichtsreduktion sind aktuell in Deutschland die Wirkstoffe Liraglutid, Semaglutid und Orlistat zugelassen. Semaglutid gilt mittlerweile als Mittel der ersten Wahl (DAG, 2023). Bis zu 20 % Gewichtsverlust können über einen Therapiezeitraum von 68 Wochen bei milden bis moderaten Nebenwirkungen erzielt werden. Wird die medikamentöse Therapie beendet und die Basistherapie nicht fortgeführt so können erneute Gewichtszunahmen bis zu 67 % beobachtet werden (Moiz et al., 2024).

Tirzepatid ist ein neues Medikament welches in Deutschland seit 2023 neben der Behandlung eines therapierefraktären Diabetes Typ 2 auch zur Adipositastherapie zugelassen ist. Eine amerikanische Studie konnte zur Therapie der Adipositas einen deutlichen Gewichtsverlust zwischen 15 % und 20 %, je nach applizierter Dosis, über einen Therapiezeitraum von 72 Wochen nachweisen (Jastreboff et al., 2022). Weitere außerhalb der EU erforschte Medikamente sind Retatrutide und Orforglipron (Moiz et al., 2024).

Bei ausbleibenden Gewichtsreduktionserfolgen durch eine medikamentöse Behandlung der angeführten Medikamente oder der Verhaltens- und Ernährungsanpassung bleibt als eine letzte Therapiemaßnahme der operative Eingriff. Eine Vorstufe zur intraabdominellen Operation bildet die endoskopische Applikation eines Magenballons. Angewendet wird dieser bei bestehender Inoperabilität zur Einleitung eines präoperativen Gewichtsverlustes um das anästhesiologische und technische perioperative Mortalitätsrisiko für eine definitive Operation zu senken. Dieses Stufenkonzept findet vor allem bei Patienten mit einem $BMI > 50 \text{ kg/m}^2$ und erheblichen Komorbiditäten Anwendung (DGAV, 2018).

Die bariatrische Operation ist ab einem BMI von 40 kg/m^2 indiziert. Liegen Begleiterkrankungen vor sollte diese ab einem BMI von 35 kg/m^2 durchgeführt werden (DGAV, 2018). Die meist angewendeten OP-Techniken mit vorliegender Evidenz sind die Sleeve-Gastrektomie und der Roux-en-Y Magenbypass (RYGB). Weitere Techniken, wie der Omegaloop oder die biliopankreatische Diversion werden nur mit 7,6 % aller Eingriffe weltweit durchgeführt. Die Magenband-OP wird aufgrund der hohen Spätkomplikationsrate und deutlich geringeren Effektivität nur noch zu 5 % weltweit angewendet (Welbourn et al., 2019).

Bei einer Magenschlauchoperation (Sleevegastrektomie) oder einem Roux-Y-Magenbypass beträgt der Gesamtgewichtsverlust nach 5 Jahren 18,8 % bzw. 25,5 %. Langzeitdaten über 10 Jahre zeigen einen Verlust des Übergewichts von 57–60 %. Die perioperative Sterblichkeit liegt bei 0,03–0,2 %, während das Risiko für schwerwiegende Komplikationen in den ersten 30 Tagen zwischen 0,8–5,6 % und 1,4–9,4 % liegt. Langfristig liegt die Re-Operationsrate zwischen 5–22,1 %. Die Re-Operationsrate des RYGB liegt hier höher. Eine erneute Gewichtszunahme nach 5 Jahren ist möglich, beträgt jedoch weniger als 5 % des Ausgangsgewichts (Arterburn et al., 2020). Bariatrische Operationen sind effektiver als die beste konservative Therapie. Heutzutage wird eine Kombinationstherapie zur effektiven Gewichtsreduktion angewendet (Müller and Blüher, 2023).

1.5 Scores in der Bariatrie

Eine Literaturrecherche zeigte, wie bereits in der S3 Leitlinie „Chirurgie der Adipositas und metabolischer Erkrankungen“ beschrieben, dass keine ausreichende Literatur hinsichtlich notfallmedizinischer Scores in der Bariatrie vorhanden sind (DGAV, 2018).

Obwohl Scores zur Therapieentscheidung für bariatrische Patienten bisher selten Anwendung finden, gibt es bisher ein paar wenige Scores, die auf diesem Gebiet bereits erprobt und validiert sind. Diese sollen im Folgenden kurz skizziert werden.

Der bis dato bekannteste Score auf diesem Gebiet ist das Edmonton Obesity Staging System (EOSS). Es wurde entwickelt um die Dringlichkeit einer Therapie der Adipositas und krankheitsbedingten Folge- und Nebenerkrankungen als auch der Mortalität darzustellen und zu klassifizieren. Das EOSS setzt sich zusammen aus einer Klasse und einer Stufe. Die Klassen sind definiert nach der offiziellen Adipositas Einteilung I-III siehe Tab. 1. Die Stufe wird eingeteilt in *mild, moderate, significant und severe* basierend auf den zu beurteilenden Merkmalen. Diese sind *obesity related riskfactors, physical symptoms, psychological symptoms und functional limitations* (Sharma and Kushner, 2009). Der Nutzen der EOSS wurde von Chiappetta et al. 2016 in einer prospektiven Beobachtungsstudie belegt. Es konnte festgestellt werden, dass es mit einem Anstieg der EOSS auch zu einem prozentualen Anstieg an postoperativen Komplikationen kam. Vor allem Patienten mit einem EOSS ≤ 3 haben ein erhöhtes Risiko für postoperative Komplikationen. Es wird empfohlen, dass adipösen Patienten mit einem EOSS = 2 die bariatrische Operation empfohlen wird um so mögliche postoperative Komplikationen zu reduzieren und das Fortschreiten von metabolisch assoziierten Erkrankungen zu unterbinden (Chiappetta et al., 2016). Auch Skulsky et al. 2020 konnten zeigen das Patienten mit einem EOSS = 3 eine Verdreifachung der Komplikationen im Vergleich zu Patienten mit einem EOSS = 2 nach einer laparoskopischen Roux-en-Y Magenbypass Operation hatten (Skulsky et al., 2020).

Die King's Obesity Staging Criteria (KOSC) wurden entwickelt um den ganzheitlichen Gesundheitszustand von adipösen Patienten darzustellen. Der Score besteht aus neun Kriterien: Atemwege, Body-Mass-Index (BMI), Herz-Kreislauf-Erkrankung, Diabetes, wirtschaftlicher Status, funktionelle Einschränkungen, Fertilität, wahrgenommener Gesundheitszustand (psychischer Zustand), Körperbild. In diesen Kategorien können Werte zwischen null und drei vergeben werden. Ein normaler Gesundheitszustand wird mit null bewertet, das Risiko einer Erkrankung mit eins, eine etablierte Erkrankung mit zwei und eine fortgeschrittene Erkrankung mit drei Punkten. Aasheim et al. (2011) führten eine Anwendbarkeitsstudie des Scores im klinischen Setting durch. Es wurden n = 144 Patienten vor und nach einer bariatrischen OP entlang der genannten Kriterien eingeschätzt und eingeordnet. Mithilfe der KOSC konnten die postoperativen gesundheitlichen Verbesserungen statistisch signifikant

($p < 0.001$) dargestellt werden. Der Interklassenkorrelationskoeffizient (ICC) der einzelnen Kategorien lag zwischen 0.62 und 0.92. Nur die Kategorie Körperbild war mit einem ICC 0.28 schlecht bewertet (Aasheim et al., 2011).

Ein weiterer Score, welcher in der bariatrischen Chirurgie entwickelt wurde, ist der IMS- Score (*individualized metabolic surgery*). Er unterteilt Patienten mit einem Typ-2 Diabetes in 3 Kategorien (*mild, moderate, severe*) basierend auf vier unabhängigen existierenden Werten: Dauer der Typ II Diabetes Erkrankung, Anzahl der Diabetesmedikamente, Insulinverwendung und Blutzuckerstabilität. Ziel ist es mit Hilfe des Scores das optimale operative Verfahren für den Patienten zu ermitteln umso die erfolgreiche Diabetes Typ II Remission zu erzielen (Saarinen et al., 2023).

Aus den dargestellten Ergebnissen zeigt sich, dass ein umfänglicher Notfallscore zur Versorgung von bariatrischen Patienten notwendig ist.

2 ZIEL DIESER ARBEIT

Ziel dieser Arbeit ist die evidenzbasierte Validierung eines von Stier et al. 2022 entwickelten Notfallscores für bariatrische Patienten. Der Hauptbestandteil wird die Reliabilitätsanalyse mit Evaluierung der Interraterreliabilität auf Basis des ICC bilden. Die Prüfung der Validität wird auf Basis der Kriteriumsvalidität mit einer Korrelationsanalyse nach Pearson (Pearson, 1896) durchgeführt. Die Inhaltsvalidität und Durchführungsobjektivität werden diskutiert.

3 MATERIAL UND METHODEN

Der Score oder auch *internal rating scale* (IRS) wurde im Rahmen einer Studie von Stier et al. (2022) entwickelt. Er diente innerhalb der Studie als Messinstrument als Einschlusskriterium der Probanden. Der Score soll die Dringlichkeit zur unmittelbaren stationären Aufnahme mit anschließender „*fast-track rescue weight reduction bridging therapy*“ und folgender bariatrischer Operation ermitteln. Der Score ist positiv (*cut-off*) ab einem Wert von 14 Punkten oder bei maximaler Punktzahl acht in den Kategorien „physical status“ oder „pulmonary status“.

Table 1
Internal Rating Scale: Necessity of immediate hospitalization with fast-track rescue weight reduction bridging therapy to obesity surgery is indicated from a total of 14 points upwards or a full score for physical or pulmonary status.

Individual status	Age	>40	1
		>50	2
Gender	Female		0
	Male		2
BMI	Obesity class III > 40 kg/m ²		2
	Obesity class III > 50 kg/m ²		3
	Obesity class III > 60 kg/m ²		4
	Obesity class III > 65 kg/m ²		5
	Obesity class III > 70 kg/m ²		6
	Obesity class III > 75 kg/m ²		8
Functional status	Independent		0
	Partially dependent		2
	Totally dependent		4
	Wheelchair-dependent		6
	Immobile		8
Physical status (modified according to ASA classification)	Normally healthy patient (according to ASA 1)		0
	Mild systemic disease (according to ASA 2)		2
	Severe systemic disease (according to ASA 3) including hypertension, dyslipidemia (well controlled under medication) and NAFLD		4
	Severe systemic disease with threat to life (according to ASA 4) including uncontrolled hypertension, dyslipidemia and NAFLD with severe hepatomegaly		6
	Severe systemic disease with threat to life and urgent need for weight loss surgery (ultima ratio) to improve physical condition (according to ASA 5)		8
Pulmonary status	Not breathless with physical activity (two flights of stairs)		0
	Breathless with physical activity (two flights of stairs)		2
	Confirmed sleep apnea		3
	Obesity hypoventilation syndrome		4
	Permanently dependent on oxygen supply		6
Type 2 diabetes status	Tracheotomy		8
	Oral antidiabetic drug(s): HbA1c < 7.5%		1
	Insulin-dependent: HbA1c < 7.5%		2
	Oral antidiabetic drug(s): HbA1c 7.5% – 10%		4
	Insulin-dependent: HbA1c 7.5% – 10%		6
	HbA1c > 10%		8

ASA, physical status classification system of the American Society of Anesthesiologists (ASA); NAFLD, non-alcoholic fatty liver disease.

Abb. 1 Der von Stier et al. (2022) entwickelte Score zur Darstellung der Dringlichkeit zur unmittelbaren stationären Aufnahme eines Patienten mit anschließender Durchführung einer „*fast-track rescue weight reduction bridging therapy*“ (Stier et al., 2022)

Der Score besteht aus fünf Hauptkategorien. Die erste Kategorie ist in drei weitere Unterkategorien unterteilt: Alter (*age*), Geschlecht (*gender*) und Body-Mass-Index (BMI). Bei Patienten über 40 Jahren wird ein Punkt vergeben, während bei einem Alter über 50 Jahren zwei Punkte vergeben werden. Für das männliche Geschlecht werden zwei Punkte und für das weibliche Geschlecht null Punkte vergeben. Der BMI wird in sechs Stufen unterteilt: Bei einem BMI über 40 kg/m² werden zwei Punkte vergeben, ab 50 kg/m² gibt es drei Punkte, ab 60 kg/m² vier Punkte, ab 65 kg/m² fünf Punkte, ab 70 kg/m² sechs Punkte und ab einem BMI von 75 kg/m² acht Punkte.

Die zweite Kategorie spiegelt den funktionellen Status (*functional status*) des Patienten wider. Für Unabhängigkeit werden null Punkte vergeben, für teilweise Abhängigkeit zwei Punkte. Komplette Abhängigkeit führt zu vier Punkten. Eine Rollstuhlmobilität wird mit sechs Punkten und Immobilität mit acht Punkten bewertet.

Die dritte Kategorie repräsentiert den physischen Zustand (*physical status*) des Patienten gemäß der ASA-Klassifikation. Für einen gesunden Patienten (ASA I) werden null Punkte vergeben. Patienten mit milden systemischen Erkrankungen (ASA II) erhalten zwei Punkte. Vier Punkte werden vergeben, wenn der Patient ernsthafte systemische Erkrankungen wie medikamentös gut kontrollierten Hypertonus und eine nicht-alkoholische Fettlebererkrankung (ASA III) hat. Eine ASA-IV-Klassifikation, die lebensbedrohliche Krankheiten wie unkontrollierten arteriellen Hypertonus umfasst, führt zu sechs Punkten. Im Falle einer akuten Lebensbedrohung im Rahmen systemischer Erkrankungen (ASA IV) werden acht Punkte vergeben.

Die folgende Kategorie erfasst den Lungenstatus (*pulmonary status*) des Patienten. Kann ein Patient zwei Etagen ohne Atemprobleme aufsteigen gibt es null Punkte. Treten Atemprobleme auf, werden zwei Punkte vergeben. Bei bestätigter Schlafapnoe gibt es drei Punkte. Sollte das Obesity hypoventilationsyndrom vorliegen gibt es vier Punkte. Ist der Patient auf dauerhafte zusätzliche Sauerstoffversorgung angewiesen werden sechs Punkte vergeben. Bei Tracheotomie acht Punkte.

Die abschließende Kategorie untersucht den Status der Diabeteserkrankung (*diabetes status*). Ein HbA1c-Wert von weniger als 7,5 % unter oraler Medikation wird mit einem Punkt bewertet, während eine Insulintherapie mit zwei Punkten bewertet wird. Ein HbA1c-Wert von über 7,5 % unter oraler Medikation ergibt vier Punkte, während eine Insulintherapie mit sechs Punkten bewertet wird. Liegt der HbA1c-Wert über 10 %, werden acht Punkte vergeben. Die zu erreichende Maximalpunktzahl liegt bei 44 Punkten über alle Kategorien.

3.1 Patientenkollektiv

Das Patientenkollektiv umfasst Daten von n = 142 Patienten, welche von April 2015–April 2017 über das Sana Klinikum Offenbach erhoben wurden. Zu diesen Daten gehören Alter, Geschlecht, BMI, HbA1c und Nebenerkrankungen. Teilweise wurden diese Daten für weitere Studien genutzt. Weitere Informationen sind unter der Nummer: NCT02616003 auf clinicaltrials.gov zu finden. Das dazugehörige Ethikvotum ist unter der Nummer FF 121/2015 registriert.

3.2 Die Rater

Insgesamt haben n = 5 Ärzte unterschiedlicher Qualifikation das Patientenkollektiv auf Basis des Scores bewertet. Ein Assistenzarzt für Chirurgie im sechsten Weiterbildungsjahr mit Schwerpunkt Orthopädie/Unfallchirurgie und ein weiterer Assistenzarzt im sechsten Weiterbildungsjahr für Chirurgie. Eine Sektionsleitung für Adipositaschirurgie als auch eine Chefärztin und ein Chefarzt für Adipositas- und metabolische Chirurgie, alle mit langjähriger wissenschaftlicher und operativer Erfahrung in der metabolischen Chirurgie. Die Zusammenstellung erfolgte randomisiert.

3.3 Statistische Methoden

Die Patientendaten wurden zunächst deskriptiv analysiert. Anschließend wurden die Häufigkeiten der Scores der einzelnen Kategorien und des Gesamtscores der Rater dargestellt. Unterschiede in der Häufigkeitsverteilung werden mit dem *Fischers Freeman Halton Exakt Test* dargestellt. Das Signifikanzniveau liegt bei $\alpha = 0,05$.

Die Reliabilität wurde mit dem Intraklassenkorrelationskoeffizient bestimmt. Um den ICC statistisch korrekt zu bestimmen musste das entsprechende Modell gewählt werden. Koo & Li (2016) fassen in ihrer Veröffentlichung zehn Modelle nach McGraw & Wong (1996) Convention respektive die sechs Modelle nach Shrout & Fleiss (1979) zusammen.

Die Auswahl des richtigen ICC-Modell basiert auf vier Fragen.

- (1) Haben alle Probanden dieselbe Gruppe von Ratern?
- (2) Haben wir zufällig ausgewählte Rater oder wurden diese bestimmt?

(3) Interessieren wir uns für die Zuverlässigkeit eines einzelnen Bewerters oder für den Durchschnittswert mehrerer Bewerter?

(4) Liegt unser Fokus auf Konsistenz oder absoluter Übereinstimmung?

Relevant in dieser Arbeit ist das „*Two-way random effects, absolute agreement, single measurements*“ Modell nach McGraw & Wong Convention (1996) oder ICC (2,1) nach Shrout and Fleiss (1979) (McGraw and Wong, 1996).

Die Rater sind zufällig auf Basis ihrer Qualifikation ausgewählt. Alle Rater vergeben einen Score für alle Probanden. Die absolute Übereinstimmung der Werte ist Gegenstand der Untersuchung. Dieses Modell eignet sich zur Bewertung klinischer Beurteilungsmethoden die für den routinemäßigen klinischen Einsatz durch beliebige Ärzte konzipiert sind. (Koo and Li, 2016).

Anschließend wurde die Varianzhomogenität zwischen den Ergebnissen der Rater mittels des Levene Test (Levene, 1960) bestimmt. Dies ist Voraussetzung für die korrekte Interpretation des ICC (Wirtz and Caspar, 2002). Die Nullhypothese des Levenetest gibt an ob die verglichenen Varianzen gleich sind. Durch ein nicht signifikantes Ergebnis kann die Nullhypothese nicht verworfen werden und es wird Varianzhomogenität angenommen. Ist das Ergebnis signifikant wird von Varianzheterogenität ausgegangen. Sollten signifikante Varianzunterschiede vorliegen, wird sich der ICC verringern und die Reliabilität wird somit unterschätzt (Wirtz and Caspar, 2002). Der Levenetest wird häufig auf Basis der Mittelwerte interpretiert. Das Signifikanzergebnis basierend auf dem Median gilt als robustere Messung und wird in dieser Arbeit verwendet (Derrick et al., 2018; Li et al., 2015; Sharma and Kibria, 2013).

Die Kriteriumsvalidität wird mit einer Korrelationsanalyse nach Pearson durchgeführt. Folgende Korrelationen wurden untersucht:

- Alter und Scores der Kategorie „age“
- Geschlecht und Scores der Kategorie „gender“
- BMI und Scores der Kategorie „BMI“
- Orthopädische Erkrankung und Scores der Kategorie „functional status“
- Pulmologische Erkrankungen und Scores der Kategorie „pulmonary status“
- HbA1c mit den Scores der Kategorie „diabetes status“

3.4 Dokumentation

Die Auswertung basiert auf einem in einer Exceltabelle dokumentierten Datensatz. Zunächst wurden die für diese Studie notwendigen Patientendaten auf Vollständigkeit geprüft und anschließend zusammengefasst. Der Score wurde in Microsoft Excel nachgebildet um die Bewertung für die Rater zu erleichtern. Anschließend wurden die Ergebnisse in die Software SPSS (IBM®) Version 29 übertragen und statistisch ausgewertet.

4 ERGEBNISSE

Insgesamt wurden n = 142 Patienten bewertet, davon waren 70 Männer (49,3 %) und 72 Frauen (50,7 %) in einem Alter von 16–72 Jahren (\bar{O} 41,9 Jahre). Der BMI lag zwischen 51,7–94,6 kg/m² (\bar{O} 68,38 kg/m²). HbA1c-Werte im Aufnahmeharbor lagen zwischen 3,3–11,3% (\bar{O} 6,35%). In Tabelle 2 finden sich gruppiert die Nebenerkrankungen der Patienten. Von diesen können eine oder mehrere pro Patienten zutreffen.

Tab. 2 Nebenerkrankungen des Patientenkollektivs

Krankheiten in Gruppen	Anzahl Fälle	in %
Lungenerkrankungen (COPD, Asthma, OHV, OSAS)	61	42,9
Tracheotomie	3	2,12
Pulmonale/Arterielle Hypertonie	65	45,7
Herzinsuffizienz	8	5,6
Rhythmusstörungen	7	4,9
Koagulopathien/Thrombosen	14	9,8
Scheitern von vorherigenen Devices	14	9,8
Infarkte (zerebral/ cardial)	5	3,5
Orthopädische Erkrankungen/ Immobilität	45	31,6
Endokrine Störungen	76	53,5
Psychologische Erkrankungen	30	21,1
NAFL/NASH	16	11,2

COPD, chronic obstructive pulmonary disease; OHV, obesity hypoventilation syndrome; OSAS obstruktives Schlafapnoesyndrome; NAFL, nichtalkoholische Fettleber; NASH, nichtalkoholische Steatosis hepatis

4.1 Häufigkeitsverteilung der Scoreergebnisse für die Kategorie „age“

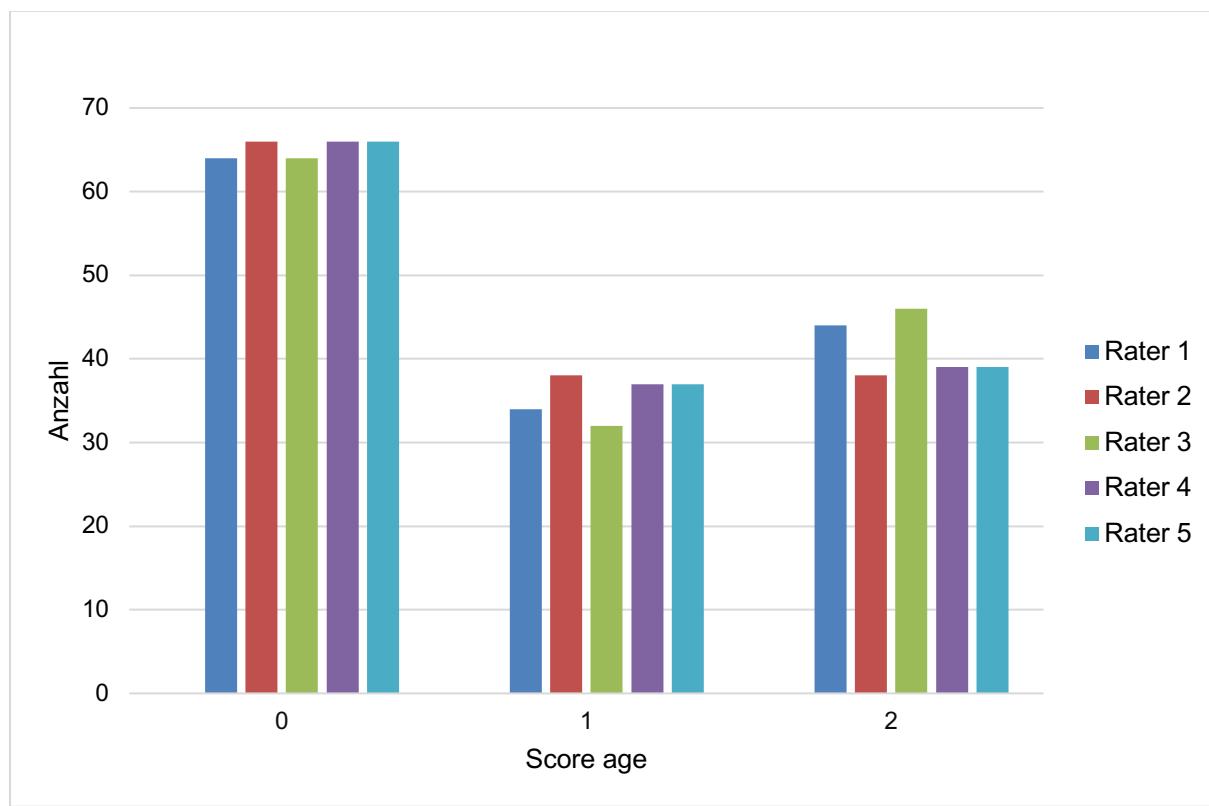


Abb. 2 Häufigkeitsverteilung der Scores für die Kategorie "age"

In der Kategorie „age“ wurde von jedem Rater jede mögliche Punktzahl vergeben. Die Verteilung der Werte zeigt keine signifikanten Differenzen ($p = 0.981$).

Tab. 3 Häufigkeitsverteilung der Scoreergebnisse für die Kategorie „age“

Score in Punkten	Rater 1	Rater 2	Rater 3	Rater 4	Rater 5
0	64	66	64	66	66
1	34	38	32	37	37
2	44	38	46	39	39
Total n =	142				142

4.2 Häufigkeitsverteilung der Scoreergebnisse für die Kategorie „gender“

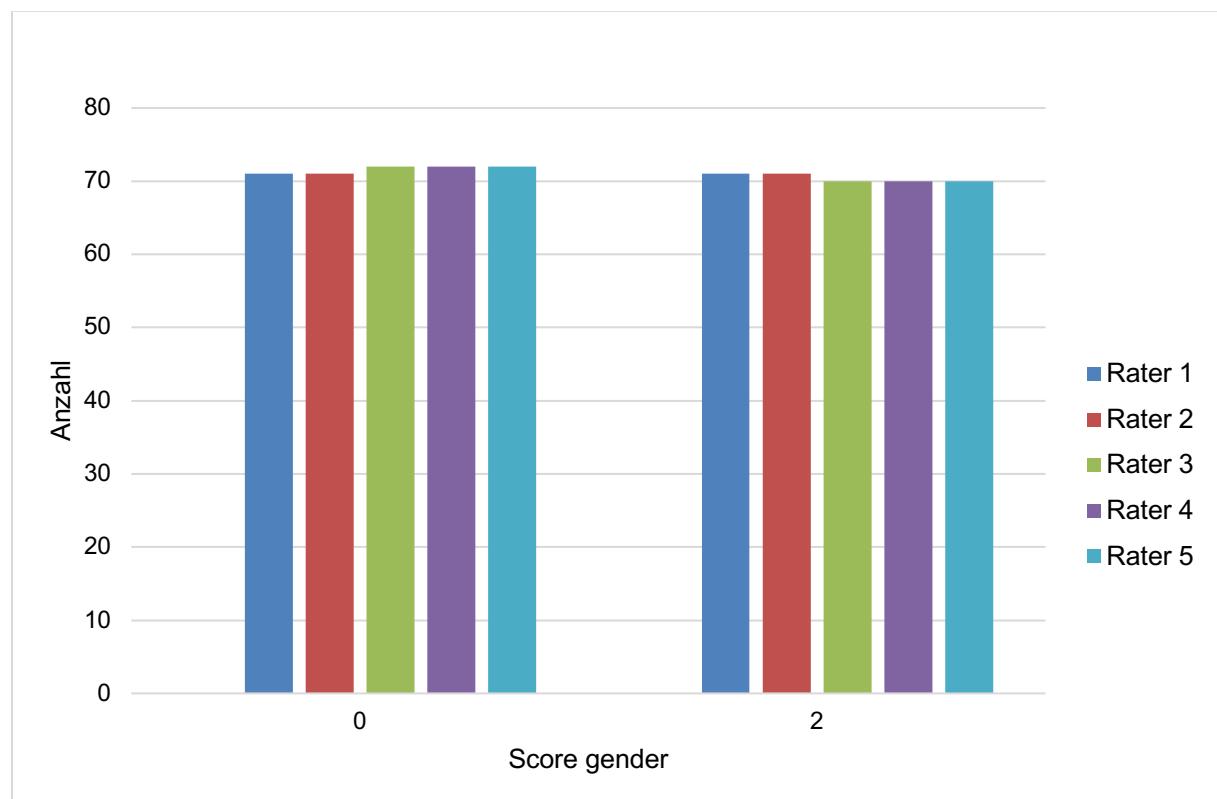


Abb. 3 Häufigkeitsverteilung der Scores für die Kategorie "gender"

Die Ergebnisse in dieser Kategorie sind statistisch gleich verteilt ($p = 1.0$).

Tab. 4 Häufigkeitsverteilung der Scoreergebnisse für die Kategorie "gender"

Score in Punkten	Rater 1	Rater 2	Rater 3	Rater 4	Rater 5
0	71	71	72	72	72
2	71	71	70	70	70
Total n =	142	142	142	142	142

4.3 Häufigkeitsverteilung der Scoreergebnisse für die Kategorie „BMI“

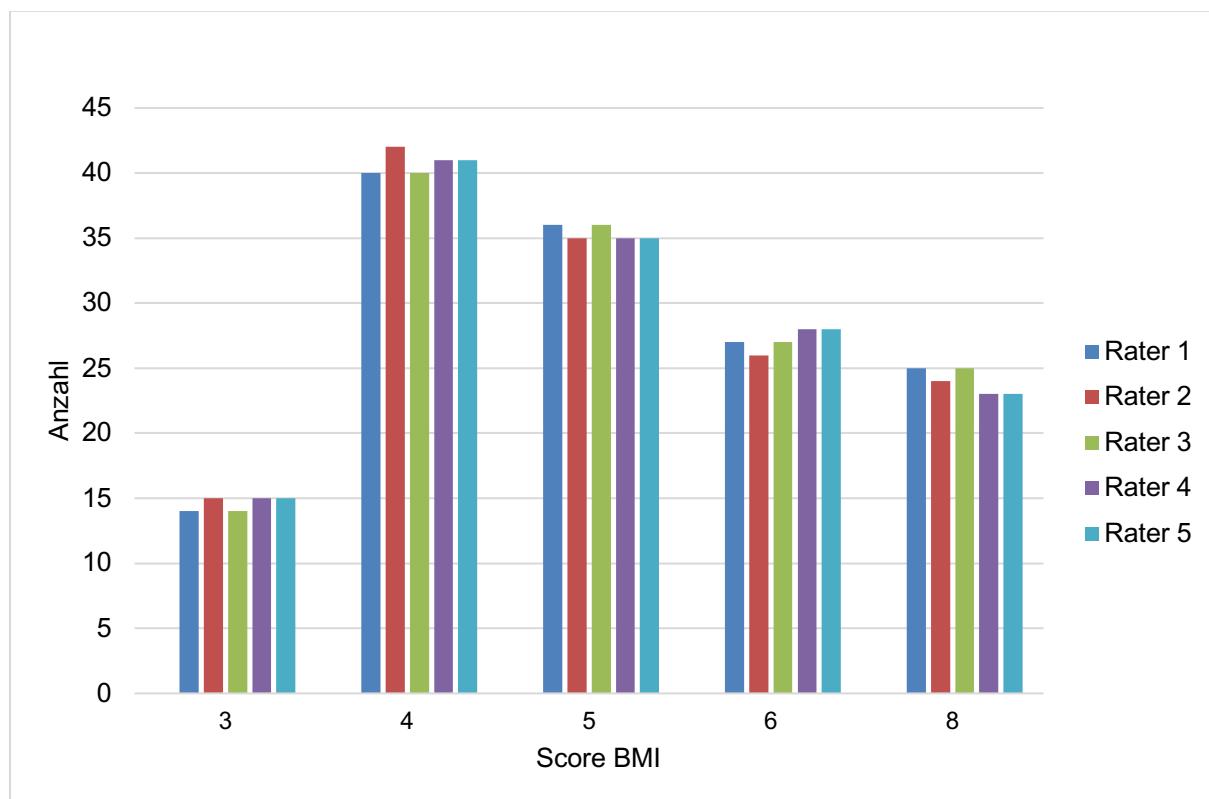


Abb. 4 Häufigkeitsverteilung für die Scores der Kategorie "BMI"

In der Kategorie „BMI“ wurde von keinem Rater null Punkte vergeben. Alle Rater haben zwischen drei und acht Punkten vergeben. Auch hier sind die Ergebnisse der Häufigkeitsverteilung statistisch gleich verteilt ($p = 1.0$).

Tab. 5 Häufigkeitsverteilung der Scoreergebnisse für die Kategorie „BMI“

Score in Punkten	Rater 1	Rater 2	Rater 3	Rater 4	Rater 5
0	0	0	0	0	0
3	14	15	14	15	15
4	40	42	40	41	41
5	36	35	36	35	35
6	27	26	27	28	28
8	25	24	25	23	23
Total n =	142	142	142	142	142

4.4 Häufigkeitsverteilung der Scoreergebnisse für die Kategorie „functional status“

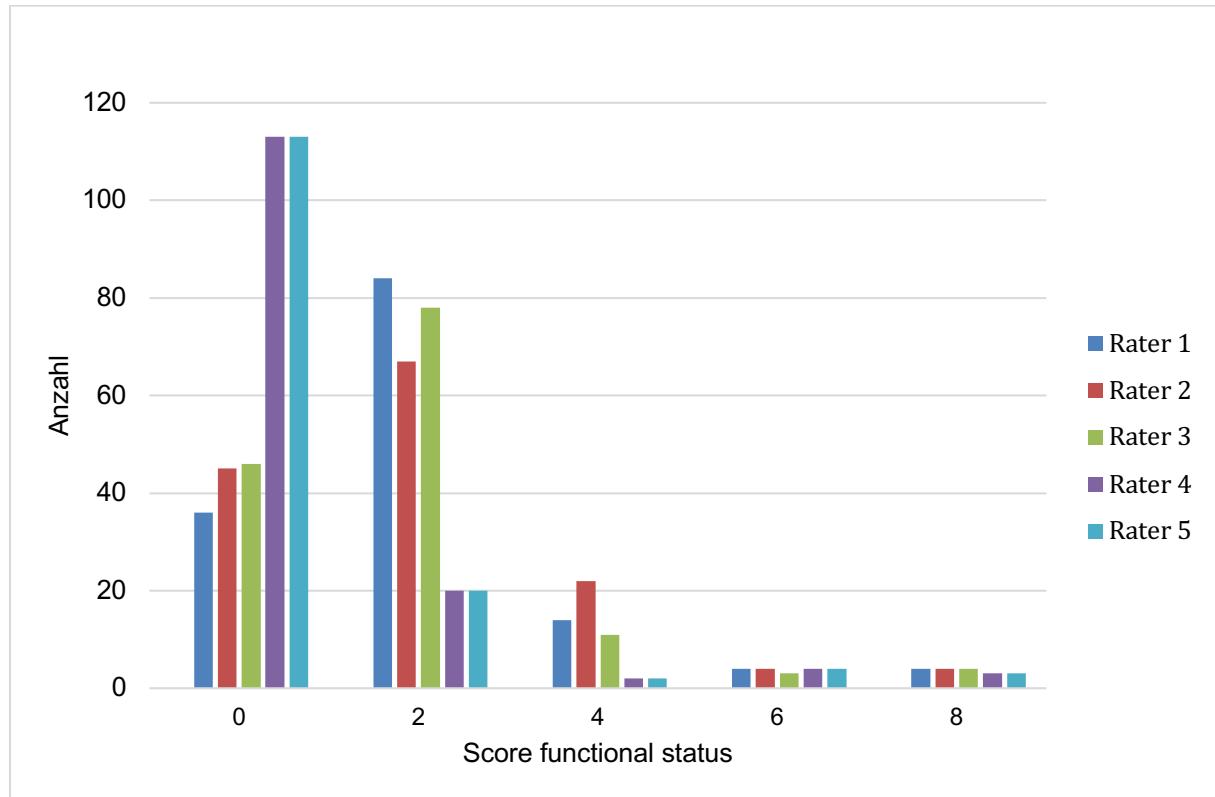


Abb. 5 Häufigkeitsverteilung für die Scores der Kategorie "functional status"

In dieser Kategorie finden sich signifikante Unterschiede zwischen den Ratern ($p < 0.001$). Rater 4 und 5 haben deutlich öfter null Punkte und deutlich weniger zwei Punkte vergeben. Bei vier Punkten zeigt ich das Rater 1, 2 und 3 öfter diese Punktzahl im Vergleich zu Rater 4 und 5 vergeben haben. Für die Punkte sechs und acht ist die Verteilung wieder ausgeglichen.

Tab. 6 Häufigkeitsverteilung für die Scoreergebnisse Kategorie "functional status"

Score in Punkten	Rater 1	Rater 2	Rater 3	Rater 4	Rater 5
0	36	45	46	113	113
2	84	67	78	20	20
4	14	22	11	2	2
6	4	4	3	4	4
8	4	4	4	3	3
Total n =	142	142	142	142	142

4.5 Häufigkeitsverteilung der Scoreergebnisse für die Kategorie „physical status“

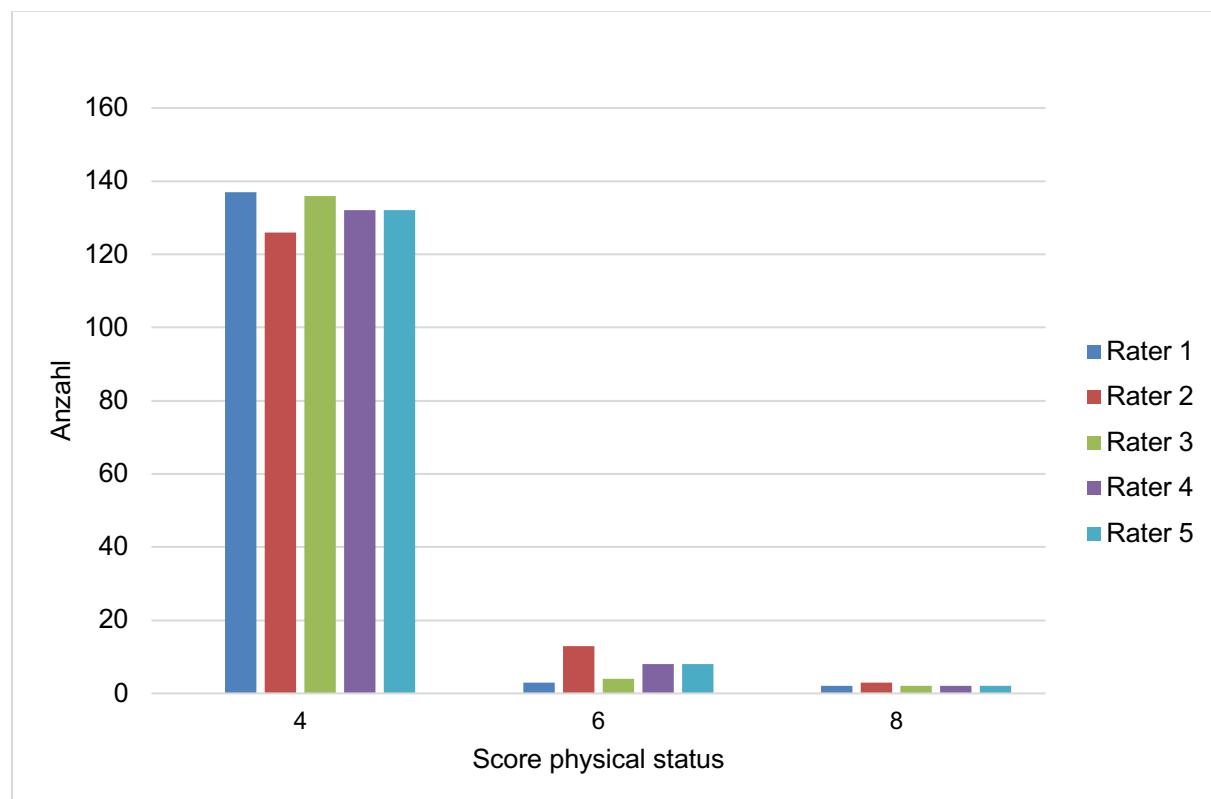


Abb. 6 Häufigkeitsverteilung für die Scores der Kategorie "physical status"

In dieser Kategorie wurden null und zwei Punkte von keinem Rater vergeben. Die größten Unterschiede zeigen sich bei sechs Punkten. Rater 2 hat diese deutlich mehr vergeben als die anderen Rater. Rater 1 und 3 haben diese am wenigsten bepunktet. Die Differenzen sind nicht signifikant ($p = 0.27$).

Tab. 7 Häufigkeitsverteilung der Scoreergebnisse für die Kategorie "physical status"

Score in Punkten	Rater 1	Rater 2	Rater 3	Rater 4	Rater 5
0	0	0	0	0	0
2	0	0	0	0	0
4	137	126	136	132	132
6	3	13	4	8	8
8	2	3	2	2	2
Total n =	142	142	142	142	142

4.6 Häufigkeitsverteilung der Scoreergebnisse für die Kategorie „pulmonary status“

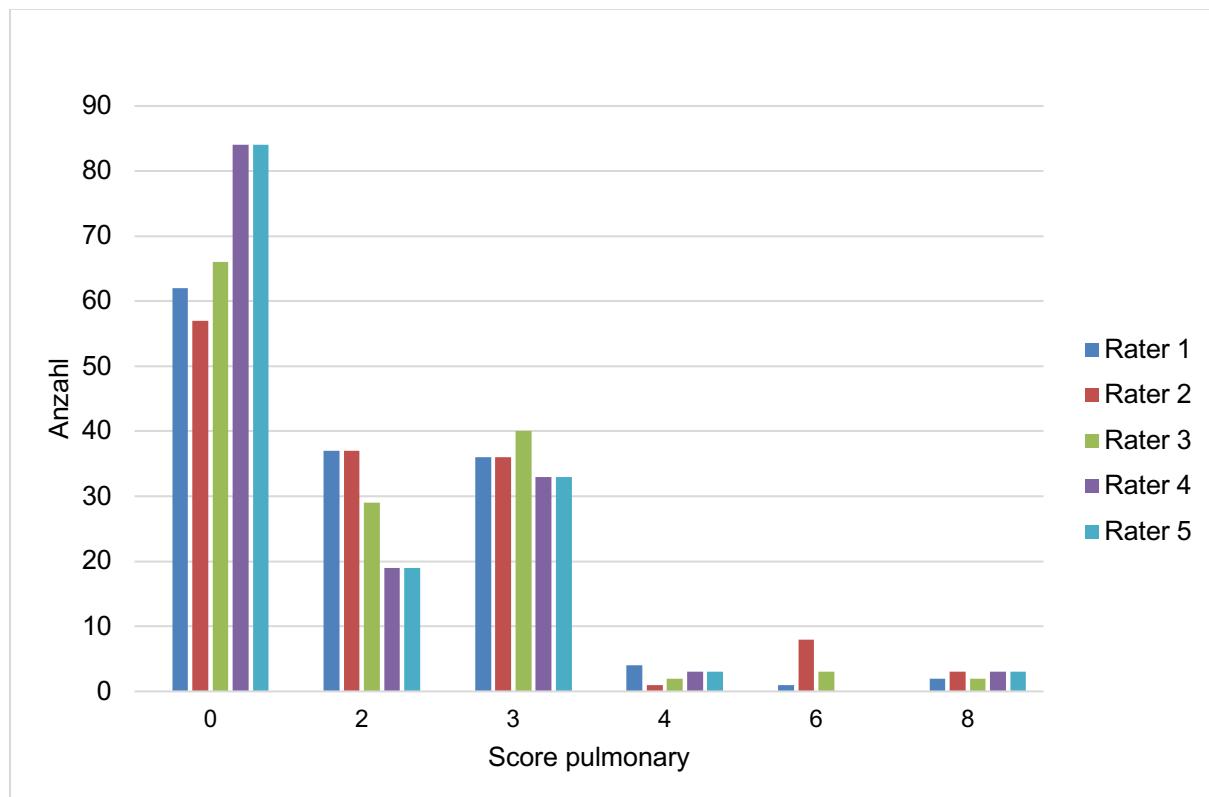


Abb. 7 Häufigkeitsverteilung für die Scores der Kategorie "pulmonary status"

Unterschiede bei der Bewertung zeigen sich vor allem bei den Punkten null, zwei und sechs. Die Unterschiede sind signifikant ($p = 0.03$).

Tab. 8 Häufigkeitsverteilung der Scoreergebnisse für die Kategorie "pulmonary status"

Score in Punkten	Rater 1	Rater 2	Rater 3	Rater 4	Rater 5
0	62	57	66	84	84
2	37	37	29	19	19
3	36	36	40	33	33
4	4	1	2	3	3
6	1	8	3	0	0
8	2	3	2	3	3
Total n =	142	142	142	142	142

4.7 Häufigkeitsverteilung der Scoreergebnisse für die Kategorie „diabetes status“

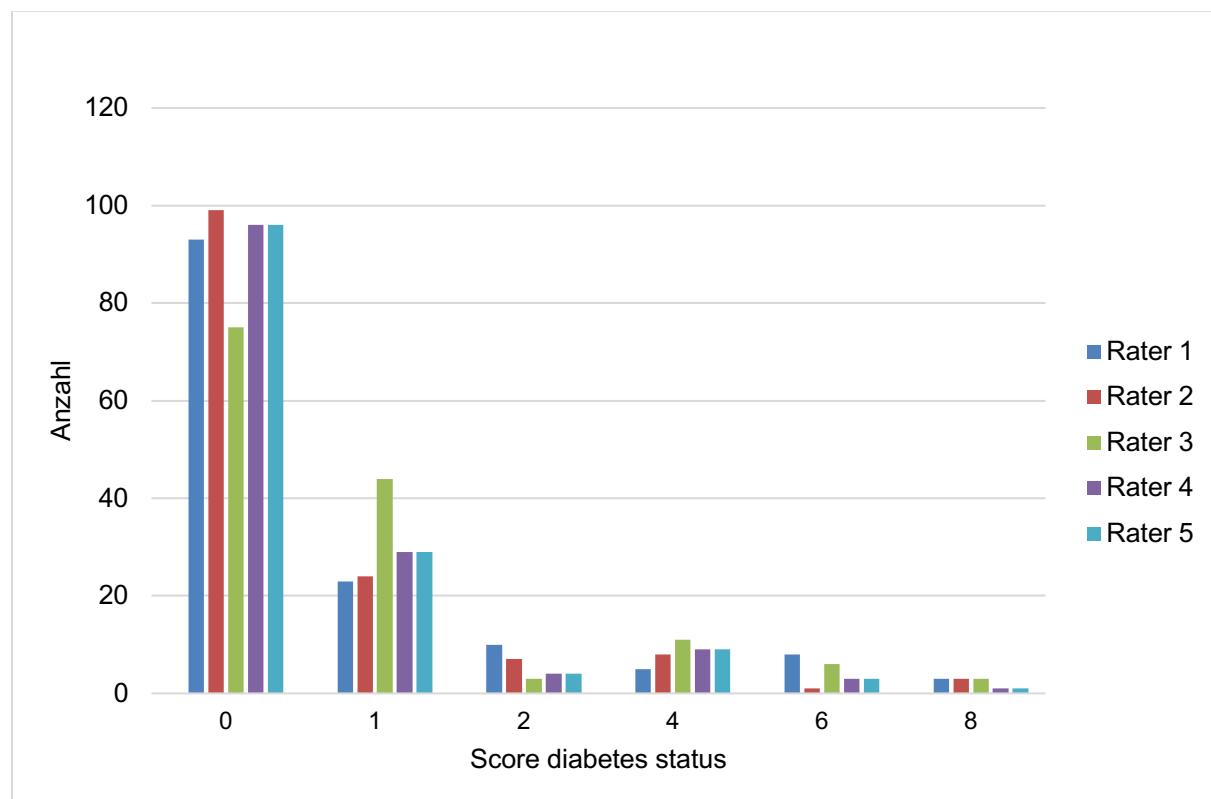


Abb. 8 Häufigkeitsverteilung für die Scores der Kategorie "diabetes status"

Nicht signifikante Unterschiede zwischen den Ratern finden sich bei allen Punkten.
Rater 3 zeigt sich als Ausreißer bei den Punkten null und eins. ($p = 0.051$)

Tab. 9 Häufigkeitsverteilung für die Scoreergebnisse der Kategorie "diabetes status"

Score in Punkten	Rater 1	Rater 2	Rater 3	Rater 4	Rater 5
0	93	99	75	96	96
1	23	24	44	29	29
2	10	7	3	4	4
4	5	8	11	9	9
6	8	1	6	3	3
8	3	3	3	1	1
Total n=	142	142	142	142	142

4.8 Ergebnisse des Gesamtscore

Rater 4 und 5 haben die exakt gleichen Ergebnisse erzielt. Die Graphen liegen übereinander und sind nicht zu differenzieren. Rater 1,2 und 3 haben sehr ähnliche Kurvenverläufe und unterscheiden sich zu denen von Rater 4 und 5. Die Unterschiede sind signifikant ($p = 0.007$). Die Ergebnisse der einzelnen Rater werden im Anhang (Abb. 10–14) in jeweils einem eigenen Histogramm dargestellt.

Tab. 10 Deskriptive Darstellung der Scoreergebnisse für den Gesamtscore der Rater

	N	Minimum	Maximum	Mean	Std. Deviation	Median
Gesamtscore Rater 1	142	8	34	15.67	4.591	15
Gesamtscore Rater 2	142	8	35	15.73	4.658	15
Gesamtscore Rater 3	142	7	34	15.61	4.705	15
Gesamtscore Rater 4	142	7	35	13.74	4.484	13
Gesamtscore Rater 5	142	7	35	13.74	4.484	13

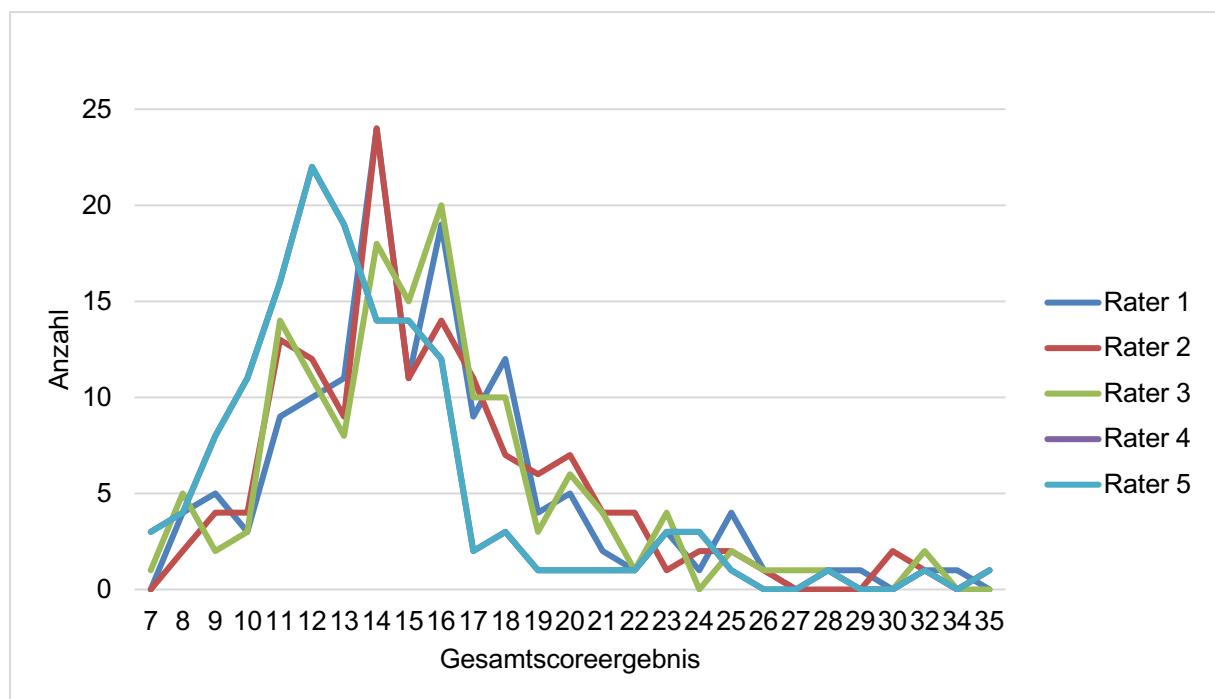


Abb. 9 Häufigkeitsverteilung für die Ergebnisse des Gesamtscores

4.9 Varianzhomogenität

Homogenität der Varianzen zwischen den Ratern war gemäß dem Levene-Test basierend auf dem Median für die Ergebnisse *age*, *gender*, *BMI*, *physical status*, *pulmonary status*, *diabetes status* und Gesamtscore erfüllt ($p > 0.05$), nicht aber für die Kategorie *functional status* ($p = <.001$).

Tab. 11 Ergebnisse der Signifikanz des Levenetest

Kategorie	p =
age	.919
gender	1.0
BMI	.994
functional status	< .001
physical status	.111
pulmonary status	.561
diabetes status	.142
Gesamtscore	.738

Die Varianzen sind gleich, wenn $p > .05$ ist.

4.10 Interklassenkorrelationskoeffizient

Die Interpretation des ICC wird in der Literatur diskutiert. Unterschiedliche Interpretationsweisen sind in der Tabelle 12 dargestellt. Die Ergebnisse dieser Arbeit werden auf der neueren Definition nach Koo & Li (2016) interpretiert. Durch die strenge Bewertung können robustere Ergebnisse bestimmt werden.

Tab. 12 ICC-Interpretation

Outcome	Koo & Li	Cicchetti and Sparrow
<i>poor</i>	<0,5	<0,4
<i>fair</i>	0,5–0,75	0,4–0,6
<i>good</i>	0,75–0,9	0,6–0,75
<i>excellent</i>	≥ 0,9	≥ 0,75

(Cicchetti and Sparrow, 1981; Koo and Li, 2016)

Die Kategorie „individual Status“ besteht aus drei Unterpunkten: *age*, *gender* und *BMI*. Für den Unterpunkt „*age*“ ist der *ICC* = 0.97 und einem Konfidenzintervall von [0.96,0.98] signifikant ($p < 0.001$).

Tab. 13 *ICC* für die Kategorie "age"

	ICC	95% Confidence Interval		Sig
		Lower Bound	Upper Bound	
Single Measures	.974	.966	.980	<.001
Average Measures	.995	.993	.996	<.001

Für den Unterpunkt „*gender*“ ist der *ICC* = 0.98 und einem Konfidenzintervall von [0.98,0.99] signifikant ($p < 0.001$).

Tab. 14 *ICC* für die Kategorie „*gender*“

	ICC	95% Confidence Interval		Sig
		Lower Bound	Upper Bound	
Single Measures	.989	.986	.991	<.001
Average Measures	.998	.997	.998	<.001

Für den Unterpunkt „*BMI*“ ist der *ICC* = 0.97 und einem Konfidenzintervall von [0.96,0.98] signifikant ($p < 0.001$).

Tab. 15 *ICC* für die Kategorie *BMI*

	ICC	95% Confidence Interval		Sig
		Lower Bound	Upper Bound	
Single Measures	.976	.969	.982	<.001
Average Measures	.995	.994	.996	<.001

Für die Kategorie „*functional status*“ ist der *ICC* = 0.58 und einem Konfidenzintervall von [0.44,0.69] signifikant ($p < 0.001$).

Tab. 16 *ICC* für die Kategorie "functional status"

	ICC	95% Confidence Interval		Sig
		Lower Bound	Upper Bound	
Single Measures	.586	.440	.698	<.001
Average Measures	.876	.797	.920	<.001

Für die Kategorie „*physical status*“ ist der ICC = 0.68 und einem Konfidenzintervall von [0.62,0.74] signifikant ($p < 0.001$).

Tab. 17 ICC für die Kategorie "physical status"

	ICC	95% Confidence Interval		Sig
		Lower Bound	Upper Bound	
Single Measures	.683	.620	.743	<.001
Average Measures	.915	.891	.935	<.001

Für die Kategorie „*pulmonary status*“ ist der ICC = 0.79 und einem Konfidenzintervall von [0.74,0.84] signifikant ($p < 0.001$).

Tab. 18 ICC für die Kategorie „pulmonary status“

	ICC	95% Confidence Interval		Sig
		Lower Bound	Upper Bound	
Single Measures	.798	.745	.844	<.001
Average Measures	.932	.905	.951	<.001

Für die Kategorie „*diabetes status*“ ist der ICC = 0.74 und einem Konfidenzintervall von [0.68,0.79] signifikant ($p < 0.001$).

Tab. 19 ICC für die Kategorie "diabetes status"

	ICC	95% Confidence Interval		Sig
		Lower Bound	Upper Bound	
Single Measures	.743	.686	.794	<.001
Average Measures	.935	.916	.951	<.001

Für den Gesamtscore ist der ICC = 0.84 und einem Konfidenzintervall von [0.75,0.89] signifikant ($p < 0.001$).

Tab. 20 ICC für den Gesamtscore

	ICC	95% Confidence Interval		Sig
		Lower Bound	Upper Bound	
Single Measures	.845	.756	.899	<.001
Average Measures	.960	.929	.976	<.001

4.11 Korrelationsanalyse nach Pearson

Für sechs von sieben Kategorien konnten Korrelationsanalysen nach Pearson berechnet werden (Tab. 21–26). Für die Kategorie „*physical status*“ konnte keine Korrelation berechnet werden. Ab Werten von $r = 0.5$ wird von einem starken Zusammenhang ausgegangen. Werte zwischen $r = 0\text{--}0.1$ sind schwach und zwischen $r = 0.1\text{--}0.5$ als moderat zu bewerten (Cohen, 1992). Es zeigen sich folgend nur positive Korrelationen. Alle Korrelationen sind signifikant ($p > 0.01$).

Tab. 21 Korrelationen der Variable Alter mit den Scorergebnissen „age“

	Age r =
Rater 1 Score „age“	.921**
Rater 2 Score „age“	.920**
Rater 3 Score „age“	.922**
Rater 4 Score „age“	.920**
Rater 5 Score „age“	.920**

**. Correlation is significant at the 0.01 level

Tab. 22 Korrelationen der Variable Geschlecht mit den Scorergebnissen "gender"

	Gender r =
Rater 1 Score „gender“	.944**
Rater 2 Score „gender“	.944**
Rater 3 Score „gender“	.958**
Rater 4 Score „gender“	.958**
Rater 5 Score „gender“	.958**

**. Correlation is significant at the 0.01 level

Tab. 23 Korrelationen der Variable BMI mit den Scoreergebnissen BMI

	BMI r =
Rater 1 Score BMI	.931**
Rater 2 Score BMI	.933**
Rater 3 Score BMI	.929**
Rater 4 Score BMI	.916**
Rater 5 Score BMI	.916**

**. Correlation is significant at the 0.01 level

Tab. 24 Korrelationen der orthopädischen Erkrankungen mit den Scoreergebnissen „*functional status*“

	Orthopädische Erkrankungen r =
Rater 1 Score „ <i>functional status</i> “	.465**
Rater 2 Score „ <i>functional status</i> “	.307**
Rater 3 Score „ <i>functional status</i> “	.519**
Rater 4 Score „ <i>functional status</i> “	.397**
Rater 5 Score „ <i>functional status</i> “	.397**

**. Correlation is significant at the 0.01 level

Tab. 25 Korrelationen der pulmonalen Erkrankungen mit den Scoreergebnissen „*pulmonary status*“

	Pulmonale Erkrankungen r =
Rater 1 Score „ <i>pulmonary status</i> “	.747**
Rater 2 Score „ <i>pulmonary status</i> “	.639**
Rater 3 Score „ <i>pulmonary status</i> “	.789**
Rater 4 Score „ <i>pulmonary status</i> “	.716**
Rater 5 Score „ <i>pulmonary status</i> “	.716**

**. Correlation is significant at the 0.01 level

Tab. 26 Korrelationen des HbA1c mit den Scoreergebnissen „*diabetes status*“

	HbA1c r=
Rater 1 Score „ <i>diabetes status</i> “	.837**
Rater 2 Score „ <i>diabetes status</i> “	.706**
Rater 3 Score „ <i>diabetes status</i> “	.877**
Rater 4 Score „ <i>diabetes status</i> “	.687**
Rater 5 Score „ <i>diabetes status</i> “	.687**

**. Correlation is significant at the 0.01 level

5 DISKUSSION

In der Kategorie „age“ sind nahezu keine Unterschiede in der Häufigkeitsverteilung zu erkennen. Der ICC liegt bei 0.98 und ist als exzellent zu bewerten. Der entsprechende Levenetest ist nicht signifikant und die Interpretation somit akkurat. Die Korrelationen nach Pearson liegt für alle Rater bei $r = 0.92$ und zeigt einen sehr starken Zusammenhang zwischen dem Alter der Patienten und den Ergebnissen der Rater dieser Kategorie. Die marginalen Unterschiede können neben den Bewertungsfehlern durch die Rater auch mit der Definition erklärt werden. So sollte anstatt > 40 Jahren und > 50 Jahren eine genauere Abgrenzung ≥ 40 und ≥ 50 verwendet werden.

In der Kategorie „gender“ sind ebenfalls nicht signifikante Unterschiede in den Scores zwischen den Ratern erkennbar. Der ICC liegt bei 0.98 und ist als exzellent zu bewerten bei nicht signifikantem Levenetest. Die Korrelationsanalyse nach Pearson liegt zwischen $r = 0.94–0.95$ und zeigt einen hohen Zusammenhang zwischen dem Geschlecht und den Ergebnissen der Rater in der Kategorie „gender“.

Die Lebenserwartung von übergewichtigen Männern ist im Vergleich zu übergewichtigen Frauen deutlich niedriger mit einem Durchschnittswert von 54.3 bzw. 60.7 Jahre (Muennig et al., 2006). Es erscheint somit sinnvoll Männern eine höhere Punktzahl (zwei) zugeben, da stark übergewichtige Männer statistisch früher sterben als Frauen.

Die Kategorie BMI zeigt eine homogene Verteilung der Scores. Zwei Punkte wurden von keinem Rater vergeben da es keinen Patienten mit einem BMI unter 50 gab. Der ICC ist mit 0.97 ein exzellentes Ergebnis bei nicht signifikantem Levenetest. Auch die Korrelation nach Pearson zeigt eine sehr gute Übereinstimmung zwischen dem erhobenem BMI und den Ergebnissen der Rater. Die Punktzahl mit einem steigendem BMI zu erhöhen ist sinnvoll durch die bereits genannte steigende Mortalität mit steigendem BMI (Kitahara et al., 2014). Zu empfehlen wäre auch hier die Anpassung der Definition. So sollte anstatt > 40 , ein ≥ 40 verwendet werden.

Die Kategorien „age“, „gender“ und BMI liefern reliable, objektive und valide Ergebnisse.

Für die Kategorie „functional status“ ergeben sich in der Häufigkeitsverteilung der Scoreergebnisse für die Punkte null, zwei und vier, signifikante Unterschiede zwischen den Ratern. Der ICC liegt bei 0.58 mit einem signifikantem Levenetest. Mit Einbeziehung des Konfidenzintervalls zeigt sich hier nach Koo und Li (2016) ein

schlechtes bis moderates Ergebnis. Durch den signifikanten Leventest kann möglicherweise von einem in Wahrheit höherem ICC ausgegangen werden (Wirtz and Caspar, 2002). Für die Korrelation nach Pearson können nur für die Ergebnisse von Rater 3 ($r = 0.51$), starke Zusammenhänge mit den orthopädischen Erkrankungen gezeigt werden. Für die anderen Rater ist der Zusammenhang nur moderat ($r = 0.30\text{--}0.46$). Diese Werte lassen auf eine geringe Objektivität der Kategorie schließen.

Es werden Punkte für Unabhängigkeit, teilweise Abhängigkeit, komplett Abhängigkeit, Rollstuhlmobilität und Immobilität vergeben. Sind die beiden letzteren Punkte klar definierte Zustände, welche auch einfach nachgewiesen werden können, so ist dies für die ersten drei Items nicht möglich. Es ist nicht trennscharf zu unterscheiden, auf was sich die Abhängigkeit bezieht. Ein rollstuhlmobiler Patient ist ebenso wie ein immobiler Patient komplett abhängig. Der eine von einem Rollstuhl und der letztere von einer Pflegefachkraft oder weiteren Personen und Hilfsmitteln. Patienten mit einer Gonarthrose können schmerzfrei mobil oder durch Gehhilfen mobil, aber von diesen abhängig sein. Diese Beispiele sollen die reduzierte Objektivität veranschaulichen. Den Ratern können sich bei der Bewertung in dieser Kategorie an keinen objektiven Daten orientieren, weil auch keine expliziten Informationen zu diesen Items erhoben wurden. Dieser Kategorie kann trotz der moderaten Ergebnisse vor allem aber aufgrund der beschriebenen fehlenden Objektivität auch keine Reliabilität und somit auch keine Validität zugesprochen werden.

Die Kategorie „*physical status*“ zeigt keine statistisch signifikanten Unterschiede in der Häufigkeitsverteilung der Scoreergebnisse. Was in dieser Kategorie besonders auffällt ist, dass null und zwei Punkte gar nicht vergeben wurden. Da alle Patienten in diesem Kollektiv morbide adipös sind, haben diese einen $\text{BMI} > 40 \text{ kg/m}^2$. Nach der ASA-Klassifikation, an welcher sich diese Kategorie orientiert, sind Patienten mit ein $\text{BMI} > 40 \text{ kg/m}^2$ per definitionem mindestens mit ASA III, also in diesem Fall mit vier Punkten zu bewerten. Der ICC für diese Kategorie ist moderat und liegt bei 0.68 mit einem nicht signifikaten Levenetest. Eine Korrelationsanalyse konnte für diese Kategorie nicht durchgeführt werden, weil kein Außenkriterium aus den erhobenen Daten ermittelt werden konnte. Da sich die Bewertung an der ASA-Klassifikation orientiert wäre hier die Korrelation zu dieser ein guter Referenzwert um die Validität zu prüfen. Die ASA-Klassifikation ist ein etabliertes und valides Einschätzungsstool zum Gesundheitszustand eines Patienten. Es ist von einer hohen Validität auszugehen und kann ggf. in einer weiteren Validierung in Zukunft geprüft werden.

Die Kategorie „*pulmonary status*“ zeigt signifikante Unterschiede in der Punkteverteilung. Rater 4 und 5 haben öfter null Punkte und weniger häufig zwei Punkte vergeben als die anderen Rater. Sechs Punkte wurden von diesen Ratern gar nicht vergeben. Der ICC ist mit 0.79 und einem nicht signifikantem Levenetest gut. Die Items dieser Kategorie sind klar und objektiv definiert. Daten, wie gut die Patienten ohne Luftnot Treppensteigen können, wurden nicht erhoben. Somit konnten die Rater bezogen, auf die ersten zwei Items, in Ihren Bewertungen spekulieren, ob der Patient zwei Etagen ohne Luftnot erklimmen kann oder nicht. Allerdings wurden Informationen zu Heimsauerstoff, respiratorischen Diagnosen und Tracheotomien erfasst. Die Rater konnten somit genaue Scores von drei bis acht Punkten vergeben.

Korrelationen zwischen den respiratorischen Erkrankungen und den Scores der Rater sind hoch und liegen zwischen $r = 0.63\text{--}0.78$. Rater 2 hat die niedrigste Korrelation, was mit den abweichenden Ergebnissen der sechs Punkte Scores zusammenhängen kann.

Die Häufigkeitsverteilung der letzten Kategorie, „*diabetes status*“ zeigt keine signifikanten Unterschiede der Scoreergebnisse. Der ICC ist gut, mit einem Wert von 0.74, bei einem nicht signifikanten Levenetest.

Die Korrelationsanalyse zeigt hohe Zusammenhänge mit Werten von $r = 0.68\text{--}0.87$. Rater 1 und 3 zeigen einen höheren Zusammenhang zum gemessenen HbA1c als die anderen Rater.

Die Items dieser Kategorie sind klar abgegrenzt. Die Objektivität ist gegeben und mit einem guten ICC-Ergebnis ist auch Reliabilität vorhanden. Starke Korrelationen des HbA1c mit den Ergebnissen machen diese Kategorie valide.

Für das Gesamtergebnis konnte eine gute Reliabilität mit einem ICC von 0.84 bestimmt werden. Es konnte jedoch keine Korrelationsanalyse mit einem anderen Konstrukt durchgeführt werden. Dieser Score ist in seiner Form bisher einzigartig und mit keinem der anderen Scores vergleichbar.

Trotz der nicht objektiven Kategorie „*functional status*“ ist der Gesamtscore reliabel und auf Basis der ermittelten Kriteriumsvalidität von sechs Kategorien auch valide. Eine Reduktion um das Kriterium „*functional status*“ oder die Löschung von Items in dieser Kategorie könnte möglicherweise die Reliabilität etwas steigern. Sinnvoll wäre allerdings eine Anpassung der Items durch eine genauere Definition. Eine Möglichkeit wäre die folgende Definition: eigenständig mobil mit Treppensteigen, in der Ebene

eigenständig mobil, mobil mit Gehilfen, rollstuhlmobil und immobil. Für die Ergebnisse dieser Anpassung auf den ICC müssten jedoch neue Daten in Zukunft erhoben und ausgewertet werden.

5.1 Limitierungen und Ausblick

Die Daten waren nicht spezifisch auf die Anwendung des Scores erhoben. Vorallem die Bewertung der Kategorie des „*functional status*“ war eingeschränkt und ließ Raum für Spekulation durch die Rater zu. Auch konnte dadurch kein Außenkriterium zur Korrelationsanalyse für die Überprüfung der Kriteriumsvalidität der Kategorie „*physical status*“ ermittelt werden.

Für eine endgültige Validierung sollte der Score auf eine Nicht-Risiko-Gruppe angewendet werden, um zu evaluieren, ob die erhobenen Gesamtwerte mit deutlicher Mehrheit signifikant unter dem aktuellen Median liegen. Zur weiteren Validierung könnten die Unterschiede zwischen einer niedrigen und einer hohen Punktzahl in Bezug auf Mortalität, prä- und postoperativen Gewichtsverlust sowie mögliche perioperative Komplikationen weiter untersucht werden. Eine weitere Möglichkeit ist die Untersuchung, inwieweit sich der Score ein Jahr nach einer bariatrischen Operation für einen Patienten verändert hat. Aasheim et al. (2011) haben diese Methodik auf die KOSC angewendet und somit den KOSC Validität bescheinigt.

6 ZUSAMMENFASSUNG

Die quantitative Testtheorie beschreibt die Erfüllung der drei Hauptgütekriterien Objektivität, Reliabilität und Validität für die Validierung eines Scores. Gegenstand der vorliegenden Arbeit ist die evidenzbasierte Validierung eines neuartigen, von Stier et al. (2022) entwickelten Notfallscores für bariatrische Patienten auf Basis dieser Gütekriterien. Es ist nach aktuellem Stand der erste Score dieser Art. Der Score soll die Dringlichkeit zur unmittelbaren stationären Aufnahme mit anschließender „*fast-track rescue weight reduction bridging therapy*“ und folgender bariatrischer Operation ermitteln. Der Score ist positiv (*cut-off*) ab einem Wert von 14 Punkten oder bei maximaler Punktzahl acht in den Kategorien „*physical status*“ oder „*pulmonary status*“. Der Score besteht aus den Kategorien „*age*“, „*gender*“, BMI, „*functional status*“, „*physical status*“, „*pulmonary status*“ und „*diabetes status*“.

Es wurden Daten von $n = 142$ Patienten gesammelt. Anschließend wurde das Patientenkollektiv von $n = 5$ ärztlichen Ratern auf Basis des genannten Scores bewertet. Eine Reliabilitätsanalyse zeigte, dass für sechs von sieben Kategorien des Scores reliable Ergebnisse erzielt werden konnten. Der ICC der einzelnen Kategorien und Rater lag zwischen 0.58 und 0.98. Nach Koo & Li (2016) sind dies moderate bis sehr gute Ergebnisse. Der Gesamtscore hatte einen ICC von 0.84 und somit ein gutes Ergebnis erzielt.

Weiterhin wurde die Objektivität des Scores diskutiert. Für fast alle Kategorien konnte ausreichende Objektivität bescheinigt werden. Lediglich für die Kategorie „*functional status*“ konnte keine Objektivität der Items beschrieben werden.

Anschließend wurde die Validität auf Basis der Kriteriumsvalidität per Pearsonkorrelationanalyse bestimmt. Für sechs von sieben Kategorien des Scores konnten positive Korrelationen berechnet werden. Die Korrelationen in der schlechtesten Kategorie lagen bei $r = 0.30$ („*functional status*“) und für die beste bei $r = 0.95$ („*gender*“). Aufgrund fehlender Daten konnte für die Kategorie „*physical status*“ keine Korrelation ermittelt werden. Für den Gesamtscore konnte aufgrund mangelnder Datenerhebung keine Konstruktvalidität bestimmt werden.

Limitierungen dieser Arbeit bestanden darin das teilweise nicht ausreichend Daten der Patienten bereit standen um genauere Ergebnisse des Scores zu ermitteln. Nichtsdestotrotz bescheinigen die vorliegenden Ergebnisse dem Gesamtscore ausreichende Objektivität, Reliabilität und Validität. Weitere Validierung wird jedoch

empfohlen vor allem in Bezug auf andere Konstrukte oder Anwendung auf eine Nicht-Risikogruppe.

7 LITERATURVERZEICHNIS

Aasheim, E.T., Aylwin, S.J.B., Radhakrishnan, S.T., Sood, A.S., Jovanovic, A., Olbers, T. und le Roux, C.W. (2011). Assessment of obesity beyond body mass index to determine benefit of treatment. Clinical Obesity 1, 77-84. <https://doi.org/https://doi.org/10.1111/j.1758-8111.2011.00017.x>

Apgar, V. (1953). A proposal for a new method of evaluation of the newborn infant. Curr Res Anesth Analg 32, 260-267.

Arterburn, D.E., Telem, D.A., Kushner, R.F. und Courcoulas, A.P. (2020). Benefits and Risks of Bariatric Surgery in Adults: A Review. Jama 324, 879-887. <https://doi.org/10.1001/jama.2020.12567>

Athayde, R.A.B., Oliveira Filho, J.R.B., Lorenzi Filho, G. und Genta, P.R. (2018). Obesity hypoventilation syndrome: a current review. J Bras Pneumol 44, 510-518. <https://doi.org/10.1590/s1806-37562017000000332>

Aune, D., Schlesinger, S., Norat, T. und Riboli, E. (2018). Body mass index, abdominal fatness, and the risk of sudden cardiac death: a systematic review and dose-response meta-analysis of prospective studies. Eur J Epidemiol 33, 711-722. <https://doi.org/10.1007/s10654-017-0353-9>

Baker, S.P., O'Neill, B., Haddon, W., Jr. und Long, W.B. (1974). The injury severity score: a method for describing patients with multiple injuries and evaluating emergency care. J Trauma 14, 187-196.

Bortz, N.D.J. (2016). Forschungs-methoden und Evaluation in den Sozial- und Humanwissenschaften, S.440-448, Bd. 5 (Springer-Verlag Berlin Heidelberg).

Bouch, D.C. und Thompson, J.P. (2008). Severity scoring systems in the critically ill. Continuing Education in Anaesthesia Critical Care & Pain 8, 181-185. <https://doi.org/10.1093/bjaceaccp/mkn033>

Brix, J.M., Andersen, B., Aydinkoc-Tuzcu, K., Beckerhinn, P., Brossard-Eitzinger, A., Cavini, A., Ciardi, C., Clodi, M., Eichner, M., Erlacher, B., et al. (2023). [Overweight and obesity in adults: general principles of treatment and conservative management]. Wien Klin Wochenschr 135, 706-720. <https://doi.org/10.1007/s00508-023-02270-9>

Brock, J. (2021). Warum es sich auch in der Pneumologie lohnt, aufs Gewicht zu achten. Pneumo News 13, 28-34. <https://doi.org/10.1007/s15033-021-2677-0>

Burki, T. (2021). European Commission classifies obesity as a chronic disease. Lancet Diabetes Endocrinol 9, 418. [https://doi.org/10.1016/s2213-8587\(21\)00145-5](https://doi.org/10.1016/s2213-8587(21)00145-5)

Chang, Y., Ryu, S., Choi, Y., Zhang, Y., Cho, J., Kwon, M.-J., Hyun, Y.Y., Lee, K.-B., Kim, H., Jung, H.-S., et al. (2016). Metabolically Healthy Obesity and Development of Chronic Kidney Disease: A Cohort Study. Ann Intern Med 164, 305-312. <https://doi.org/10.7326/m15-1323>

Chiappetta, S., Stier, C., Squillante, S., Theodoridou, S. und Weiner, R.A. (2016). The importance of the Edmonton Obesity Staging System in predicting postoperative

outcome and 30-day mortality after metabolic surgery. *Surg Obes Relat Dis* 12, 1847-1855. <https://doi.org/10.1016/j.soard.2016.02.042>

Cicchetti, D.V. und Sparrow, S.A. (1981). Developing criteria for establishing interrater reliability of specific items: applications to assessment of adaptive behavior. *Am J Ment Defic* 86, 127-137.

Clodi, M., Toplak, H., Resl, M., Brix, J., Leitner, D.R., Harreiter, J., Hoppichler, F., Wascher, T.C., Schindler, K. und Ludvik, B. (2023). [Obesity and type 2 diabetes (Update 2023)]. *Wien Klin Wochenschr* 135, 91-97. <https://doi.org/10.1007/s00508-023-02184-6>

Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112 (1), 155–159.

DAG (2023): ADIPOSITAS-MEDIKAMENTE: ANTWORTEN AUF HÄUFIGE FRAGEN. https://adipositas-gesellschaft.de/adipositas-medikamente-fragen-und-antworten/#_edn6. Abgerufen 04.06.2024.

Derrick, B., Ruck, A.-M., Toher, D. und White, P. (2018). Tests for equality of variances between two samples which contain both paired observations and independent observations.

DGAV, D.G.f.A.-u.V.e.V. (2018): S3-Leitlinie: Chirurgie der Adipositas und metabolischer Erkrankungen, AWMF-Register Nr. 088-001 Abgerufen 04.08.2024.

Di Giacinto, I., Guarnera, M., Esposito, C., Falcetta, S., Cortese, G., Pascarella, G., Sorbello, M. und Cataldo, R. (2021). Emergencies in obese patients: a narrative review. *Journal of Anesthesia, Analgesia and Critical Care* 1, 13. <https://doi.org/10.1186/s44158-021-00019-2>

Eguchi, Y., Hyogo, H., Ono, M., Mizuta, T., Ono, N., Fujimoto, K., Chayama, K., Saibara, T. und Jsg, N. (2012). Prevalence and associated metabolic factors of nonalcoholic fatty liver disease in the general population from 2009 to 2010 in Japan: a multicenter large retrospective study. *Journal of Gastroenterology* 47, 586-595. <https://doi.org/10.1007/s00535-012-0533-z>

Field, A.E., Coakley, E.H., Must, A., Spadano, J.L., Laird, N., Dietz, W.H., Rimm, E. und Colditz, G.A. (2001). Impact of overweight on the risk of developing common chronic diseases during a 10-year period. *Arch Intern Med* 161, 1581-1586. <https://doi.org/10.1001/archinte.161.13.1581>

Geiling, J. (2010). Critical care of the morbidly obese in disaster. *Crit Care Clin* 26, 703-714. <https://doi.org/10.1016/j.ccc.2010.06.001>

Hauner, H., Moss, A., Berg, A., Bischoff, S.C., Colombo-Benkmann, M., Ellrott, T., Heintze, C., Kanthak, U., Kunze, D., Stefan, N., Teufel, M., Wabitsch, M., & Wirth, A. (2014): Interdisziplinäre Leitlinie der Qualität S3 zur „Prävention und Therapie der Adipositas“, 050/001. https://register.awmf.org/assets/guidelines/050-001I_S3_Adipositas_Prävention_Therapie_2014-11-abgelaufen.pdf. Abgerufen 08.06.2024.

Heesen, T. (2005). Prüfung der Reliabilität, Validität und Änderungssensitivität der Kurzform des Funktionsfragebogens Bewegungsapparat (SMFA-D) bei Patienten mit

Gonarthrose und Knieendoprothesenimplantation. Med. Dissertation, Universität Würzburg.

Holle, R. (1995). Methoden zur Konstruktion und Evaluierung klinischer Scores. Habilitationsschrift. Ruprecht-Karls-Universität Heidelberg Institut für Medizinische Biometrie und Informatik Abteilung Medizinische Biometrie.

Hossain, P., Kawar, B. und El Nahas, M. (2007). Obesity and diabetes in the developing world--a growing challenge. *N Engl J Med* 356, 213-215. <https://doi.org/10.1056/NEJMp068177>

Hubert, H.B., Feinleib, M., McNamara, P.M. und Castelli, W.P. (1983). Obesity as an independent risk factor for cardiovascular disease: a 26-year follow-up of participants in the Framingham Heart Study. *Circulation* 67, 968-977. <https://doi.org/10.1161/01.cir.67.5.968>

Jastreboff, A.M., Aronne, L.J., Ahmad, N.N., Wharton, S., Connery, L., Alves, B., Kiyosue, A., Zhang, S., Liu, B., Bunck, M.C., et al. (2022). Tirzepatide Once Weekly for the Treatment of Obesity. *N Engl J Med* 387, 205-216. <https://doi.org/10.1056/NEJMoa2206038>

Khan, S.S., Ning, H., Wilkins, J.T., Allen, N., Carnethon, M., Berry, J.D., Sweis, R.N. und Lloyd-Jones, D.M. (2018). Association of Body Mass Index With Lifetime Risk of Cardiovascular Disease and Compression of Morbidity. *JAMA Cardiology* 3, 280-287. <https://doi.org/10.1001/jamacardio.2018.0022>

Kitahara, C.M., Flint, A.J., Berrington de Gonzalez, A., Bernstein, L., Brotzman, M., MacInnis, R.J., Moore, S.C., Robien, K., Rosenberg, P.S., Singh, P.N., et al. (2014). Association between Class III Obesity (BMI of 40–59 kg/m²) and Mortality: A Pooled Analysis of 20 Prospective Studies. *PLOS Medicine* 11, e1001673. <https://doi.org/10.1371/journal.pmed.1001673>

Knaus, W.A., Draper, E.A., Wagner, D.P. und Zimmerman, J.E. (1985). APACHE II: a severity of disease classification system. *Crit Care Med* 13, 818-829.

Koo, T.K. und Li, M.Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine* 15, 155-163. <https://doi.org/https://doi.org/10.1016/j.jcm.2016.02.012>

Krampen, D.D. (2022): Dorsch Lexikon der Psychologie. <https://dorsch.hogrefe.com/stichwort/testoekonomie#>. Abgerufen 28.06.2024.

Levene, H., I. Olkin, S. G. Ghurye, W. Hoeffding, W. G. Madow, and H. B. Mann (eds), (1960). Robust test for equality of variances. In Contributions to Probability and Statistics: Essays in Honour of Harold Hotelling.

Li, X., Qiu, W., Morrow, J., DeMeo, D.L., Weiss, S.T., Fu, Y. und Wang, X. (2015). A Comparative Study of Tests for Homogeneity of Variances with Application to DNA Methylation Data. *PLoS One* 10, e0145295. <https://doi.org/10.1371/journal.pone.0145295>

Manning, S. (2020). The Crashing Obese Patient. *Emerg Med Clin North Am* 38, 857-869. <https://doi.org/10.1016/j.emc.2020.06.013>

McGraw, K. und Wong, S. (1996). Forming Inferences About Some Intraclass Correlation Coefficients. Psychological Methods 1, 30-46. <https://doi.org/10.1037/1082-989X.1.1.30>

Moiz, A., Levett, J.Y., Filion, K.B., Peri, K., Reynier, P. und Eisenberg, M.J. (2024). Long-Term Efficacy and Safety of Once-Weekly Semaglutide for Weight Loss in Patients Without Diabetes: A Systematic Review and Meta-Analysis of Randomized Controlled Trials. Am J Cardiol 222, 121-130. <https://doi.org/10.1016/j.amjcard.2024.04.041>

Muennig, P., Lubetkin, E., Jia, H. und Franks, P. (2006). Gender and the burden of disease attributable to obesity. Am J Public Health 96, 1662-1668. <https://doi.org/10.2105/ajph.2005.068874>

Müller, T.D. und Blüher, M. (2023). [Obesity treatment: will pharmacotherapies replace metabolic surgery in the future?]. Inn Med (Heidelb) 64, 629-635. <https://doi.org/10.1007/s00108-023-01530-0>

Oprita, B., Aignatoiae, B. und Gabor-Postole, D.A. (2014). Scores and scales used in emergency medicine. Practicability in toxicology. J Med Life 7 Spec No. 3, 4-7.

Pati, S., Irfan, W., Jameel, A., Ahmed, S. und Shahid, R.K. (2023). Obesity and Cancer: A Current Overview of Epidemiology, Pathogenesis, Outcomes, and Management. Cancers (Basel) 15. <https://doi.org/10.3390/cancers15020485>

Pearson, K., Henrici, Olaus Magnus Friedrich Erdmann (1896). VII. Mathematical contributions to the theory of evolution.& III. Regression, heredity, and panmixia. Philosophical Transactions of the Royal Society of London Series A, Containing Papers of a Mathematical or Physical Character 187, 253-318. <https://doi.org/10.1098/rsta.1896.0007>

Polyzos, S.A., Kountouras, J. und Mantzoros, C.S. (2019). Obesity and nonalcoholic fatty liver disease: From pathophysiology to therapeutics. Metabolism 92, 82-97. <https://doi.org/10.1016/j.metabol.2018.11.014>

Powell-Wiley, T.M., Poirier, P., Burke, L.E., Després, J.-P., Gordon-Larsen, P., Lavie, C.J., Lear, S.A., Ndumele, C.E., Neeland, I.J., Sanders, P., et al. (2021). Obesity and Cardiovascular Disease: A Scientific Statement From the American Heart Association. Circulation 143, e984-e1010. <https://doi.org/doi:10.1161/CIR.0000000000000973>

Saarinen, I., Grönroos, S., Hurme, S., Peterli, R., Helmiö, M., Bueter, M., Strandberg, M., Wölnerhanssen, B.K. und Salminen, P. (2023). Validation of the individualized metabolic surgery score for bariatric procedure selection in the merged data of two randomized clinical trials (SLEEVEPASS and SM-BOSS). Surg Obes Relat Dis 19, 522-529. <https://doi.org/10.1016/j.soard.2022.10.036>

Schienkiewitz, A., Kuhnert, R., Blume, M. und Mensink, G.B.M. (2022). Overweight and obesity among adults in Germany - Results from GEDA 2019/2020-EHIS. Journal of Health Monitoring, 21--28.

Schuster, H.P. und Dick, W. (1994). Scoring systems in emergency medicine? Der Anaesthesist 43, 30-35. <https://doi.org/10.1007/s001010050031>

Sharma, A.M. und Kushner, R.F. (2009). A proposed clinical staging system for obesity. *Int J Obes (Lond)* 33, 289-295. <https://doi.org/10.1038/ijo.2009.2>

Sharma, D. und Kibria, B.M.G. (2013). On some test statistics for testing homogeneity of variances: a comparative study. *Journal of Statistical Computation and Simulation* 83, 1944-1963. <https://doi.org/10.1080/00949655.2012.675336>

Sims, E.A., Danforth, E., Jr., Horton, E.S., Bray, G.A., Glennon, J.A. und Salans, L.B. (1973). Endocrine and metabolic effects of experimental obesity in man. *Recent Prog Horm Res* 29, 457-496. <https://doi.org/10.1016/b978-0-12-571129-6.50016-6>

Skulsky, S.L., Dang, J.T., Battiston, A., Switzer, N.J., Birch, D.W., Sharma, A.M. und Karmali, S. (2020). Higher Edmonton Obesity Staging System scores are associated with complications following laparoscopic Roux-en-Y gastric bypass. *Surgical Endoscopy* 34, 3102-3109. <https://doi.org/10.1007/s00464-019-07067-4>

Stier, C., Koschker, A.C., Kim, M., Stier, R., Chiappetta, S. und Stein, J. (2022). Fast-track rescue weight reduction therapy to achieve rapid technical operability for emergency bariatric surgery in patients with life-threatening inoperable severe obesity - A proof of concept study. *Clin Nutr ESPEN* 50, 238-246. <https://doi.org/10.1016/j.clnesp.2022.05.009>

Teasdale, G. und Jennett, B. (1974). Assessment of coma and impaired consciousness. A practical scale. *Lancet* 2, 81-84. [https://doi.org/10.1016/s0140-6736\(74\)91639-0](https://doi.org/10.1016/s0140-6736(74)91639-0)

Welbourn, R., Hollyman, M., Kinsman, R., Dixon, J., Liem, R., Ottosson, J., Ramos, A., Våge, V., Al-Sabah, S., Brown, W., et al. (2019). Bariatric Surgery Worldwide: Baseline Demographic Description and One-Year Outcomes from the Fourth IFSO Global Registry Report 2018. *Obesity Surgery* 29, 782-795. <https://doi.org/10.1007/s11695-018-3593-1>

WHO, Obesity, C.o. und World Health, O. (2000). *Obesity : preventing and managing the global epidemic : report of a WHO consultation* (Geneva: World Health Organization).

Wirtz, M. und Caspar, F. (2002). Beurteilerübereinstimmung und Beurteilerreliabilität: Methoden zur Bestimmung und Verbesserung der Zuverlässigkeit von Einschätzungen mittels Kategoriensystemen und Ratingskalen S. 157-198 (Hogrefe).

WOA, T.L., Rachel Jackson-Leach, Jaynaide Powis, Hannah Brinsden and Maggie Gray (2023): *World Obesity Atals*. <https://www.worldobesity.org/resources/resource-library/world-obesity-atlas-2023>. Abgerufen 04.08.2024.

8 ABBILDUNGSVERZEICHNIS

Abb. 1	Der von Stier et al. (2022) entwickelte Score zur Darstellung der Dringlichkeit zur unmittelbaren stationären Aufnahme eines Patienten mit anschließender Durchführung einer “fast-track rescue weight reduction bridging therapy” (Stier et al., 2022)	12
Abb. 2	Häufigkeitsverteilung der Scores für die Kategorie "age"	18
Abb. 3	Häufigkeitsverteilung der Scores für die Kategorie "gender"	19
Abb. 4	Häufigkeitsverteilung für die Scores der Kategorie "BMI"	20
Abb. 5	Häufigkeitsverteilung für die Scores der Kategorie "functional status"	21
Abb. 6	Häufigkeitsverteilung für die Scores der Kategorie "physical status"	22
Abb. 7	Häufigkeitsverteilung für die Scores der Kategorie "pulmonary status"	23
Abb. 8	Häufigkeitsverteilung für die Scores der Kategorie "diabetes status"	24
Abb. 9	Häufigkeitsverteilung für die Ergebnisse des Gesamtscores	25
Abb. 10.	Gesamtscore- Ergebnisse von Rater 1	48
Abb. 11.	Gesamtscore- Ergebnisse von Rater 2	48
Abb. 12.	Gesamtscore- Ergebnisse von Rater 3	49
Abb. 13.	Gesamtscore- Ergebnisse von Rater 4	49
Abb. 14.	Gesamtscore- Ergebnisse von Rater 5	49

9 TABELLENVERZEICHNIS

Tab. 1 Gewichtsklassifikation nach (WHO et al., 2000)	5
Tab. 2 Nebenerkrankungen des Patientenkollektivs	17
Tab. 3 Häufigkeitsverteilung der Scoreergebnisse für die Kategorie „age“	18
Tab. 4 Häufigkeitsverteilung der Scoreergebnisse für die Kategorie "gender"	19
Tab. 5 Häufigkeitsverteilung der Scoreergebnisse für die Kategorie „BMI“	20
Tab. 6 Häufigkeitsverteilung für die Scoreergebnisse Kategorie "functional status"	21
Tab. 7 Häufigkeitsverteilung der Scoreergebnisse für die Kategorie "physical status"	22
Tab. 8 Häufigkeitsverteilung der Scoreergebnisse für die Kategorie "pulmonary status"	23
Tab. 9 Häufigkeitsverteilung für die Scoreergebnisse der Kategorie "diabetes status"	24
Tab. 10 Deskriptive Darstellung der Scoreergebnisse für den Gesamtscore der Rater	25
Tab. 11 Ergebnisse der Signifikanz des Levenetest	26
Tab. 12 ICC-Interpretation	26
Tab. 13 ICC für die Kategorie "age"	27
Tab. 14 ICC für die Kategorie „gender“	27
Tab. 15 ICC für die Kategorie BMI	27
Tab. 16 ICC für die Kategorie "functional status"	27
Tab. 17 ICC für die Kategorie "physical status"	28
Tab. 18 ICC für die Kategorie „pulmonary status“	28
Tab. 19 ICC für die Kategorie "diabetes status"	28
Tab. 20 ICC für den Gesamtscore	28
Tab. 21 Korrelationen der Variable Alter mit den Scorergebnissen „age“	29
Tab. 22 Korrelationen der Variable Geschlecht mit den Scorergebnissen "gender"	29

Tab. 23 Korrelationen der Variable BMI mit den Scoreergebnissen BMI	29
Tab. 24 Korrelationen der orthopädischen Erkrankungen mit den Scoreergebnissen "functional status"	30
Tab. 25 Korrelationen der pulmonalen Erkrankungen mit den Scoreergebnissen „pulmonary status“	30
Tab. 26 Korrelationen des HbA1c mit den Scoreergebnissen „diabetes status“	30

10 LEBENSLAUF

PERSONALIEN

Name und Vorname: Alaeddin, Ramie Kamal Mohammed

Geburtsdatum: 09.12.1989

Geburtsort: Leipzig

The image consists of a vertical stack of approximately 15 horizontal bars. Each bar is solid black and has a different length. Some bars contain one or more white rectangular voids. The top bar is the longest. Below it, there are two shorter bars, one with a small white gap in its center. The third bar is long and solid black. The fourth bar is also long and solid black. The fifth bar is shorter and solid black. The sixth bar is very long and solid black. The seventh bar is solid black with a white gap in its center. The eighth bar is solid black with a white gap in its center. The ninth bar is solid black with a white gap in its center. The tenth bar is solid black with a white gap in its center. The eleventh bar is solid black with a white gap in its center. The twelfth bar is solid black with a white gap in its center. The thirteenth bar is solid black with a white gap in its center. The fourteenth bar is solid black with a white gap in its center. The fifteenth bar is solid black with a white gap in its center.



11 DANKSAGUNG

Frau PD Dr. Susanne Blank danke ich sehr für die Betreuung der Arbeit und Befähigung zur Promotion.

Frau Dr. Christine Stier danke ich für die Überlassung des Themas dieser Arbeit, der dazugehörigen Patientendaten und die durchgehende Motivation zum wissenschaftlichen Arbeiten.

Besonderen Dank schulde ich Frau PD Dr. Chiapetta, Dr. Rheinwalt, Dr. Pouwels und Frau Dr. Stier für das raten des Patientenkollektivs.

Klara Schönfeld, Jennifer Schmitz und Nicole Ostermann danke ich für die konstruktiven Anmerkungen und das Korrekturlesen.

12 APPENDIX




Abb. 10 Gesamtscore- Ergebnisse von Rater 1




Abb. 11 Gesamtscore- Ergebnisse von Rater 2




Abb. 12 Gesamtscore- Ergebnisse von Rater 3




Abb. 13 Gesamtscore- Ergebnisse von Rater 4




Abb. 14 Gesamtscore- Ergebnisse von Rater 5