Aus der Klinik für Audiologie und Phoniatrie
der Medizinischen Fakultät Charité – Universitätsmedizin Berlin

DISSERTATION

Evaluation des langjährigen Einsatzes von VHI-9i und SUM in der multiparametrischen phoniatrischen Stimmfunktionsdiagnostik

Evaluation of the long-term use of VHI-9i and VEM in multidimensional phoniatric voice function diagnostics

zur Erlangung des akademischen Grades
Doctor rerum medicinalium (Dr. rer. medic.)

vorgelegt der Medizinischen Fakultät
Charité – Universitätsmedizin Berlin

von
Felix Max Ludwig Caffier
aus Berlin

Datum der Promotion: 03.03.2023

Inhaltsverzeichnis

Abkürzungsverzeichnis

2D	2-dimensional
3D	3-dimensional
AUC	Fläche unter der ROC-Kurve (engl. *Area Under the Curve*)
AVA	Automatisierte Vokalanalyse
B	Behauchtheit (engl. *Breathiness*)
CCC	korrekt klassifizierte Fälle (engl. *Correctly Classified Cases*)
DiVAS	Digital Video Archiving and Analysis Software
DSI	Dysphonie-Schweregrad-Index (engl. *Dysphonia Severity Index*)
ELS	Europäische Laryngologische Gesellschaft (engl. *European Laryngological Society*)
FPR	Falsch-positiv-Rate (engl. *False Positive Rate*)
GRB[AS]	Gesamt-Heiserkeit, Rauigkeit, Behauchtheit [, Asthenie, Gepresstheit] (engl. *Grade, Roughness, Breathiness [, Asthenia, Strain]*)
H	Heiserkeit (engl. *Overall Grade of Hoarseness*; G)
J	Youden-Index
MPT	maximale Phonationsdauer (engl. *Maximum Phonation Time*)
R	Rauigkeit (engl. *Roughness*)
RBH	Rauigkeit, Behauchtheit, Gesamt-Heiserkeit (engl. *Overall Grade of Hoarseness, Roughness, Breathiness*; GRB)
ROC	Operationscharakteristik (engl. *Receiver Operating Characteristic*)
SPSS	IBM-Statistiksoftware (engl. *Statistical Package for the Social Sciences*)
SUM	Stimmumfangsmaß (engl. *Vocal Extent Measure*; VEM)
TNR	Richtig-negativ-Rate (engl. *True Negative Rate*)
TOLMS	transorale CO2-Laser-Mikrochirurgie (engl. *CO_2 transoral laser microsurgery*)
TPR	Richtig-positiv-Rate (engl. *True Positive Rate*)
VAPP	Aktivitäts- und Teilhabeprofil (engl. *Voice Activity and Participation Profile*)
VEM	siehe SUM
VHI	Stimmstörungsindex mit 30 Einzelfragen (engl. *Voice Handicap Index*)
VHI-12	Kurzform des VHI (12 Einzelfragen)
VHI-10	Kurzform des VHI (10 Einzelfragen)
VHI-9i	internationale Kurzform des VHI (9 Einzelfragen)
VHIs	Selbsteinschätzung der aktuellen Stimmbeeinträchtigung (engl. *current self-reported vocal impairment*)
VLS	Videolaryngostroboskopie
VoiSS	Stimmsymptomskala (engl. *Voice Symptom Scale*)
VPQ	Stimmleistungsfragebogen (engl. *Vocal Performance Questionnaire*)
VRP	Stimmumfangsprofil (engl. *Voice Range Profile*)
V-RQOL	Fragebogen zur stimmbezogenen Lebensqualität (engl. *Voice-related Quality of Life*)

Abstract

The three studies in this cumulative dissertation deal with multidimensional phoniatric voice function diagnostics and focus on the short forms of the Voice Handicap Index (VHI-9i, VHI-12) and the Vocal Extent Measure (VEM).

The first study addressed the validation of the VHI-9i as a subjective diagnostic tool in voice analysis. The questionnaire was found to be highly reliable. The VHI-9i was successfully validated by correlation with other established vocal parameters and statistical classification methods. The current self-assessed voice impairment (VHIs) proved to be the most appropriate candidate for classification. Based on our results, we recommend the following VHI-9i categories: severity level 0 (no impairment): 0 ≤ 7 points; level 1 (mild impairment): 8 ≤ 16 points; level 2 (moderate impairment): 17 ≤ 26 points; level 3 (severe impairment): 27 ≤ 36 points.

The second study investigated the relationships between the vocal parameters Dysphonia Severity Index (DSI), VEM, VHI-12, and the auditory-perceptual assessment of roughness (R), breathiness (B), and overall grade of hoarseness (G) in patients with various voice disorders. Particular emphasis was placed on the VEM as a new, easy-to-use and potentially more stable parameter than the established DSI. The study showed that all investigated vocal parameters are suitable for evaluating therapy success. The VEM as an objective measure for quantifying vocal performance closely mirrored subjective VHI-12 and GRB assessments.

The third study examined oncological and vocal outcomes after transoral CO_2 laser microsurgery (TOLMS) in T1a vocal fold cancer patients. Therefore, voices were evaluated pre- and post-therapeutically using a multidimensional approach. Oncological treatment results via TOLMS were excellent: The 5-year Kaplan-Meier estimates for recurrence-free, overall, and disease-specific survival were 71.4%, 94.4%, and 100%. Postoperatively, all subjective vocal parameters (VHI-9i, VHIs, GRB) improved significantly. Among the objective parameters, only the VEM improved significantly in the overall cohort, all cordectomy types, and both genders. Additionally, the VEM represented subjective voice impairment the best because of its high correlation with the VHI-9i.

We conclude that VHI-9i and VEM are excellent additions to the established DSI, GRB, and the original VHI questionnaire. Both parameters reliably quantify subjective voice impairment as well as objective vocal performance and are suitable for documenting treatment results. Their widespread use in multidimensional phoniatric voice function diagnostics seems reasonable and desirable, since they expand and enhance current voice diagnostics.

Zusammenfassung

Die drei Studien dieser kumulativen Dissertation widmen sich der multiparametrischen phoniatrischen Stimmfunktionsdiagnostik, wobei der besondere Fokus auf den Kurzfassungen des Voice Handicap Index (VHI-9i, VHI-12) sowie dem Stimmumfangsmaß (SUM) liegt.

Die erste Studie beschäftigte sich mit der Validierung des VHI-9i als subjektives diagnostisches Instrument in der Stimmanalyse. Der Fragebogen stellte sich als äußerst reliabel heraus. Durch die Korrelation mit anderen etablierten Stimmfunktionsparametern und die Klassifikation mittels statistischer Verfahren konnte der VHI-9i erfolgreich validiert werden, wobei sich die Selbsteinschätzung der aktuellen Stimmbeeinträchtigung (VHIs) als geeignetster Kandidat für eine Klassifizierung erwies. Aufgrund unserer Studienergebnisse empfehlen wir die folgende VHI-9i-Klassifikation: Schweregrad 0 (keine Stimmstörung): $0 \leq 7$ Punkte; Schweregrad 1 (geringgradige Stimmstörung): $8 \leq 16$ Punkte; Schweregrad 2 (mittelgradige Stimmstörung): $17 \leq 26$ Punkte; Schweregrad 3 (hochgradige Stimmstörung): $27 \leq 36$ Punkte.

Die zweite Studie untersuchte die Zusammenhänge zwischen den Stimmfunktionsparametern Dysphonia Severity Index (DSI), SUM, VHI-12 und der auditiv-perzeptiven Beurteilung der Rauigkeit (R), Behauchtheit (B) bzw. Gesamtheiserkeit (H) bei Patienten mit verschiedensten stimmlichen Beeinträchtigungen. Besonderes Augenmerk lag dabei auf dem SUM als einem neuen, einfach handhabbaren und potentiell stabileren Parameter im Vergleich zum etablierten DSI. Die Studie zeigte, dass prinzipiell alle untersuchten Stimmfunktionsparameter zur Evaluation des Therapieerfolgs geeignet sind, wobei das SUM als objektives Maß zur Quantifizierung der stimmlichen Leistungsfähigkeit auch die subjektiven Einschätzungen mittels VHI-12 und RBH-Systematik widerspiegelt.

Die dritte Studie befasste sich mit dem onkologischen und stimmbezogenen Outcome nach transoraler CO_2-Lasermikrochirurgie (TOLMS) bei Patienten mit T1a-Stimmlippenkarzinomen. Dazu wurden die Stimmen prä- und posttherapeutisch multiparametrisch untersucht. Mittels TOLMS ließen sich sehr gute onkologische Behandlungsergebnisse erzielen: Die 5-Jahres-Wahrscheinlichkeiten für das rezidivfreie, das Gesamt- und das krankheitsspezifische Überleben nach Kaplan-Meier betrugen 71,4%, 94,4% und 100%. Alle subjektiven Stimmfunktionsparameter (VHI-9i, VHIs, RBH) verbesserten sich postoperativ signifikant. Von den objektiven Parametern verbesserte sich nur das SUM signifikant in der Gesamtkohorte, bei allen Chordektomie-Typen und bei getrennter Betrachtung beider Geschlechter. Außerdem repräsentierte das SUM die subjektive Stimmbeeinträchtigung am besten aufgrund seiner hohen Korrelation mit dem VHI-9i.

Insgesamt kann aus den Studien geschlussfolgert werden, dass VHI-9i und SUM hervorragende Ergänzungen zu den etablierten Parametern DSI, RBH sowie der originalen VHI-Langfassung darstellen. Beide Parameter quantifizieren zuverlässig die subjektive stimmliche Beeinträchtigung bzw. die objektive Leistungsfähigkeit der Stimme und eignen sich zur verlässlichen Dokumentation von Behandlungsergebnissen. Ihr universeller Einsatz in der multiparametrischen phoniatrischen Stimmfunktionsdiagnostik erscheint sinnvoll und erstrebenswert, da sie die aktuelle Stimmdiagnostik erweitern und verbessern.

1. Einleitung

Die menschliche Stimmfunktion ist komplex und erfordert nach dem Protokoll der Europäischen Laryngologischen Gesellschaft (ELS) den Einsatz verschiedener Parameter, um sie möglichst genau und umfassend zu beschreiben [1]. Man unterscheidet dabei im Allgemeinen zwei Arten: Einerseits die subjektiven Parameter, die sowohl durch die Patienten[1] selbst als auch durch geschultes medizinisches Personal mit Hilfe von Fragebögen sowie Stimmaufnahmen erfasst werden, andererseits die objektiven bzw. akustisch-aerodynamischen Parameter, die durch spezielle technische Geräte am Computer aufgezeichnet werden. Subjektive Parameter werden eingesetzt, da die individuell empfundene Einschränkung der Stimmfunktion nicht zwangsläufig mit dem objektiv erhobenen Stimmbefund übereinstimmen muss. Beispielsweise werden Sänger oder Schauspieler, die im Rahmen ihres Berufes Höchstleistungen mit ihrer Stimme erbringen müssen, eine objektiv gesehen leichte Heiserkeit subjektiv anders bewerten als IT-Mitarbeiter oder Rentner. Die Selbstbeurteilung ist also ein wichtiger Bestandteil, um die spezifische Einschränkung der Lebensqualität zu messen und eine maßgeschneiderte Therapie zusammen mit dem Patienten zu planen.

Zur Erfassung der subjektiven Stimmbeeinträchtigung, also des „Voice Handicaps“, haben sich der Voice Handicap Index (VHI) und seine Kurzformen durchgesetzt. Der VHI [2] ist ein etablierter Fragebogen, der physische, psychische und emotionale Aspekte einer Stimmstörung erfasst. Alle 30 Einzelfragen werden auf einer Likert-Skala von 0 bis 4 beantwortet (0: nie, 1: selten, 2: manchmal, 3: oft, 4: immer). Die Gesamtpunktzahl aller Fragen entspricht einer von 4 Dysphonie-Schweregraden (0-14: keine Dysphonie, 15-28: leichte Dysphonie, 29-50: mittelschwere Dysphonie, 51-120: schwere Dysphonie). Da die Beantwortung des VHI verhältnismäßig viel Zeit in Anspruch nehmen kann, wurden Kurzformen entwickelt. Diese sollen die Akzeptanz der psychometrischen Diagnostik bei Patienten und medizinischem Personal erhöhen, indem sie die Bearbeitungszeit verkürzen, ohne an Aussagekraft zu verlieren. In den nachfolgend beschriebenen Studien wurden sowohl der VHI-12 mit 12 Einzelfragen als auch der VHI-9i mit 9 Einzelfragen eingesetzt [3]. Der VHI-12-Gesamtwert von bis zu 48 Punkten wird ebenfalls 4 Schweregraden zugeordnet (0-7: keine Dysphonie, 8-14: leichte Dysphonie,

[1] Aus Gründen der Lesbarkeit werden Personengruppen in einer neutralen Form und wenn möglich im Plural genannt (Ärzte, Patienten, Sänger etc.), wobei immer sowohl weibliche als auch männliche Formen und Personen aller Geschlechter gemeint sind. Ausnahmen bilden jene Passagen im Text, in denen studienbedingt explizit zwischen Männern und Frauen unterschieden werden muss, etwa um Zusammenhänge zwischen Messungen und Geschlecht zu untersuchen. Auf Varianten des Genderns mit Sternen, Doppelpunkten etc. wurde unter anderem deshalb bewusst verzichtet, da sie nicht den amtlichen Rechtschreibregeln genügen und nicht in allen Bundesländern zugelassen sind.

15-22: mittelschwere Dysphonie, 23-48: schwere Dysphonie). Der VHI-10 mit 10 Einzelfragen findet teilweise im amerikanischen Ausland Anwendung, spielte in unseren Studien aber keine Rolle. Er wird allerdings manchmal in Studien als Vergleichsparameter für andere Selbstevaluationsfragenbögen genutzt und soll deshalb nicht unerwähnt bleiben. Die Gesamtpunktzahl des VHI-9i liegt zwischen 0 und 36, jedoch beruht die bisherige 4-stufige Schweregradeinteilung lediglich auf den Perzentilen einer repräsentativen Untersuchung von 716 Patienten [4]. Das Hauptziel der ersten Studie war es deshalb, die Einteilung der Schweregrade zu validieren und die Grenzwerte der einzelnen Schweregradklassen durch statistische Berechnungen entweder zu bestätigen oder anzupassen. Außerdem enthielten unsere Fragebögen zum VHI-12 und VHI-9i die Zusatzfrage „Wie schätzen Sie Ihre Stimme heute ein?“. Der aktuelle Zustand der eigenen Stimme sollte dabei mit nur einer einzigen Zahl, dem sogenannten VHIs, auf einer Skala von 0 bis 3 zusammenfassend bewertet werden (0: normal, 1: leicht gestört, 2: mittelgradig gestört, 3: hochgradig gestört). Des Weiteren wurden alle Probanden nach ihrem Beruf und Stimmgebrauch im Alltag gefragt, um die Stimmbelastung klassifizieren zu können.

Eine andere Methode der subjektiven Stimmevaluation stellt die Beurteilung der Rauigkeit (R), Behauchtheit (B) sowie Gesamtheiserkeit (H) im sog. RBH-System dar [5-7], das aus der GRBAS-Skala hervorgegangen ist [8, 9]. Die Stimme wird hier nicht von den Patienten selbst, sondern durch geschultes ärztliches und paramedizinisches Fachpersonal unabhängig voneinander bewertet. Während der Untersuchungen wurden die Patienten gebeten, den standardisierten Text „Der Nordwind und die Sonne" zu lesen, während die Parameter R, B und H auf einer 4-stufigen Ordinalskala von 0 bis 3 bewertet und anschließend gemittelt wurden (Gruppenurteil). Dabei stellt ein Wert von 0 eine nicht vorhandene, 1 eine geringgradig, 2 eine mittel- und 3 eine hochgradige Beeinträchtigung dar. Der Wert für H wird dabei oft als Goldstandard bzw. Referenz für das gesamte Untersuchungsergebnis verwendet [10].

Für die Bestimmung der objektiven akustisch-aerodynamischen Parameter müssen die Stimmen unter vergleichbaren Bedingungen aufgezeichnet sowie durch spezielle Hard- und Software analysiert werden. Alle Stimmaufnahmen fanden in einem schallgedämmten Labor mit einer Geräuschkulisse < 40 dB(A) statt. Das Headset (XION GmbH, Berlin, Deutschland) hielt ein selbstkalibrierendes Mikrofon (Frequenzgang 70 Hz-20 kHz, Dynamikbereich 40-120 dB(A)) auf einem konstanten Abstand von 30 cm zum Mund [11]. Das Mikrofonsignal wurde digitalisiert und per USB an einen Computer übertragen. Dort erfolgte die Datenverarbeitung entweder durch die Software DiVAS (XION GmbH) oder das Programm LingWAVES (WEVOSYS GmbH, Baunach, Deutschland). Aus den Messungen lassen sich objektive Parameter wie der Dysphonia Severity Index (DSI) und das Stimmumfangsmaß (SUM) berechnen. Der DSI wird aus der höchstmöglich

produzierbaren Grundfrequenz, der niedrigsten Phonationsintensität, der maximalen Phonationsdauer (MPT) sowie dem Jitter berechnet [12]. Die Probanden sollten während der Messung auf /a/ oder /na/ phonieren (ca. 3 Sekunden lang auf einer Tonhöhe in angenehmer Lautstärke), wobei der stabilste von 3 Versuchen für die DSI-Berechnung ausgewählt wurde. Gemäß der Gonnermann-Klassifikation [13] kann das Ergebnis einem von 4 Dysphonie-Schweregraden zugeordnet werden (normal: ≥ 4,2; geringgradig: < 4,2 bis ≥ 1,8; mittelgradig: < 1,8 bis ≥ -1,2; hochgradig: < -1,2).

Da der multidimensionale DSI einerseits durch Extremwerte (z.B. höchste Grundfrequenz, niedrigste Phonationsintensität) sowie andererseits durch Alter und Geschlecht [14-16] beeinflusst werden kann, wurde das eindimensionale SUM entwickelt [17]. Im Gegensatz zum DSI, der die Beeinträchtigung der Stimme beschreibt, stellt das SUM ein Leistungsmaß für die Stimmkapazität dar [18-20]. Das SUM berechnet sich aus dem Verhältnis von Fläche und Umfang des Stimmumfangsprofils (VRP). Der Ablauf der VRP-Messungen erfolgte standardisiert wie in früheren Publikationen beschrieben [21, 22]. Eine kleine bzw. große Stimmkapazität ist durch ein niedriges bzw. hohes SUM gekennzeichnet. Im Normalfall liegen die gemessenen Werte zwischen 0 und 120, können aber durch stark beeinträchtigte oder außergewöhnlich „große“ Stimmen niedriger bzw. höher ausfallen. Das SUM wurde über die proprietäre AVA-Software berechnet [23, 24] und wird in Anlehnung an Müller [22] in vier Klassen eingeteilt. Man unterscheidet eine normale Stimmkapazität (≥ 108) von einer leicht reduzierten (< 108 bis ≥ 93), mäßig (< 93 bis ≥ 69) sowie stark reduzierten (< 69) Stimmkapazität. Während der vor mehr als 2 Jahrzehnten entwickelte DSI in der multiparametrischen Stimmfunktionsdiagnostik weltweit benutzt wird, handelt es sich beim SUM um einen relativ jungen Parameter, der noch nicht international etabliert ist. Umso wichtiger war es deshalb, das SUM in allen 3 Studien dieser Dissertationsschrift anzuwenden und seine Performance insbesondere im Vergleich mit dem DSI zu analysieren.

Neben der subjektiven und objektiven Stimmevaluation wurden die laryngealen Befunde mittels 2D- bzw. 3D-Videolaryngostroboskopie (VLS) erfasst [25, 26], um zwischen organischer und funktioneller Dysphonie zu unterscheiden. Dabei kamen starre transorale oder flexible transnasale hochauflösende Videolaryngoskope mit integriertem Mikrofon (XION GmbH) zum Einsatz. Mittels zuschaltbarer Stroboskopie wurden die phonatorischen Stimmlippenschwingungen visualisiert, wie z.B. eine reduzierte oder aufgehobene Randkantenverschieblichkeit.

Der Schwerpunkt der ersten Arbeit lag auf der Validierung des VHI-9i als diagnostisches Instrument in der subjektiven Stimmfunktionsanalyse. Dazu musste der Fragebogen zuerst auf seine Reliabilität geprüft werden. Durch die Korrelation mit anderen etablierten Parametern und die Klassifikation mittels statistischer Verfahren sollte der VHI-9i anschließend mit Hilfe von

Daten aus über 10 Jahren klinischer Praxis validiert werden. Die zweite Studie untersuchte die Zusammenhänge zwischen den Stimmfunktionsparametern DSI, SUM, VHI-12 und RBH bei Patienten mit stimmlichen Beeinträchtigungen aufgrund verschiedenster Erkrankungen. Besonderes Augenmerk lag dabei auf dem SUM als einem neuen, einfach handhabbaren und potentiell stabileren Parameter im Vergleich zum etablierten DSI. In der dritten Studie sollte das onkologische und stimmbezogene Outcome nach transoraler CO_2-Lasermikrochirurgie (TOLMS) [27-30] bei Patienten mit T1a-Stimmlippenkarzinomen untersucht werden. Dazu wurde prä- und posttherapeutisch eine komplette multiparametrische Stimmdiagnostik nach ELS-Protokoll durchgeführt, erweitert um die Parameter SUM und VHI-9i. Neben der Analyse der Auswirkungen patienten-, tumor- und behandlungsbezogener Faktoren auf die Krankheitskontrolle und das Überleben war die Stimmfunktion das besonders ausführlich untersuchte Ergebniskriterium.

2. Material und Methodik

2.1 Validierung und Klassifizierung des VHI-9i

Im Zeitraum Mai 2009 bis März 2021 wurden insgesamt 17.660 aufeinanderfolgende Fälle dokumentiert. Für die Studie berücksichtigt wurden nur Erstuntersuchungen mit vollständiger multidimensionaler Stimmfunktionsdiagnostik. Jittermessungen > 5% wurden ausgeschlossen, wie in der Literatur empfohlen [31]. Für die statistische Analyse blieben 3661 vollständige und eindeutige Fälle übrig. 416 Patienten wurden gebeten, denselben VHI-9i-Fragebogen ein zweites Mal innerhalb einer Woche auszufüllen, um durch den Vergleich der Antworten die Reliabilität des VHI-9i zu etablieren. Eine zwischenzeitliche therapeutische Intervention fand nicht statt.

Mit der Berufsanamnese wurde die stimmliche Belastung im Berufsalltag eingeschätzt. Nach Koufman und Isaacson [32] werden 4 Stufen mit folgenden Stimmanforderungen unterschieden: Hochleistungsstimmberufe (Stufe 1), Berufssprecher (Stufe 2), Nicht-Berufssprecher (Stufe 3), sowie Berufe ohne besonderen Stimmbedarf (Stufe 4). Die Stimmen wurden außerdem von einem erfahrenen Phoniater und der leitenden Logopädin unabhängig voneinander mittels RBH-System bewertet. Neben der subjektiven Stimmevaluation wurden objektive Stimmumfangsprofil-messungen im Stimmlabor durchgeführt. Die DiVAS- und AVA-Software ermittelten automatisch aus den gemessenen Werten die objektiven Parameter DSI bzw. SUM.

Die statistische Analyse wurde mit IBM SPSS Version 26.0.0.1 durchgeführt. Um die Reliabilität des VHI-9i-Fragebogens nachzuweisen, wurden nur die absoluten Unterschiede in der VHI-9i-Gesamtpunktzahl zwischen Test und Retest verglichen, da Unterschiede in den Einzelfragen für

den klinischen Alltag nicht relevant sind. Die Bias-Untersuchung erfolgte mittels t-Test für verbundene Stichproben. Korrelationen wurden über Pearsons r berechnet. Bei der Untersuchung der Altersabhängigkeit des VHI-9i kam eine Regressionsanalyse zum Einsatz. Die Analyse der geschlechtsspezifischen Unterschiede erfolgte durch t-Tests für unabhängige Stichproben. Der nichtparametrische Kruskal-Wallis H-Test wurde verwendet, um eventuelle Abhängigkeiten vom beruflichen Stimmgebrauch festzustellen.

Nach der Reliabilitätsanalyse musste die Klassifizierung des VHI-9i validiert werden. Korrelationen mit den Parametern VHIs, DSI, VEM, R, B sowie H wurden mittels Spearmans rho (ρ) bestimmt, um den am besten geeigneten Klassifikator zu finden. Die Ausgewogenheit dieses potentiellen Klassifikators in Bezug auf Sensitivität (d.h. Richtig-positiv-Rate, TPR) und Spezifität (d.h. Richtig-negativ-Rate, TNR) wurde anschließend mittels ROC-Kurven (Receiver Operating Characteristic) untersucht, die den Zusammenhang zwischen TPR und Falsch-Positiv-Rate (FPR = 1 - TNR) grafisch darstellt. Da die Grenzen für 4 Schweregrade validiert werden sollten, es sich bei ROC aber um einen binären Klassifikator handelt, mussten die Kurven dreimal erstellt werden. Durch die Berechnung der Fläche unter der Kurve (AUC) konnte die Eignung der Stimmfunktionsparameter als Klassifikator eingeschätzt werden. Als Ausgangspunkt für die ROC-Auswertung wurde der Youden-Index (J) [33] verwendet, der immer dann am höchsten wird (Max J), wenn Sensitivität und Spezifität optimal ausgeglichen sind (J = TPR - FPR = TPR + TNR - 1). Als weiterer Anhaltspunkt wurde der Wert bestimmt, an dem die Anzahl der korrekt klassifizierten Fälle (CCCs) am größten war (Max CCC). Der CCC wird wie folgt berechnet:

CCC = TPR * (Fälle innerhalb der Klasse) + TNR * (Fälle außerhalb der Klassengrenze)

Anschließend wurde der Median zwischen Max J und Max CCC berechnet, um plausible Schwellenwerte zu finden, die sich auch mit unserer jahrzehntelangen klinischen Erfahrung im Einsatz des VHI-9i decken.

2.2 Multiparametrische Diagnostik in der phoniatrischen Stimmbeurteilung

An der zweiten Studie nahmen 152 Patienten mit verschiedensten stimmlichen Beeinträchtigungen teil. Je nach diagnostizierter Pathologie und dem bisherigen Krankheitsverlauf erhielten die Betroffenen entweder einen chirurgischen Eingriff (direkte Mikrolaryngoskopie mit Befundabtragung in Vollnarkose) oder eine konservative Stimmübungstherapie (logopädische Behandlung, 20 Einheiten zu je 45 Minuten, 2x/Woche). Die operativen Interventionen wurden von insgesamt 3 erfahrenen Phonochirurgen der Klinik für Audiologie und Phoniatrie an der

Charité – Universitätsmedizin Berlin durchgeführt, unter totaler intravenöser Anästhesie (TIVA mit Propofol/Remifentanil) und nach einheitlichen operativen Standards [34].
Da sowohl bei der initialen prä-interventionellen Visite als auch 3 Monate nach der Behandlung multiparametrische Daten aufgezeichnet wurden, ergaben sich insgesamt 304 Datensätze für die weitere Analyse. Das Ergebnis der Behandlungen wurde durch den prä- und posttherapeutischen Vergleich von Stimmfunktionsdiagnostik bzw. VLS-Untersuchungen bestimmt. Die digitale VLS definierte das Vorliegen einer organischen oder funktionellen Dysphonie. Auf Grundlage objektiv erfasster akustischer und aerodynamischer Parameter wurden über die Programme LingWAVES und AVA der etablierte DSI [12, 13] bzw. das SUM berechnet [17, 24]. Die subjektive Selbsteinschätzung der Stimme durch die Patienten erfolgte mit Hilfe des VHI-12 Fragebogens [3, 35]. Auch die Gesamtbeeinträchtigung der Stimme (VHIs) sowie das Ausmaß des beruflichen bzw. privaten Stimmgebrauchs wurden ebenfalls über diesen Fragebogen erfasst. Des Weiteren wurde jede Stimmaufnahme durch 3 erfahrene Prüfer (1 Facharzt für Phoniatrie, 1 klinischer Linguist, 1 biomedizinischer Ingenieur) unabhängig voneinander auf der RBH-Skala bewertet. Die Beurteilung der Komponente H als Einschätzung des Gesamtheiserkeitsgrades galt dabei als maßgebliches Kriterium für den Therapieerfolg. Die Parameter H, VHI-12, VHIs, DSI und SUM wurden vor und 3 Monate nach der Intervention miteinander verglichen, ebenso deren prä- und posttherapeutische Veränderungen (Δ). Die statistische Auswertung erfolgte unter Verwendung der Software R (Version 3.6.0). Als statistische Methoden kamen die Berechnung des Spearman'schen Rangkorrelationskoeffizienten (r_s) sowie der t-Test für verbundene Stichproben zum Einsatz. Das Signifikanzlevel wurde bei 5% bzw. $\alpha = 0{,}05$ festgelegt.

2.3 Einsatz des SUM zur Therapieevaluation bei T1a Glottiskarzinomen

Im Zeitraum zwischen Juni 2009 und Oktober 2019 wurden insgesamt 60 konsekutive Patienten mit histologisch bestätigtem pT1a cN0 cM0 Glottiskarzinom erfasst. Personen mit Tis, T1b und T2 Glottiskarzinomen wurden nicht in diese prospektive Studie aufgenommen. 9 Probanden mussten aufgrund unvollständiger Behandlungsdokumentation bzw. nicht wahrgenommener Nachfolgeuntersuchungen ausgeschlossen werden, sodass 51 Fälle übrigblieben. Die Untersuchungen fanden am Tag vor der laserchirurgischen Intervention sowie 3 Monate nach in-sano Resektion und abgeschlossener Wundheilung statt. Die TOLMS wurde in Mikrolaryngoskopie unter Verwendung des AcuPulse 30W/40 ST CO_2-Lasersystems in Vollnarkose durchgeführt. Dabei wurden etablierte Sicherheitsvorkehrungen eingehalten, z.B. Verwendung von Schutzbrillen, Lasertuben, feuchten Abdeckungen, Beatmung mit O2-Konzentration < 40 %. Durch die Applikation einer subepithelialen Infusion mit NaCl-Lösung konnten mögliche tumoröse

Infiltrationen in tiefere Strukturen beurteilt und das gesunde umliegende Gewebe vor thermischer Schädigung geschützt werden. Die Laser-Chordektomien erfolgten unter Einhaltung eines Sicherheitsabstandes ≤ 1 mm, wobei Randschnitte bei unsicherer vollständiger Tumorentfernung entnommen wurden. Der jeweilige Resektionstyp wurde gemäß ELS-Protokoll [36] bestimmt. Die histopathologische Untersuchung erfolgte entsprechend der Richtlinien des American Joint Committee on Cancer (AJCC) für das Tumor-Staging [37]. Patienten mit R1-Status wurden für eine Nachresektion vorgesehen.

Die Interventionsergebnisse wurden anhand prä- und postoperativer VLS sowie Stimmfunktionsdiagnostik gemäß ELS-Protokoll analysiert [1, 38]. Die auditiv-perzeptive Beurteilung der Stimmaufnahmen erfolgte mittels RBH-System, die subjektive Selbsteinschätzung der Stimmbeeinträchtigung mittels VHI-9i-Fragebogen (v.a. Gesamtpunktzahl, VHIs). VRP-Messungen und akustisch-aerodynamische Analysen wurden mit Hilfe der DiVAS-Software realisiert, wobei insbesondere die Parameter MPT, Jitter, DSI und SUM erfasst bzw. berechnet wurden. Die statistische Auswertung erfolgte unter Verwendung der Software R (Version 4.0.1). Stärke und Richtung des Zusammenhangs zwischen prä- und postoperativen Messungen wurden mittels Spearman-Rangkorrelation (r_s) ermittelt. Eine etwaige Verbesserung der Stimmfunktionsparameter wurde über den Wilcoxon Signed-Rank-Test geprüft. Die Analyse der Auswirkungen patienten-, tumor- und behandlungsbezogener Faktoren auf die Krankheitskontrolle und das Überleben erfolgte mittels Kaplan-Meier-Methode [39].

3. Ergebnisse

3.1 Validierung und Klassifizierung des VHI-9i

Bei den 416 Test-Retest-Paaren der Reliabilitätskohorte lag das Durchschnittsalter bei 50 ± 17 Jahren (mean ± SD), wobei Männer (56 ± 16) im Allgemeinen älter waren als Frauen (46 ± 17). 253 Studienteilnehmer hatten Berufe ohne besonderen Stimmbedarf (Stufe 4; 60,8%), 78 waren Nicht-Berufssprecher (Stufe 3; 18,7%), 59 Berufssprecher (Stufe 2; 14,2%), während lediglich 26 einen Hochleistungsstimmberuf ausübten (Stufe 1; 6,3%). Der durchschnittliche Zeitabstand zwischen dem Ausfüllen des ersten und zweiten Fragebogens betrug 3,3 Tage (Median 2 Tage). Der mittlere Unterschied im VHI-9i-Gesamtwert betrug lediglich 0,25 ± 3,52 Punkte (mean ± SD). Der t-Test für verbundene Stichproben ergab keine signifikanten Unterschiede zwischen den Test-Retest-Ergebnissen ($p = 0{,}146$). Beide Gesamtwerte korrelierten sehr gut miteinander ($r = 0{,}919$, $p < 0{,}01$). In nur 5% der Fälle zeigte sich eine Differenz von mehr als 7 Punkten. Das

Konfidenzintervall von 7 kann damit als Mindestgröße für eine signifikante Änderung der Punktzahl der Selbsteinschätzung, z. B. nach Interventionen, angenommen werden. Weder das Geschlecht ($p = 0,589$) noch der berufliche Stimmgebrauch ($p = 0,701$) hatten Einfluss auf die Zuverlässigkeit des Fragebogens. Allerdings zeigte sich eine leichte Altersabhängigkeit, wobei für jedes Lebensjahr die Differenz um 0,016 Punkte stieg ($p = 0,028$).

Von den 3661 Teilnehmern der Validierungskohorte mit vollständiger multidimensionaler Stimmdiagnostik waren 1456 männlich (39,8%) und 2205 weiblich (60,2%). Das Durchschnittsalter betrug 48 ± 17 Jahre, wobei Männer etwas älter waren als Frauen (50 ± 18 vs. 47 ± 17). Da bereits gezeigt wurde, dass die Reliabilität des VHI-9i-Fragebogens nicht von Geschlecht oder beruflicher Stimmbelastung abhängt und auch das Alter nur eine vernachlässigbare Rolle spielt, wurden diese Aspekte in den weiteren Analysen nicht gesondert berücksichtigt. Gemäß der bisherigen perzentilbasierten VHI-9i-Klassifikation hatten 15,5% der Probanden gesunde Stimmen ($0 \leq 5$ Punkte), 25,7% eine leichte ($6 \leq 13$), 32,3% eine mittelgradige ($14 \leq 22$) und 26,5% eine hochgradige Dysphonie ($23 \leq 36$). Teilt man die aktuelle Studiendatenbank ebenfalls nach Perzentilen ein, so zeigt sich ein anderes Bild: 25% der Kohorte hatten 0-9 Punkte, 50% bis 16 Punkte und 75% bis zu 22 Punkte.

Der VHI-9i korrelierte am stärksten mit dem VHIs, auch wenn die Beziehung zwischen beiden nur moderat war ($\rho = 0,592$). Die Korrelation des VHI-9i mit allen anderen Parametern war deutlich schwächer. DSI und SUM korrelierten ebenfalls moderat miteinander ($\rho = 0,663$). Die Verteilung der Probanden auf die Schweregrade der Parameter H und R war recht ähnlich, lediglich das Ergebnis für B unterschied sich, da hier über 50% aller Fälle als "gesund" eingestuft wurden. H und R wiesen die stärkste Korrelation untereinander auf ($\rho = 0,871$).

Die VHI-9i-Gesamtpunktzahl wurde anschließend mit den Schweregraden für VHIs, DSI, VEM und H in Beziehung gesetzt. Grundsätzlich stiegen mit wachsender VHI-9i-Punktzahl auch die Schweregrade, wobei sich deren Grenzen aufgrund mangelnder Trennschärfe überschnitten. Das traf insbesondere für DSI und SUM zu, weshalb beide Parameter für eine Klassifizierung des VHI-9i ungeeignet scheinen. Die ROC-Analyse zeigte, dass der VHIs am besten für die Klassifizierung geeignet ist, da die AUC-Werte hier am höchsten waren.

Gemäß der Reliabilitätsanalyse sollte jeder Schweregrad mindestens einen Bereich von 7 Punkten umfassen. Leider konnten weder bei der Klassifizierung mittels Max J noch mittels Max CCC Bereichsgrenzen gefunden werden, die dieses Kriterium erfüllten. Deshalb wurden zusätzlich die Mediane zwischen Max J und Max CCC berechnet. Abhängig davon, welcher der Stimmfunktionsparameter VHIs, DSI, VEM oder H als potentieller Klassifikator diente, unterschieden sich die damit gefundenen potentiellen neuen Grenzen der VHI-9i-Schweregrade. Die erste Grenze, die die

Schweregrade 0 und 1 trennt, lag einheitlich zwischen 7 und 8 Punkten. Die Grenze zwischen den Schweregraden 1 und 2 war weniger einheitlich (zwischen 14 und 20 Punkten), weshalb dieser zweite Grenzwert nicht durch die ROC-Klassifizierung bestimmt werden konnte, sondern anhand des 50%-Quartils festgesetzt wurde (16 Punkte). Die Mediane der Grenze zwischen den Schweregraden 2 und 3 lagen - abgesehen vom SUM - wieder relativ einheitlich zwischen 26 und 28 Punkten. Abgesehen vom zweiten Grenzwert konnten alle Schweregradobergrenzen durch den VHIs J-Median definiert werden.

3.2 Multiparametrische Diagnostik in der phoniatrischen Stimmbeurteilung

Von den 152 Studienteilnehmern waren 102 weiblich und 50 männlich, wobei beide Geschlechter in Bezug auf soziodemografische Merkmale, diagnostizierte laryngeale Pathologien sowie Verteilung der unterschiedlichen Heiserkeitsgrade (H) vergleichbar waren. Das Durchschnittsalter (mean ± SD) lag bei 50 ± 17 Jahren, wobei Männer (43 ± 17) im Allgemeinen etwas jünger waren als Frauen (45 ± 15). 86 Patienten gaben an, in ihrem Beruf eine hohe stimmliche Belastung zu haben (z.B. Lehrer, Schauspieler, Sänger), 66 dagegen eine geringe (z.B. Büroangestellte, IT-Spezialisten, Arbeiter). Bei 101 Studienteilnehmern (66 Frauen, 35 Männer) ergab die VLS organisch-strukturelle Veränderungen. 51 Probanden (36 Frauen, 15 Männer) hatten eine normale Kehlkopfanatomie, litten aber an einer funktionellen Dysphonie, die durch eine Überlastung der Stimme hervorgerufen wurde. Prätherapeutisch wiesen 46 Probanden keine Heiserkeit (H0) auf, wobei die spezifische Analyse zeigte, dass sich darunter 13 Patienten mit kleinen glottalen Veränderungen befanden, 4 mit arytenoidalen Pathologien und 29 mit funktioneller Dysphonie.

Bei allen untersuchten Stimmfunktionsparametern ließen sich posttherapeutische Verbesserungen nachweisen. In der subjektiven auditiv-perzeptiven Analyse waren die Stimmen im Mittel signifikant weniger rau, behaucht und heiser ($p < 0,001$). Der maßgebliche Referenzwert H sank von 1,2 ± 0,9 auf 0,7 ± 0,7 (mean ± SD). Dabei reduzierte sich die Heiserkeitsbeurteilung bei den meisten Patienten um einen Heiserkeitsgrad, bei 29 Probanden von H1 auf H0 bzw. bei 22 Probanden von H2 auf H1. Die Veränderungen im VHI-12 waren ebenfalls signifikant ($p < 0,001$). Während sich die mittlere Gesamtpunktzahl von 15 ± 12 auf 10 ± 8 Punkte verringerte, verbesserte sich der mittlere VHIs-Wert von 1,5 ± 0,8 auf 0,6 ± 0,5. Seitens der objektiven Parameter verzeichnete der DSI einen signifikanten mittleren Anstieg von 2,2 ± 4,2 auf 3,7 ± 2,3 ($p < 0,001$), wobei sich beide Werte auf dem Niveau einer geringgradigen Dysphonie bewegen. Außerdem vergrößerte sich das SUM von 60 ± 53 auf 79 ± 32, was einer wesentlichen Verbesserung der Stimmkapazität entspricht ($p < 0,001$). Während Alter und Geschlecht keinen relevanten Einfluss auf das Behandlungsergebnis hatten, spielte die Art der therapeutischen Intervention eine große

Rolle. Im Vergleich zeigte sich, dass konservative Therapien zur qualitativen Wiederherstellung der Stimmfunktion führten, jedoch quantitativ kleinere Verbesserungen lieferten. Dagegen hatten die phonochirurgischen Interventionen größeren Einfluss auf die Stimmfunktion, da sie mit einer messbar höheren Verbesserung fast aller Parameter einhergingen.

Die Stimmfunktionsparameter H, VHI-12, VHIs, DSI und SUM korrelierten signifikant miteinander ($p < 0,01$), wobei die Stärke der linearen Beziehungen nur schwach bis mäßig ausgeprägt wer. Die Korrelationskoeffizienten änderten sich im Allgemeinen posttherapeutisch nur geringfügig, mit Ausnahme des DSI, dessen Korrelationen mit H bzw. SUM nach der Behandlung deutlich abnahmen. Die schwachen Beziehungen zwischen DSI und VHI-12 bzw. VHIs zeigten posttherapeutisch keine relevanten Veränderungen. Dagegen korrelierte das SUM sowohl mit dem VHI-12 als auch mit H mäßig bis stark negativ, unabhängig vom Therapiestatus. Die Untersuchung der therapiebedingten Veränderungen (Δ) ergab einen moderaten Zusammenhang zwischen ΔDSI und ΔSUM ($r_s = 0,5$; $p < 0,01$). Während ΔDSI bezüglich der subjektiven Parameter nur schwach mit ΔH korrelierte, waren die Beziehungen zu ΔVHI-12 bzw. ΔVHIs nicht signifikant. Im Gegensatz dazu korrelierte ΔSUM stärker mit ΔH und zeigte auch in der Analyse mit ΔVHI-12 bzw. ΔVHIs signifikante Zusammenhänge.

3.3 Einsatz des SUM zur Therapieevaluation bei T1a Glottiskarzinomen

Bei den 51 Patienten (43 männlich, 8 weiblich) lag das Durchschnittsalter (mean ± SD) bei 65 ± 12 Jahren, wobei Männer (68 ± 10) im Allgemeinen älter waren als Frauen (52 ± 14). Hinsichtlich des Noxenabusus wurde Tabak- sowie Alkoholkonsum von 39 bzw. 43 Patienten angegeben. Wie anhand der präoperativen VLS festgestellt werden konnte, zeigten die Läsionen keine Seitenpräferenz: In 28 Fällen war die rechte, in 23 Fällen die linke Stimmlippe befallen. Die gesamte Stimmlippe war bei 26 Patienten involviert, während in 14 bzw. 11 Fällen lediglich zwei Drittel bzw. ein Drittel der Stimmlippenlänge betroffen war. Überwiegend zeigten sich exophytische ($n = 29$) bzw. hyperkeratotisch flache Befunde ($n = 20$), nur 2 Läsionen waren ulzerierend. Neben fehlender oder reduzierter phonatorischer Stimmlippenbeweglichkeit bestand insbesondere bei exophytischem Tumorwachstum eine Glottisschluss-Insuffizienz.

Präoperativ bestanden stimmlich mittelgradige Beeinträchtigungen, die sich sowohl in den subjektiven Parametern der auditiven Perzeption (im Mittel R2 B1 H2) und der Selbstbeurteilung (VHI-9i 18 ± 8), als auch in den objektiven akustisch-aerodynamischen Parametern widerspiegelten (SUM 64 ± 33; DSI 1,2 ± 2,4; MPT 13 ± 6 s). Die Korrelation zwischen SUM und DSI war signifikant ($r_s = 0,51$), wobei das SUM stärker mit den subjektiven Parametern VHI-9i und RBH korrelierte. Bei 24 Patienten wurde eine subepitheliale Chordektomie (Typ I), bei 18

Patienten eine subligamentäre (Typ II) und bei 9 Patienten eine transmuskuläre Chordektomie (Typ III) mittels TOLMS durchgeführt. Das Gewebe der pT1a Plattenepithelkarzinome war histopathologisch in 15 Fällen gut differenziert (G1), in 34 mäßig (G2) und in 2 Fällen schlecht differenziert (G3). Durch die Primäroperation wurde das pT1a bei 29 Patienten vollständig entfernt (R0). In 22 Fällen wurde bei unklarem Schnittrand mit fraglicher R1-Situation eine zweite Exzision indiziert, wobei 17 Patienten in der Nachresektion keinen Restbefund aufwiesen. In den verbleibenden 5 Fällen lagen maligne Läsionen vor (3 pT1, 1 Tis, 1 SIN III), die im Rahmen der zweiten TOLMS vollständig entfernt werden konnten.

Verglichen mit den präoperativen Befunden war drei Monate nach in-sano Resektion und vollständiger Heilung eine deutliche Verbesserung der Stimmfunktion zu verzeichnen. In der RBH-Klassifizierung wurden die Stimmen als weniger rau ($R_{prä}$ 1,8 ± 0,7 vs. R_{post} 1,2 ± 0,7), behaucht ($B_{prä}$ 1,0 ± 0,6 vs. B_{post} 0,6 ± 0,6) und heiser ($H_{prä}$ 1,9 ± 0,7 vs. H_{post} 1,3 ± 0,7) beurteilt. Die mittlere VHI-9i Gesamtpunktzahl verringerte sich von 18 ± 8 auf 9 ± 9 Punkte, und auch der mittlere VHIs-Wert verbesserte sich (2 ± 1 vs. 1 ± 1). Sämtliche subjektiven Verbesserungen zeigten sich sowohl für die Gesamtkohorte als auch für alle Chordektomie-Typen signifikant. In Bezug auf die objektiven Parameter war das SUM ebenfalls in der Gesamtpopulation (64 ± 33 vs. 83 ± 31) sowie bei sämtlichen Chordektomie-Typen signifikant verbessert. Im Gegensatz dazu erreichten weder die DSI-Zunahme (1,2 ± 2,4 vs. 1,5 ± 2,3) noch die Jitter-Abnahme (0,9 ± 1,1 vs. 0,6 ± 0,4) in der Gesamtkohorte das Signifikanzniveau, lediglich bei Frauen sowie Chordektomie-Typ III (jeweils $p < 0,05$). SUM und DSI korrelierten negativ mit dem VHI-9i (r_s = -0,29 vs. r_s = -0,11), wobei nur die Korrelation zwischen VHI-9i und SUM signifikant war ($p < 0,05$). Die MPT zeigte unspezifische, ungerichtete Veränderungen ohne jegliche Signifikanz.

Während des postoperativen follow-up von durchschnittlich 45 ± 26 Monaten traten bei 10 Patienten Lokalrezidive auf, in 2 weiteren Fällen sekundäre pT1a Glottiskarzinome auf der kontralateralen Stimmlippe. Alle rezidivierenden und sekundären Malignome wurden stadiengerecht erfolgreich behandelt. 2 Patienten verstarben (einer interkurrent, einer aufgrund eines sekundären Pankreaskarzinoms). Die 5-Jahres-Wahrscheinlichkeiten für das rezidivfreie, das Gesamt- und das krankheitsspezifische Überleben nach Kaplan-Meier betrugen 71,4%, 94,4% und 100%.

4. Diskussion

Zusammenfassend zeigten die Analysen des langjährigen Einsatzes von VHI-9i und SUM an der Klinik für Audiologie und Phoniatrie der Charité – Universitätsmedizin Berlin, dass beide

Parameter für die subjektive bzw. objektive Beurteilung der Stimmfunktion sehr gut geeignet sind. Unsere erste Studie [40] demonstrierte, dass der VHI-9i ein reliabler Fragebogen ist, der sich nicht umsonst seit mehr als zehn Jahren als wertvolles Diagnoseinstrument bewährt hat. Die hohe Akzeptanz bei Patienten und Personal zeigt auch die hohe Zahl von 17.660 konsekutiv ausgefüllten Fragebögen. Allerdings sollte die bisherige, im klinischen Alltag praktizierte perzentilbasierte Einteilung an die aktuellen Ergebnisse der statistischen Analyse angepasst werden. Die Test-Retest-Analyse zeigte, dass die Reliabilität des Fragebogens unabhängig vom Geschlecht und der beruflichen Stimmbelastung ist. Auch das Alter hatte für den klinischen Gebrauch keinen relevanten Einfluss: Für jedes Lebensjahr erhöhte sich das Retest-Ergebnis um 0,016 Punkte, was hochgerechnet einer Differenz von 1 Punkt zwischen einem Jugendlichen und einem älteren Menschen entspricht. Die Reliabilitätsanalyse zeigte auch, dass für den VHI-9i ein Unterschied von mindestens 7 Punkten eine überzufällige Veränderung anzeigt. Der Bereich der Schweregrade muss also mindestens 7 Punkte umfassen, um die Möglichkeit von Retest-Artefakten zu minimieren. Bei der Validierung mittels ROC-Analyse führte weder die Optimierung für Max J noch für Max CCC zu einer Klassifizierung, die diese Anforderung erfüllte. Die Berechnung des J-Medians ergab jedoch zufriedenstellende Ergebnisse für die klinische Anwendung.

Der VHIs erwies sich als geeignetster Kandidat für eine Validierung des VHI-9i, da die Korrelation zwischen beiden Parametern am stärksten war. DSI, SUM und RBH spielten letztlich keine Rolle in unserer Empfehlung für die überarbeitete VHI-9i-Klassifizierung, weil sie schwächer mit dem VHI-9i korrelierten, nicht praktikable Schweregradgrenzen und eine schlechtere Trennschärfe aufwiesen. Das bestätigt auch die Ergebnisse aktueller Studien, die zeigen, dass o.g. Stimmparameter verschiedene Aspekte der Stimmfunktion eines Patienten messen, nicht redundant sind, sondern sich gegenseitig ergänzen [19, 22, 41]. Die neue Grenze von 7 Punkten für den niedrigsten Schweregrad 0 entspricht direkt dem J-Median-Ergebnis des VHIs für gesunde Stimmen. Die Obergrenze für den Schweregrad 1 entspricht mit 16 Punkten dem Wert des 50%-Quartils, da kein anderer Klassifikator geeignete Werte lieferte. Schweregrad 2 wird nach oben mit 26 Punkten durch den VHIs J-Median begrenzt. Aufgrund dieser statistischen Ergebnisse empfehlen wir folgende neue VHI-9i-Klassifikation:

- Schweregrad 0 (keine Stimmstörung): $0 \leq 7$ Punkte
- Schweregrad 1 (geringgradige Stimmstörung): $8 \leq 16$ Punkte
- Schweregrad 2 (mittelgradige Stimmstörung): $17 \leq 26$ Punkte
- Schweregrad 3 (hochgradige Stimmstörung): $27 \leq 36$ Punkte

Neben dem VHI und seinen Kurzformen können eine Reihe weitere Fragenbögen zur

Selbstevaluation der Stimmfunktion eingesetzt werden. Dazu zählt beispielsweise der besonders in englischsprachigen Raum verbreitete und validierte Voice-related Quality of Life Fragebogen (V-RQOL), der dem VHI-9i in Fragenanzahl und Bewertungsschema recht ähnlich ist [42]. VHI und V-RQOL korrelieren zwar stark, sind allerdings durch den unterschiedlichen Fokus der Einzelfragen nicht direkt miteinander vergleichbar [43]. Ähnlich verhält es sich mit dem Voice Activity and Participation Profile (VAPP), das ebenfalls eine starke Korrelation mit dem VHI aufweist [44]. Das VAPP benötigt zudem deutlich mehr Einzelfragen als der VHI-9i, was den diagnostischen Prozess unnötig in die Länge zieht. Diesen Nachteil teilt das VAPP mit der 2003 entwickelten Voice Symptom Scale (VoiSS) [45], weshalb der VHI-9i aus Effizienzgründen zu bevorzugen ist. Eine hohe Korrelation besteht außerdem zwischen dem VHI-10 und dem Vocal Performance Questionnaire (VPQ) [46]. Grundsätzlich scheinen alle genannten Fragebögen für die klinische Diagnostik geeignet zu sein, wobei sie durch ihre unterschiedlich fokussierten Einzelfragen abweichende Individualergebnisse liefern können. Kurze Fragebögen wie der VHI-9i haben einen entscheidenden Zeitvorteil, ohne an klinisch relevanter Aussagekraft zu verlieren.

Die zweite Studie [10] zeigte, dass sich in unserer multiparametrischen Stimmfunktionsdiagnostik alle untersuchten Parameter nach erfolgreicher konservativer oder operativer Stimmtherapie verbesserten und daher prinzipiell zur Evaluation des Therapieerfolgs geeignet sind. Insgesamt bestätigten unsere Daten die Ergebnisse anderer Studien, in denen Behandlungseffekte bei Patienten mit verschiedenen Stimmproblemen untersucht wurden [18, 47-51]. Wie erwartet hatte die Phonomikrochirurgie den größten Einfluss auf die Verbesserung der Stimmfunktion, wobei insbesondere bei Sängern mit funktioneller Dysphonie auch geringere quantitative Verbesserungen durch konservative Therapien zur Wiedererlangung der künstlerischen Fähigkeiten führten. Wie in der Literatur beschrieben, profitierten bei organisch-strukturellen Dysphonien vor allem jüngere Patienten mit kurzer Krankheitsdauer und eher kleinen, überlastungsbedingten Veränderungen der Lamina propria von konservativer Stimmtherapie [52-54].
Bei genauerer Analyse der einzelnen Stimmparameter bestätigte unsere Arbeit u.a. frühere Studien zur Eignung des DSI für die Beschreibung des Therapieerfolgs [55-58]. Da der DSI die Dysphonie als Negativkriterium beschreibt und durch seine multiparametrische Berechnung fehleranfällig ist [14-17], sollten auch alternative Parameter wie das VRP-basierte SUM angewendet werden, die weniger störanfällig sind und die stimmliche Leistungsfähigkeit als Positivkriterium erfassen. Im Therapieverlauf verbesserten sich die SUM-Werte unserer Patienten deutlich, was sich mit den wenigen Studien deckt, die diesen relativ neuen Parameter bereits eingesetzt haben und signifikant höhere posttherapeutische SUM-Werte beobachteten [18-20].

Auf Seiten der subjektiven Parameter konnte der VHI-12 überzeugen, da er die stimmliche Beeinträchtigung im Rahmen der Selbstbeurteilung durch die Patienten nachvollziehbar wiedergab. Unsere Daten korrespondierten mit den Ergebnissen zahlreicher Studien, die verschiedene Versionen des VHI-Fragebogens zur Untersuchung des Behandlungserfolgs bei organischen bzw. funktionellen Dysphonien einsetzten [47, 48, 59, 60]. VHI-Kurzformen stellen damit eine von Patienten und medizinischem Personal akzeptierte, effiziente und praktikable Lösung zur Selbstevaluation dar [3, 61]. Auch die auditiv-perzeptive Beurteilung der Stimme mittels RBH-System kann als zuverlässig gelten, insbesondere wenn ein Gruppenurteil durch erfahrenes medizinisches Fachpersonal gefällt wird [7, 19, 62]. In unserer Studie korrelierte der maßgebliche Referenzwert H besser mit dem SUM als mit dem DSI, weshalb der SUM die professionelle Expertenevaluation am besten widerspiegelt. Der signifikante negative Zusammenhang bedeutet, dass das SUM mit abnehmender Heiserkeit zunimmt. Für den DSI konnte kein signifikanter Zusammenhang mit dem VHI-12 gefunden werden, sowie nur eine schwache negative Korrelation mit H, die nach der Therapie sogar noch abnahm. Damit scheint der DSI zwar geeignet zu sein, den Therapieerfolg anzuzeigen, kann aber die subjektiv erlebte Verbesserung des Stimmproblems nur schlecht wiedergeben. Dagegen standen die therapieinduzierten Veränderungen des SUM mit denen des VHI-12 in einem stärkeren Zusammenhang. Somit wird das SUM als objektives Maß zur Quantifizierung der stimmlichen Leistungsfähigkeit zusätzlich durch die subjektive Selbsteinschätzung und die auditiv-perzeptive Expertenbeurteilung validiert.

Die dritte Studie [63] bestätigte die bisher bekannten Forschungsergebnisse, dass die transorale CO_2-Lasermikrochirurgie mit einem sehr guten onkologischen Outcome bei T1a Glottiskarzinomen einhergeht [29, 64-66]. Im stimmbezogenen Outcome zeigten alle Stimmfunktionsparameter postoperative Veränderungen. Wie in der Literatur beschrieben, war die individuelle Stimmfunktion nach TOLMS je nach Ausmaß des präoperativen Tumorbefundes besser, ähnlich oder leicht eingeschränkt [67-69]. Im Allgemeinen verbessert sich die Stimmqualität während der langfristigen Nachsorge [70, 71], was entsprechend der Beobachtungen dieser Studie vor allem daran liegt, dass narbige Veränderungen im Stimmlippenbereich nach Monaten wieder schwingungsfähig werden. Das in der Literatur angewendete Pooling der Resektionstypen erscheint dagegen unseres Erachtens nicht sinnvoll, wie das unterschiedliche Outcome in den Chordektomie-Subgruppen zeigt. Unsere Ergebnisse bestätigten des Weiteren, dass der Fokus auf den Stimmerhalt die Rate der Kontrollmikrolaryngoskopien mit umschriebenen Nachresektionen erhöhen kann, wenn bei der TOLMS-Primäroperation ein sehr kleiner Sicherheitsabstand intendiert wird, der in histologisch unsicheren Tumorrändern resultiert [72, 73]. Im Einklang mit

der Literatur zeigte sich, dass Re-Operationen durch engmaschige Überwachung des Lokalbefundes mittels VLS entweder vermieden oder rechtzeitig indiziert werden können [74, 75].
Die subjektiven Stimmfunktionsparameter RBH, VHI-9i und VHIs verbesserten sich postoperativ grundsätzlich signifikant, was den Daten der Literatur entspricht [67, 76, 77]. Bei den objektiven akustisch-aerodynamischen Maßen zeigten unsere Ergebnisse ein differenzierteres Bild. Die unspezifischen, ungerichteten und nicht signifikanten Veränderungen der MPT stimmen mit Hamzany et al. überein [71], woraus geschlossen werden kann, dass der Einsatz aerodynamischer Parameter für die Outcome-Bewertung bei T1a Glottiskarzinomen nicht sinnvoll ist. Im Gegensatz dazu erschienen die objektiven akustischen Parameter unserer Untersuchungen sehr gut geeignet, die postoperative Stimmfunktion zu beurteilen. Dies ist jedoch nicht unumstritten, da sich in der Literatur sowohl Nachweise für Verschlechterungen [67], Verbesserungen [69, 78, 79] als auch keine relevanten Veränderungen [68, 71] finden lassen. Während das SUM in der Stimmdiagnostik als vergleichsweise neuer Parameter noch nicht weit verbreitet ist, stellt der DSI eine etablierte diagnostische Alternative dar. Der DSI kann jedoch leicht durch z.B. Alter oder Geschlecht beeinflusst werden [14, 15], was auch in der vorliegenden Studie bestätigt wurde. Im Gegensatz dazu zeigte sich das SUM weniger störanfällig gegenüber DSI-relevanten Extremwerten (z.B. höchste Frequenz, niedrigste Intensität), und auch der Einfluss des Alters scheint geringer zu sein. SUM und DSI korrelierten signifikant miteinander, wobei das SUM die subjektiven stimmlichen Beeinträchtigungen aufgrund der höheren Korrelationen mit dem VHI-9i besser widerspiegelte. Das SUM verbesserte sich als einziger objektiver Stimmparameter sowohl in der Gesamtkohorte, bei allen Chordektomie-Typen als auch bei getrennter Betrachtung beider Geschlechter signifikant. Dieser Parameter erwies sich daher als besonders prädestiniert dafür, die stimmliche Leistungsfähigkeit nach T1a-Exzision zu quantifizieren.
Zusammenfassend für die dritte Studie ist festzustellen, dass die Berücksichtigung phonochirurgischer Prinzipien bei transoraler CO_2-Lasermikrochirurgie von T1a Glottiskarzinomen sehr gute onkologische und stimmbezogene Ergebnisse liefert. Das SUM als sensitiver objektiver Parameter spiegelt die subjektive Selbsteinschätzung am besten wider. Die Stimmdiagnostik nach ELS-Protokoll ist zeitaufwändiger, wobei dieser Mehraufwand im Rahmen einer evidenzbasierten Therapie gerechtfertigt ist und sich als notwendig für den Nachweis des Stimmerhalts erweist.

Insgesamt kann aus unseren Arbeiten geschlussfolgert werden, dass VHI-9i und SUM hervorragende Ergänzungen zu den etablierten Parametern DSI, RBH sowie der originalen VHI-Langfassung darstellen. Beide Parameter quantifizieren zuverlässig die subjektive stimmliche

Beeinträchtigung bzw. die objektive Leistungsfähigkeit der Stimme und eignen sich zur verlässlichen Dokumentation von Behandlungsergebnissen. Ihr universeller Einsatz in der multiparametrischen phoniatrischen Stimmfunktionsdiagnostik erscheint sinnvoll und erstrebenswert, da sie in der klinischen Anwendung die von der ELS empfohlenen Kriterien zur Stimmdiagnostik erweitern und verbessern.

5. Literaturverzeichnis

1. Dejonckere PH, Bradley P, Clemente P, Cornut G, Crevier-Buchman L, Friedrich G, Van De Heyning P, Remacle M, Woisard V, Committee on Phoniatrics of the European Laryngological S. *A basic protocol for functional assessment of voice pathology, especially for investigating the efficacy of (phonosurgical) treatments and evaluating new assessment techniques. Guideline elaborated by the Committee on Phoniatrics of the European Laryngological Society (ELS)*. Eur Arch Otorhinolaryngol. 2001;258(2):77-82.
2. Jacobson BH, Johnson A, Grywalski C, Silbergleit A, Jacobson G, Benninger MS, Newman CW. *The Voice Handicap Index (VHI): Development and Validation*. Am J Speech Lang Pathol 1997;6(3):66-70.
3. Nawka T, Leeuw IMVD, De Bodt M, Guimaraes I, Holmberg EB, Rosen CA, Schindler A, Woisard V, Whurr R, Konerding U. *Item Reduction of the Voice Handicap Index Based on the Original Version and on European Translations*. Folia Phoniatr Logo. 2009;61(1):37-48.
4. Nawka T, Rosanowski F, Gross M. *[How to render an expert opinion on dysphonia]*. Laryngorhinootologie. 2014;93(9):591-8.
5. Schonweiler R, Wubbelt P, Hess M, Ptok M. *[Psychoacoustic scaling of acoustic voice parameters by multicenter voice ratings]*. Laryngorhinootologie. 2001;80(3):117-22.
6. Ptok M, Schwemmle C, Iven C, Jessen M, Nawka T. *[On the auditory evaluation of voice quality]*. Hno. 2006;54(10):793-802.
7. Hanschmann H, Berger R. *[Perceptual and acoustic evaluation of hoarseness]*. Laryngorhinootologie. 2011;90(2):68-70.
8. Wendler J, Rauhut A, Krüger H. *Classification of voice qualities*. Journal of Phonetics. 1986;14 (3-4):483-8.
9. Anders LC, Hollien H, Hurme P, Sonninen A, Wendler J. *Perception of Hoarseness by Several Classes of Listeners*. Folia Phoniatr. 1988;40(2):91-100.
10. Seipelt M, Moller A, Nawka T, Gonnermann U, Caffier F, Caffier PP. *Monitoring the Outcome of Phonosurgery and Vocal Exercises with Established and New Diagnostic Tools*. Biomed Res Int. 2020;2020:4208189.
11. Schutte HK, Seidner W. *Recommendation by the Union-of-European-Phoniatricians (Uep) - Standardizing Voice Area Measurement Phonetography*. Folia Phoniatr. 1983;35(6):286-8.
12. Wuyts FL, De Bodt MS, Molenberghs G, Remacle M, Heylen L, Millet B, Van Lierde K, Raes J, Van de Heyning PH. *The dysphonia severity index: an objective measure of vocal quality based on a multiparameter approach*. J Speech Lang Hear Res. 2000;43(3):796-809.
13. Gonnermann U. *Quantifizierbare Verfahren zur Bewertung von Dysphonien [Quantifiable Techniques for Evaluation of Dysphonia]*. Frankfurt/Main, Germany: Peter Lang; 2007.
14. Hakkesteegt MM, Brocaar MP, Wieringa MH, Feenstra L. *Influence of age and gender on the dysphonia severity index. A study of normative values*. Folia Phoniatr Logop. 2006;58(4):264-73.
15. Aichinger P, Feichter F, Aichstill B, Bigenzahn W, Schneider-Stickler B. *Inter-device reliability of DSI measurement*. Logoped Phoniatr Vocol. 2012;37(4):167-73.
16. Barsties VLB, Ulozaite-Staniene N, Maryn Y, Petrauskas T, Uloza V. *The Influence of Gender and Age on the Acoustic Voice Quality Index and Dysphonia Severity Index: A Normative Study*. J Voice. 2019;33(3):340-5.
17. Caffier PP, Moller A, Forbes E, Muller C, Freymann ML, Nawka T. *The Vocal Extent Measure: Development of a Novel Parameter in Voice Diagnostics and Initial Clinical Experience*. Biomed Res Int. 2018;2018:3836714.
18. Caffier PP, Salmen T, Ermakova T, Forbes E, Ko SR, Song W, Gross M, Nawka T. *Phonomicrosurgery in Vocal Fold Nodules: Quantification of Outcomes in Professional and Non-Professional Voice Users*. Med Probl Perform Art. 2017;32(4):187-94.

19. Salmen T, Ermakova T, Möller A, Seipelt M, Weikert S, Rummich J, Gross M, Nawka T, Caffier PP. *The Value of Vocal Extent Measure (VEM) Assessing Phonomicrosurgical Outcomes in Vocal Fold Polyps*. J Voice. 2017;31(1):114.e7-.e15.
20. Salmen T, Ermakova T, Schindler A, Ko SR, Göktas Ö, Gross M, Nawka T, Caffier PP. *Efficacy of microsurgery in Reinke's oedema evaluated by traditional voice assessment integrated with the Vocal Extent Measure (VEM)*. Acta Otorhinolaryngol Ital. 2018;38(3):194-203.
21. Freymann ML, Mathmann P, Rummich J, Muller C, Neumann K, Nawka T, Caffier PP. *Gender-specific reference ranges of the vocal extent measure in young and healthy adults*. Logoped Phoniatr Vocol. 2020;45(2):73-81.
22. Müller C, Caffier F, Nawka T, Müller M, Caffier PP. *Pathology-Related Influences on the VEM: Three Years' Experience since Implementation of a New Parameter in Phoniatric Voice Diagnostics*. Biomed Res Int. 2020;2020:5309508.
23. Möller A. *Vocal extent measure as a new parameter in instrumental voice diagnostics. (Unpublished bachelor thesis)*. Stralsund, Germany: Fachhochschule Stralsund – University of Applied Sciences; 2010.
24. Caffier PP, Möller A. *Das Stimmumfangsmaß SUM als neuer Parameter in der objektiven Stimmdiagnostik*. Sprache · Stimme · Gehör. 2016;40(04):183-7.
25. Caffier PP, Schmidt B, Gross M, Karnetzky K, Nawka T, Rotter A, Seipelt M, Sedlmaier B. *A comparison of white light laryngostroboscopy versus autofluorescence endoscopy in the evaluation of vocal fold pathology*. Laryngoscope. 2013;123(7):1729-34.
26. Caffier PP, Nawka T, Ibrahim-Nasr A, Thomas B, Muller H, Ko SR, Song W, Gross M, Weikert S. *Development of three-dimensional laryngostroboscopy for office-based laryngeal diagnostics and phonosurgical therapy*. Laryngoscope. 2018;128(12):2823-31.
27. Strong MS, Jako GJ. *Laser surgery in the larynx. Early clinical experience with continuous CO 2 laser*. Ann Otol Rhinol Laryngol. 1972;81(6):791-8.
28. Steiner W. *Results of curative laser microsurgery of laryngeal carcinomas*. Am J Otolaryngol. 1993;14(2):116-21.
29. Ledda GP, Puxeddu R. *Carbon dioxide laser microsurgery for early glottic carcinoma*. Otolaryngol Head Neck Surg. 2006;134(6):911-5.
30. Harris AT, Tanyi A, Hart RD, Trites J, Rigby MH, Lancaster J, Nicolaides A, Taylor SM. *Transoral laser surgery for laryngeal carcinoma: has Steiner achieved a genuine paradigm shift in oncological surgery?* Ann R Coll Surg Engl. 2018;100(1):2-5.
31. Titze IR, Liang HX. *Comparison of F(O) Extraction Methods for High-Precision Voice Perturbation Measurements*. J Speech Hear Res. 1993;36(6):1120-33.
32. Koufman JA, Isaacson G. *The spectrum of vocal dysfunction*. Otolaryngol Clin North Am. 1991;24(5):985-8.
33. Youden WJ. *Index for rating diagnostic tests*. Cancer. 1950;3(1):32-5.
34. Rosen CA, Simpson CB. *Operative Techniques in Laryngology*. Berlin: Springer; 2008.
35. Nawka T, Wiesmann U, Gonnermann U. *Validation of the German version of the voice handicap index (VHI)*. Hno. 2003;51(11):921-9.
36. Remacle M, Van Haverbeke C, Eckel H, Bradley P, Chevalier D, Djukic V, de Vicentiis M, Friedrich G, Olofsson J, Peretti G, Quer M, Werner J. *Proposal for revision of the European Laryngological Society classification of endoscopic cordectomies*. Eur Arch Otorhinolaryngol. 2007;264(5):499-504.
37. Amin MB, Greene FL, Edge SB, Compton CC, Gershenwald JE, Brookland RK, Meyer L, Gress DM, Byrd DR, Winchester DP. *The Eighth Edition AJCC Cancer Staging Manual: Continuing to build a bridge from a population-based to a more "personalized" approach to cancer staging*. CA Cancer J Clin. 2017;67(2):93-9.
38. Ternstrom S, Pabon P, Sodersten M. *The Voice Range Profile: Its Function, Applications, Pitfalls and Potential*. Acta Acust United Ac. 2016;102(2):268-83.
39. Kaplan EL, Meier P. *Nonparametric Estimation from Incomplete Observations*. Journal of the American Statistical Association. 1958;53(282):457-81.

40. Caffier F, Nawka T, Neumann K, Seipelt M, Caffier PP. *Validation and Classification of the 9-Item Voice Handicap Index (VHI-9i)*. J Clin Med. 2021;10(15).
41. Woisard V, Bodin S, Yardeni E, Puech M. *The voice handicap index: correlation between subjective patient response and quantitative assessment of voice*. J Voice. 2007;21(5):623-31.
42. Hogikyan ND, Sethuraman G. *Validation of an instrument to measure voice-related quality of life (V-RQOL)*. J Voice. 1999;13(4):557-69.
43. Portone CR, Hapner ER, McGregor L, Otto K, Johns MM, 3rd. *Correlation of the Voice Handicap Index (VHI) and the Voice-Related Quality of Life Measure (V-RQOL)*. J Voice. 2007;21(6):723-7.
44. Tutya AS, Zambon F, Oliveira G, Behlau M. *Comparison of V-RQOL, VHI and VAPP scores in teachers*. Revista da Sociedade Brasileira de Fonoaudiologia. 2011;16(3):273-81.
45. Deary IJ, Wilson JA, Carding PN, MacKenzie K. *VoiSS: a patient-derived Voice Symptom Scale*. J Psychosom Res. 2003;54(5):483-9.
46. Deary IJ, Webb A, Mackenzie K, Wilson JA, Carding PN. *Short, self-report voice symptom scales: psychometric characteristics of the voice handicap index-10 and the vocal performance questionnaire*. Otolaryngol Head Neck Surg. 2004;131(3):232-5.
47. Ropero Rendon MDM, Ermakova T, Freymann ML, Ruschin A, Nawka T, Caffier PP. *Efficacy of Phonosurgery, Logopedic Voice Treatment and Vocal Pedagogy in Common Voice Problems of Singers*. Adv Ther. 2018;35(7):1069-86.
48. Reetz S, Bohlender JE, Brockmann-Bauser M. *Do Standard Instrumental Acoustic, Perceptual, and Subjective Voice Outcomes Indicate Therapy Success in Patients With Functional Dysphonia?* J Voice. 2019;33(3):317-24.
49. Ruotsalainen J, Sellman J, Lehto L, Verbeek J. *Systematic review of the treatment of functional dysphonia and prevention of voice disorders*. Otolaryngol Head Neck Surg. 2008;138(5):557-65.
50. Sielska-Badurek E, Osuch-Wojcikiewicz E, Sobol M, Kazanecka E, Rzepakowska A, Niemczyk K. *Combined Functional Voice Therapy in Singers With Muscle Tension Dysphonia in Singing*. J Voice. 2017;31(4):509 e23- e31.
51. Zeitels SM, Hillman RE, Desloge R, Mauri M, Doyle PB. *Phonomicrosurgery in singers and performing artists: treatment outcomes, management theories, and future directions*. Ann Otol Rhinol Laryngol Suppl. 2002;190:21-40.
52. LeBorgne WD, Donahue EN. *Voice Therapy as Primary Treatment of Vocal Fold Pathology*. Otolaryngol Clin North Am. 2019;52(4):649-56.
53. Sahin M, Gode S, Dogan M, Kirazli T, Ogut F. *Effect of voice therapy on vocal fold polyp treatment*. Eur Arch Otorhinolaryngol. 2018;275(6):1533-40.
54. Agarwal J, Wong A, Karle W, Naunheim M, Mori M, Courey M. *Comparing short-term outcomes of surgery and voice therapy for patients with vocal fold polyps*. Laryngoscope. 2019;129(5):1067-70.
55. Hakkesteegt MM, Brocaar MP, Wieringa MH. *The applicability of the dysphonia severity index and the voice handicap index in evaluating effects of voice therapy and phonosurgery*. J Voice. 2010;24(2): 199-205.
56. Nemr K, Simoes-Zenari M, de Souza GS, Hachiya A, Tsuji DH. *Correlation of the Dysphonia Severity Index (DSI), Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V), and Gender in Brazilians With and Without Voice Disorders*. J Voice. 2016;30(6):765 e7- e11.
57. Uloza V, Latoszek BBV, Ulozaite-Staniene N, Petrauskas T, Maryn Y. *A comparison of Dysphonia Severity Index and Acoustic Voice Quality Index measures in differentiating normal and dysphonic voices*. Eur Arch Otorhinolaryngol. 2018;275(4):949-58.
58. Brockmann-Bauser M, Balandat B, Bohlender JE. *Immediate Lip Trill Effects on the Standard Diagnostic Measures Voice Range Profile, Jitter, Maximum Phonation Time, and Dysphonia Severity Index*. J Voice. 2020;34(6):874-83.
59. Young VN, Smith LJ, Rosen C. *Voice outcome following acute unilateral vocal fold paralysis*. Ann Otol Rhinol Laryngol. 2013;122(3):197-204.
60. Yilmaz T. *Surgical treatment of glottic web using butterfly mucosal flap technique: Experience on 12 patients*. Laryngoscope. 2019;129(6):1423-7.

61. Rosen CA, Lee AS, Osborne J, Zullo T, Murry T. *Development and validation of the voice handicap index-10*. Laryngoscope. 2004;114(9):1549-56.
62. Dicks P, Nawka T. *Reliabilität der auditiv-perzeptiven Beurteilung der Heiserkeit organischer Stimmstörungen mittels visueller Analogskala und Ordinalskala unter Einsatz natürlicher Ankerstimmen*. Sprache Stimme Gehör. 2014;38(03):137-42.
63. Song W, Caffier F, Nawka T, Ermakova T, Martin A, Mürbe D, Caffier PP. *T1a Glottic Cancer: Advances in Vocal Outcome Assessment after Transoral CO(2)-Laser Microsurgery Using the VEM*. J Clin Med. 2021;10(6).
64. Canis M, Ihler F, Martin A, Matthias C, Steiner W. *Transoral laser microsurgery for T1a glottic cancer: review of 404 cases*. Head Neck. 2015;37(6):889-95.
65. Wiegand S. *Evidence and evidence gaps of laryngeal cancer surgery*. GMS Curr Top Otorhinolaryngol Head Neck Surg. 2016;15:Doc03.
66. Batra A, Goyal A, Goyal M, Goel S. *Oncological Outcomes Following Transoral CO2 Laser Microsurgery for T1 Glottic Cancer*. Indian J Otolaryngol Head Neck Surg. 2019;71(Suppl 1):542-7.
67. Vilaseca I, Huerta P, Blanch JL, Fernandez-Planas AM, Jimenez C, Bernal-Sprekelsen M. *Voice quality after CO2 laser cordectomy--what can we really expect?* Head Neck. 2008;30(1):43-9.
68. Lester SE, Rigby MH, MacLean M, Taylor SM. *'How does that sound?': objective and subjective voice outcomes following CO(2) laser resection for early glottic cancer*. J Laryngol Otol. 2011;125(12): 1251-5.
69. Friedman AD, Hillman RE, Landau-Zemer T, Burns JA, Zeitels SM. *Voice outcomes for photoangiolytic KTP laser treatment of early glottic cancer*. Ann Otol Rhinol Laryngol. 2013;122(3):151-8.
70. Chu PY, Hsu YB, Lee TL, Fu S, Wang LM, Kao YC. *Longitudinal analysis of voice quality in patients with early glottic cancer after transoral laser microsurgery*. Head Neck. 2012;34(9):1294-8.
71. Hamzany Y, Crevier-Buchman L, Lechien JR, Bachar G, Brasnu D, Hans S. *Multidimensional Voice Quality Evaluation After Transoral CO2 Laser Cordectomy: A Prospective Study*. Ear Nose Throat J. 2021;100(1_suppl):27S-32S.
72. Burns JA, Har-El G, Shapshay S, Maune S, Zeitels SM. *Endoscopic laser resection of laryngeal cancer: is it oncologically safe? Position statement from the American Broncho-Esophagological Association*. Ann Otol Rhinol Laryngol. 2009;118(6):399-404.
73. Aluffi Valletti P, Taranto F, Chiesa A, Pia F, Valente G. *Impact of resection margin status on oncological outcomes after CO2 laser cordectomy*. Acta Otorhinolaryngol Ital. 2018;38(1):24-30.
74. Peretti G, Piazza C, Cocco D, De Benedetto L, Del Bon F, Redaelli De Zinis LO, Nicolai P. *Transoral CO(2) laser treatment for T(is)-T(3) glottic cancer: the University of Brescia experience on 595 patients*. Head Neck. 2010;32(8):977-83.
75. Bertino G, Degiorgi G, Tinelli C, Cacciola S, Occhini A, Benazzo M. *CO(2) laser cordectomy for T1-T2 glottic cancer: oncological and functional long-term results*. Eur Arch Otorhinolaryngol. 2015;272(9):2389-95.
76. Fink DS, Sibley H, Kunduk M, Schexnaildre M, Kakade A, Sutton C, McWhorter AJ. *Subjective and objective voice outcomes after transoral laser microsurgery for early glottic cancer*. Laryngoscope. 2016;126(2):405-7.
77. Lee HS, Kim JS, Kim SW, Noh WJ, Kim YJ, Oh D, Hong JC, Lee KD. *Voice outcome according to surgical extent of transoral laser microsurgery for T1 glottic carcinoma*. Laryngoscope. 2016;126(9):2051-6.
78. Peretti G, Piazza C, Balzanelli C, Mensi MC, Rossini M, Antonelli AR. *Preoperative and postoperative voice in Tis-T1 glottic cancer treated by endoscopic cordectomy: an additional issue for patient counseling*. Ann Otol Rhinol Laryngol. 2003;112(9 Pt 1):759-63.
79. Roh JL, Kim DH, Kim SY, Park CI. *Quality of life and voice in patients after laser cordectomy for Tis and T1 glottic carcinomas*. Head Neck. 2007;29(11):1010-6.

Eidesstattliche Versicherung

„Ich, Felix Max Ludwig Caffier, versichere an Eides statt durch meine eigenhändige Unterschrift, dass ich die vorgelegte Dissertation mit dem Thema: *Evaluation des langjährigen Einsatzes von VHI-9i und SUM in der multiparametrischen phoniatrischen Stimmfunktionsdiagnostik – Evaluation of the long-term use of VHI-9i and VEM in multidimensional phoniatric voice function diagnostics* selbstständig und ohne nicht offengelegte Hilfe Dritter verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel genutzt habe.
Alle Stellen, die wörtlich oder dem Sinne nach auf Publikationen oder Vorträgen anderer Autoren/innen beruhen, sind als solche in korrekter Zitierung kenntlich gemacht. Die Abschnitte zu Methodik (insbesondere praktische Arbeiten, Laborbestimmungen, statistische Aufarbeitung) und Resultaten (insbesondere Abbildungen, Graphiken und Tabellen) werden von mir verantwortet.

Ich versichere ferner, dass ich die in Zusammenarbeit mit anderen Personen generierten Daten, Datenauswertungen und Schlussfolgerungen korrekt gekennzeichnet und meinen eigenen Beitrag sowie die Beiträge anderer Personen korrekt kenntlich gemacht habe (siehe Anteilserklärung). Texte oder Textteile, die gemeinsam mit anderen erstellt oder verwendet wurden, habe ich korrekt kenntlich gemacht.

Meine Anteile an etwaigen Publikationen zu dieser Dissertation entsprechen denen, die in der untenstehenden gemeinsamen Erklärung mit dem Erstbetreuer, angegeben sind. Für sämtliche im Rahmen der Dissertation entstandenen Publikationen wurden die Richtlinien des ICMJE (International Committee of Medical Journal Editors; www.icmje.og) zur Autorenschaft eingehalten. Ich erkläre ferner, dass ich mich zur Einhaltung der Satzung der Charité – Universitätsmedizin Berlin zur Sicherung Guter Wissenschaftlicher Praxis verpflichte.

Weiterhin versichere ich, dass ich diese Dissertation weder in gleicher noch in ähnlicher Form bereits an einer anderen Fakultät eingereicht habe.

Die Bedeutung dieser eidesstattlichen Versicherung und die strafrechtlichen Folgen einer unwahren eidesstattlichen Versicherung (§§156, 161 des Strafgesetzbuches) sind mir bekannt und bewusst.“

_____

Datum Unterschrift

Anteilserklärung an den erfolgten Publikationen

Felix Max Ludwig Caffier hatte folgenden Anteil an den eingereichten Publikationen:

Publikation 1: **Caffier F**, Nawka T, Neumann K, Seipelt M, Caffier PP.
Validation and Classification of the 9-Item Voice Handicap Index (VHI-9i).
J Clin Med 2021; 10(15):3325.
https://doi.org/10.3390/jcm10153325

Beitrag im Einzelnen: inhaltliche Ausarbeitung des Forschungsprojektes, Durchführung von Literaturrecherchen, Hauptanteil bei der Datenverarbeitung, Mitarbeit bei der statistischen Analyse, Erstellung der Tabellen 1 bis 6, Erstellung der Abbildungen 1 bis 6, Hauptanteil bei der Verfassung des Manuskripts und im Review-Prozess.

Publikation 2: Seipelt M, Möller A, Nawka T, Gonnermann U, **Caffier F**, Caffier PP.
Monitoring the Outcome of Phonosurgery and Vocal Exercises with Established and New Diagnostic Tools.
Biomed Res Int 2020 Jan 23;2020:4208189.
https://doi.org/10.1155/2020/4208189

Beitrag im Einzelnen: Durchführung von Literaturrecherchen, Mitarbeit bei der Datenverarbeitung und statistischen Analyse, Erstellung der Abbildungen 1 bis 4, Mitarbeit bei der Verfassung des Manuskripts und im Review-Prozess.

Publikation 3: Song W, **Caffier F**, Nawka T, Ermakova T, Martin A, Mürbe D, Caffier PP.
T1a Glottic Cancer: Advances in Vocal Outcome Assessment after Transoral CO2-Laser Microsurgery Using the VEM.
J Clin Med 2021; 10(6):1250.
https://doi.org/10.3390/jcm10061250

Beitrag im Einzelnen: Durchführung von Literaturrecherchen, Mitarbeit bei der Datenauswertung, Erstellung der Abbildungen 1 bis 3, Mitarbeit bei der Erstellung der Abbildung 4, Mitarbeit bei der Verfassung des Manuskripts und im Review-Prozess.

_____

Unterschrift, Datum und Stempel des betreuenden Hochschullehrers

_____

Unterschrift des Doktoranden

Auszug aus der Journal Summary List: Publikation 1

Journal Data Filtered By: **Selected JCR Year: 2019** Selected Editions: SCIE,SSCI
Selected Categories: **"MEDICINE, GENERAL and INTERNAL"**
Selected Category Scheme: WoS
Gesamtanzahl: 165 Journale

Rank	Full Journal Title	Total Cites	Journal Impact Factor	Eigenfactor Score
1	NEW ENGLAND JOURNAL OF MEDICINE	347,451	74.699	0.660800
2	LANCET	256,199	60.392	0.437300
3	JAMA-JOURNAL OF THE AMERICAN MEDICAL ASSOCIATION	158,632	45.540	0.290050
4	Nature Reviews Disease Primers	7,567	40.689	0.032310
5	BMJ-British Medical Journal	118,586	30.223	0.145170
6	ANNALS OF INTERNAL MEDICINE	58,033	21.317	0.091210
7	JAMA Internal Medicine	17,260	18.652	0.086180
8	PLOS MEDICINE	32,312	10.500	0.065990
9	Journal of Cachexia Sarcopenia and Muscle	3,553	9.802	0.007860
10	Cochrane Database of Systematic Reviews	67,763	7.890	0.134360
11	CANADIAN MEDICAL ASSOCIATION JOURNAL	15,212	7.744	0.016160
12	JOURNAL OF TRAVEL MEDICINE	2,659	7.089	0.006360
13	MAYO CLINIC PROCEEDINGS	15,627	6.942	0.024990
14	JOURNAL OF INTERNAL MEDICINE	10,912	6.871	0.014180
15	BMC Medicine	15,204	6.782	0.042500
16	MEDICAL JOURNAL OF AUSTRALIA	11,075	6.112	0.011070
17	Translational Research	4,043	5.411	0.008350
18	JOURNAL OF THE ROYAL SOCIETY OF MEDICINE	4,214	5.238	0.002580
19	JAMA Network Open	2,239	5.032	0.007660

Rank	Full Journal Title	Total Cites	Journal Impact Factor	Eigenfactor Score
20	Deutsches Arzteblatt International	4,817	4.796	0.007380
21	ANNALS OF FAMILY MEDICINE	5,567	4.686	0.010880
22	JOURNAL OF GENERAL INTERNAL MEDICINE	20,229	4.597	0.026960
23	AMERICAN JOURNAL OF MEDICINE	24,975	4.529	0.024230
24	Journal of Personalized Medicine	617	4.433	0.001950
25	AMERICAN JOURNAL OF PREVENTIVE MEDICINE	23,547	4.420	0.040180
26	European Journal of Internal Medicine	4,933	4.329	0.010280
27	AMYLOID-JOURNAL OF PROTEIN FOLDING DISORDERS	1,486	4.323	0.002920
28	BRITISH JOURNAL OF GENERAL PRACTICE	6,669	4.190	0.008670
29	Frontiers in Medicine	3,034	3.900	0.009870
30	PREVENTIVE MEDICINE	17,316	3.788	0.030080
31	PALLIATIVE MEDICINE	5,413	3.739	0.008460
32	AMERICAN JOURNAL OF CHINESE MEDICINE	3,531	3.682	0.002970
33	MEDICAL CLINICS OF NORTH AMERICA	3,161	3.529	0.004080
34	EUROPEAN JOURNAL OF CLINICAL INVESTIGATION	6,344	3.481	0.006590
35	PANMINERVA MEDICA	806	3.467	0.000660
36	Journal of Clinical Medicine	5,214	3.303	0.010940
37	ANNALS OF MEDICINE	4,510	3.243	0.005190
38	CANADIAN FAMILY PHYSICIAN	3,833	3.112	0.005150

Publikation 1

Caffier F, Nawka T, Neumann K, Seipelt M, Caffier PP.
Validation and Classification of the 9-Item Voice Handicap Index (VHI-9i).
J Clin Med 2021; 10(15):3325. https://doi.org/10.3390/jcm10153325

JCR Impact Factor 2020: 4.242 (Q1; 39/168 Medicine, General & Internal)

Journal of
Clinical Medicine

MDPI

Article

Validation and Classification of the 9-Item Voice Handicap Index (VHI-9i)

Felix Caffier [1,†], Tadeus Nawka [1,†], Konrad Neumann [2], Matthias Seipelt [3] and Philipp P. Caffier [1,*,†]

[1] Department of Audiology and Phoniatrics, Charité-Universitätsmedizin Berlin, Charitéplatz 1, D-10117 Berlin, Germany; felix.caffier@charite.de (F.C.); tadeus.nawka@charite.de (T.N.)
[2] Institute of Biometry and Clinical Epidemiology, Charité-Universitätsmedizin Berlin, Campus Charité Mitte, Charitéplatz 1, D-10117 Berlin, Germany; konrad.neumann@charite.de
[3] Department of Otorhinolaryngology, Ernst von Bergmann Klinikum Potsdam, Charlottenstr. 72, D-14467 Potsdam, Germany; MatthiasSeipelt@web.de
* Correspondence: philipp.caffier@charite.de
† Corporate Member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Campus Charité Mitte.

Abstract: The international nine-item Voice Handicap Index (VHI-9i) is a clinically established short-scale version of the original VHI, quantifying the patients' self-assessed vocal handicap. However, the current vocal impairment classification is based on percentiles. The main goals of this study were to establish test–retest reliability and a sound statistical basis for VHI-9i severity levels. Between 2009 and 2021, 17,660 consecutive cases were documented. A total of 416 test–retest pairs and 3661 unique cases with complete multidimensional voice diagnostics were statistically analyzed. Classification candidates were the overall self-assessed vocal impairment (VHIs) on a four-point Likert scale, the dysphonia severity index (DSI), the vocal extent measure (VEM), and the auditory–perceptual evaluation (GRB scale). The test–retest correlation of VHI-9i total scores was very high (r = 0.919, $p < 0.01$). Reliability was excellent regardless of gender or professional voice use, with negligible dependency on age. The VHIs correlated best with the VHI-9i, whereas statistical calculations proved that DSI, VEM, and GRB are unsuitable classification criteria. Based on ROC analysis, we suggest modifying the former VHI-9i severity categories as follows: 0 (healthy): $0 \leq 7$; 1 (mild): $8 \leq 16$; 2 (moderate): $17 \leq 26$; and 3 (severe): $27 \leq 36$.

Keywords: Voice Handicap Index (VHI-9i); international short scale; VHI-9i severity levels; test–retest reliability; validation of classification ranges; self-assessed vocal impairment (VHIs); hoarseness; dysphonia severity categories; voice diagnostics

Citation: Caffier, F.; Nawka, T.; Neumann, K.; Seipelt, M.; Caffier, P.P. Validation and Classification of the 9-Item Voice Handicap Index (VHI-9i). *J. Clin. Med.* **2021**, *10*, 3325. https://doi.org/10.3390/jcm10153325

Academic Editor: Renee Speyer

Received: 31 May 2021
Accepted: 25 July 2021
Published: 28 July 2021

1. Introduction

A patient's self-assessment of his or her own voice is an important tool for diagnosing voice disorders and vocal treatment outcomes [1,2]. Only the patients themselves can quantify how much a voice disorder impacts their daily lives. For instance, mild hoarseness affects professional voice users such as opera singers in a different way than non-professional voice users such as office workers [3,4].

The Voice Handicap Index (VHI) was developed and validated as a statistically robust method to measure the subjective impact of voice disorders [5]. The original questionnaire consists of 30 items (VHI-30) addressing functional, physical and emotional impairments in the context of dysphonia according to the patient's own experience. Each question is answered on a scale from 0 (never) to 4 (always), resulting in an overall score ranging from 0 to 120. The VHI-30 was translated and validated cross-culturally to form international variants (e.g., [6–11]) which were proven to be equivalent with each other [12,13].

From our own clinical experience, many patients and medical staff perceive the original 30-item questionnaire as rather time-consuming. To increase overall acceptance and practicability, shortened versions with fewer items were developed. A 12-item questionnaire [14,15] was soon followed by another reduction to 10 items [16,17]. Since 2009, the

commonly used variant at the Charité-Universitätsmedizin Berlin is the VHI-9i international questionnaire [14]. It consists of only nine items, after item reduction based on the original VHI-30 and European translations. A detailed discussion of the item and scale development can be found in the original VHI-9i publication [14]. In everyday diagnostic practice, the German translation of the VHI-9i is widely used by laryngologists and phoniatricians in German-speaking countries (e.g., [18–22]). Despite its clinical adoption, the reliability and validity of this VHI short scale as well as its classification have not yet been statistically verified. Instead, the current classification scale is based on the 25th, 50th, and 75th percentiles, dividing the scores into four severity classes. Thus far, clinical experience seems to plausibly reflect the self-perceived voice impairment. However, to overcome this arbitrary percentile-based exploration, we looked for a sound statistical basis for VHI-9i severity levels by revising the current cut-off points. In the context of expert opinion, thorough classifications of vocal parameters are essential for the assessment of dysphonia. In addition, a reliable and valid VHI-9i severity classification is needed to improve clinician-rated evaluations of treatment outcomes (e.g., better characterization of the quantified extent of subjective vocal impairment, more comprehensible assessment of individual pre- vs. post-therapeutic comparisons).

This study aims to address these shortcomings. Initially, we investigated whether the VHI-9i produces reliable results independent of age, gender or professional voice use. Next, the questionnaire validity was examined. For this purpose, the relationship between VHI-9i total scores and other established vocal parameters was statistically analyzed to establish cut-off values for healthy voices and mild to severe dysphonia. For external criteria, we intended to use objective acoustic–aerodynamic voice function diagnostics including voice range profile (VRP) measurements, dysphonia severity index (DSI) and vocal extent measure (VEM) calculations, as well as the subjective auditory–perceptual evaluation of voices by experienced examiners (GRB scale). Furthermore, the overall self-assessed vocal impairment (VHIs) served as an internal criterion.

2. Materials and Methods

2.1. Study Design and Patients

This study was conducted in accordance with the Declaration of Helsinki and approved by the local ethical review board. Selection criteria involved informed consent and the completion of the standard phoniatric examination procedures. After taking the medical history, all patients presenting in the Department of Audiology and Phoniatrics, Charité-Universitätsmedizin Berlin, Germany, received a digital videolaryngostroboscopy to assess the laryngeal findings and to establish a medical diagnosis. Subsequently, multidimensional voice function diagnostics were carried out as recommended by the European Laryngological Society (ELS) [1], starting with subjective evaluations (GRB, VHI-9i) and followed by objective voice function diagnostics (VRP, DSI, VEM). For subjective vocal self-assessment, patients completed the VHI-9i questionnaire. To estimate the voice use of every study participant, we also asked about their occupation and categorized them according to Koufman and Isaacson [23]: elite vocal performers (Level 1; e.g., actors, singers, voice artists), professional voice users (Level 2; e.g., teachers, politicians, moderators), non-vocal professionals (Level 3; e.g., lawyers, medical personnel, civil service employees), and non-vocal non-professionals (Level 4; e.g., IT staff, office workers, mechanics).

Between May 2009 and March 2021, a total of 17,660 consecutive cases were documented in the clinical database. To analyze the reliability of the VHI-9i, 718 patients were asked to complete the same questionnaire for a second time, without therapeutical intervention. The retest form had to be returned within one week to study the differences between the original answers and the retest. The second VHI-9i questionnaire was returned by 517 patients, corresponding to a response rate of 72%. Some questionnaires containing unanswered items or ambiguous checkmarks (e.g., between items) had to be excluded, resulting in 416 test–retest pairs.

The remaining 16,942 consecutive cases were analyzed to establish the validity of the questionnaire and to calculate statistically valid classification ranges. Since the VHI-9i should be compared with other established vocal parameters, only 7766 cases with complete multi-dimensional diagnostic assessment were considered. Cases with unreliable perturbation measures (jitter > 5%) were excluded, as recommended in the literature [1,24], resulting in a sample size of 6882. After another exclusion of follow-up visits, 3661 complete and unique cases were left for statistical analysis.

2.2. Subjective Examination Instruments

The VHI-9i represents an item-reduced short scale of the established VHI-30 [14], available in several languages (i.e., Dutch, English, French, German, Italian, Portuguese and Swedish). In this study, the German translation of the questionnaire was used (see Appendix A). Study participants were asked to answer all 9 items on a scale from 0 to 4 (0: never, 1: almost never, 2: sometimes, 3: almost always, 4: always), resulting in a total score between 0 and 36. The total score was then assigned to one of four dysphonia severity categories, ranging from 0 (healthy; $0 \leq 5$), 1 (mild; $6 \leq 13$), 2 (moderate; $14 \leq 22$), to 3 (severe; $23 \leq 36$). However, these categories correspond to a classification proposed by Nawka et al., based on the percentiles of a representative investigation of 716 patients [25]. Since these classification ranges have not yet been validated, statistical calculation of potential cut-off values for the VHI-9i classification was a main goal of this study.

Additionally, participants were asked to rate their overall voice impairment at present on a scale from 0 to 3 (0: normal, 1: mild, 2: moderate, 3: severe), the VHI summary assessment (VHIs). This index allows patients to assess how they feel about their voice with only one number. The relationship between VHI-9i and VHIs scores was examined to determine whether patients would rate themselves differently when asked about specific situations in their lives (VHI-9i items) or directly about their overall impairment (VHIs).

Apart from self-assessment, voices were also evaluated by auditory–perceptual assessment using the GRB system [26–28]. Based on the GRBAS scale, our department developed the modified GRB classification [29,30]. Only the first three criteria are used, focusing on the overall grade of hoarseness (G) and both main pathophysiological hoarseness components: roughness (R) and breathiness (B). The assessment of voice quality can be carried out more quickly and easily. Therefore, this system has become established in German-speaking countries and is also recommended in the ELS protocol [1]. Patients were asked to read the standardized text "The north wind and the sun" (German version), while the perceived G, R and B were scored on a scale from 0 to 3. To increase objectivity, each voice recording was rated independently by one experienced phoniatric physician and one senior speech–language therapist. The means were used for further exploration. While the degree of G serves as the overall indicator of dysphonia in the original GRBAS scale, it is regarded as gold standard for hoarseness evaluation in the GRB system presented here [31].

2.3. Objective Acoustic Assessment

For objective external validation criteria, we applied acoustic–aerodynamic voice function diagnostics. Voice recordings of all participants were conducted at the voice lab of our outpatient department, which is a sound-treated room with a background noise <40 dB(A). Study participants were asked to wear a head-mounted microphone with a stable mouth–microphone distance of 30 cm [32]. The equipment used for this purpose was the XION microphone headset (model number 352,009,010; XION GmbH, Berlin, Germany), which enables the realization of speech and singing VRP measurements and voice analyses under reproducible conditions. Technical microphone specifications include a frequency response of 70 Hz–20 kHz and a dynamic range of 40–120 dB(A). The microphone headset incorporates a calibrated audio interface that transmits digitized data to the PC via USB. The built-in electronics ensure the automatic calibration of the microphone connection without additional adjustments. The audio was processed via the DiVAS 2.8 software using the Singing Voice Analysis module (product number 350,020,013) and the Speaking

Voice Analysis module (product number 350,020,024; XION GmbH, Berlin, Germany). VRP measurements were performed to show the functional interactions of different components of voice generation regarding vocal frequency and intensity [33,34]. The detailed procedure of VRP recordings is described in previous publications [35,36].

The established parameter DSI was automatically calculated as a weighted combination of the highest possible fundamental frequency, the lowest phonation intensity, maximum phonation time and jitter [37]. Regarding jitter, the waveform matching method was used for fundamental frequency extraction as it meets the high-precision criterion of being able to extract a 1% frequency change per cycle with a 1% accuracy, as long as the signal-to-noise ratio is greater than about 40 dB and concomitant amplitude modulations are below about 5% [24]. Measurements were conducted in a standing position. Subjects were asked to produce a sustained vowel (/na/ or /a/) for about 3 seconds at comfortable pitch and loudness. The most stable recording out of 3 trials was chosen for DSI calculation. Based on Gonnermann's investigation of 495 subjects [38], the DSI scores were sorted into 4 severity categories, discriminating healthy voices ($\geq$4.2) from mildly (<4.2 to $\geq$1.8), moderately (<1.8 to ≥ -1.2), or severely (<-1.2) dysphonic voices. Since the DSI quantifies dysphonia as a negative criterion and involves the risk of imprecise results due to its multidimensional data acquisition, the one-dimensional parameter VEM was recently developed [35].

VEM calculation was performed automatically after VRP recording via the proprietary AVA software [39,40]. The VEM quantifies a subject's dynamic performance and frequency range. It is calculated as a relation of the area and perimeter of the VRP and describes the vocal function by an interval-scaled value without unit, usually between 0 and 120. These limits may be exceeded at both ends by either severely impaired or exceptionally capable voices with a large ambitus and dynamic range. A small vocal capacity is described by a low VEM, a large VRP by a high VEM. The VEM emphasizes the vocal abilities and enables a classification of voice performance as a positive criterion [21,31,41]. Based on Müller's investigation of 994 subjects [36], the resulting VEM scores were divided into percentiles, distinguishing a normal vocal capacity ($\geq$108) from mildly reduced (<108 to $\geq$93), moderately (<93 to $\geq$69) and severely reduced (<69) vocal capacities.

Table 1 summarizes the severity classification of different objective and subjective vocal parameters by reference range. In contrast to the ordinally scaled GRB and VHIs, the classifications of metrically scaled parameters (VEM, VHI-30, VHI-9i) are based on the percentiles of the respective study cohorts (Level 0: 100th percentile/4th quartile; Level 1: 75th percentile/3rd quartile; Level 2: 50th percentile/2nd quartile; Level 3: 25th percentile/1st quartile).

Table 1. Severity classification of different vocal parameters, assessed by study participants (VHI-30, VHI-9i, VHIs), experienced clinicians (GRB), and acoustic–aerodynamic analysis (VEM, DSI). Although all parameters share the same classification scale (0–3), equal levels of severity among different parameters do not imply equivalence (* classification ranges based on percentiles).

Level of Severity	VHI-30 *[25]	VHI-9i *[25]	VHIs	Grade (G)	VEM *[36]	DSI [38]
0: healthy	$0 \leq 14$	$0 \leq 5$	0	0	$\geq$108	$\geq$4.2
1: mild	$15 \leq 28$	$6 \leq 13$	1	1	93 < 108	1.8 < 4.2
2: moderate	$29 \leq 50$	$14 \leq 22$	2	2	69 < 93	$-1.2 < 1.8$
3: severe	$51 \leq 120$	$23 \leq 36$	3	3	<69	<-1.2

3. Data Analysis

Statistical analysis was performed using IBM SPSS version 26.0.0.1. To establish the questionnaire as reliable, the absolute differences in total VHI-9i scores between test and retest were compared. An analysis of the differences of every single item in the questionnaire is individually important, but only the total scores are relevant in diagnostic practice. Paired-sample *t*-tests were used to check for biases, and correlations were established

through Pearson's r. To test the dependency of the VHI-9i total score on age, a regression analysis was performed. Gender differences were analyzed through independent sample t-tests. We checked for a dependency on voice use by means of the nonparametric Kruskal–Wallis H-test.

Before the cut-off points for the VHI-9i severity categories could be validated, the correlations between the VHI-9i and the severity classifications for VHIs, DSI, VEM, G, R and B had to be determined using Spearman's rho (ρ), in order to choose which of them was best suited for classification. These vocal parameters had to be balanced in terms of sensitivity (i.e., true positive rate, TPR) and specificity (i.e., true negative rate, TNR) when applied to the VHI-9i scores. Receiver operator characteristic (ROC) curves were used, which plot the TPR against the false positive rate (FPR = 1 − TNR). Since ROC is a binary classifier, the curves had to be plotted three times to establish possible cut-off points for every severity level (0 vs. 1–3, 0–1 vs. 2–3, 0–2 vs. 3). The area under the curve (AUC) was used to rank the performance of every curve to distinguish between two severity classes. Values between 0.8 and 0.9 are considered excellent, 0.7 to 0.8 acceptable, 0.5 to 0.7 poor.

Several methods exist to determine good class boundaries from ROC curves. As a starting point, we used Youden's index (J) [42]. The highest J (Max J) is achieved when sensitivity and specificity are at optimal balance (J = TPR − FPR = TPR + TNR − 1). As a second possible class boundary, we determined the point where the number of correctly classified cases (CCCs) was the highest. The CCC is calculated as follows:

CCC	= TPR	* (n cases of classifying index above class boundary)
	+ TNR	* (n cases of classifying index below class boundary)

To find plausible cut-off values or categories of reasonable size, we selected a value between the two suggested class boundaries based on the median between Max J and Max CCC, also taking into account well over a decade of clinical experience with the VHI-9i.

4. Results

4.1. Test–Retest Reliability

After eliminating all incomplete questionnaires, 416 test–retest pairs were left. The mean age (±SD) was 50 (±17), with males skewing generally older at 56 (±16) compared to female patients at 46 (±17) years of age. A total of 26 participants (6.3%) were classified as elite vocal performers, 59 as professional voice users (14.2%), 78 as non-vocal professionals (18.7%) and 253 as non-vocal non-professionals (60.8%). An overview of the test–retest population is given in Figure 1 and Table 2.

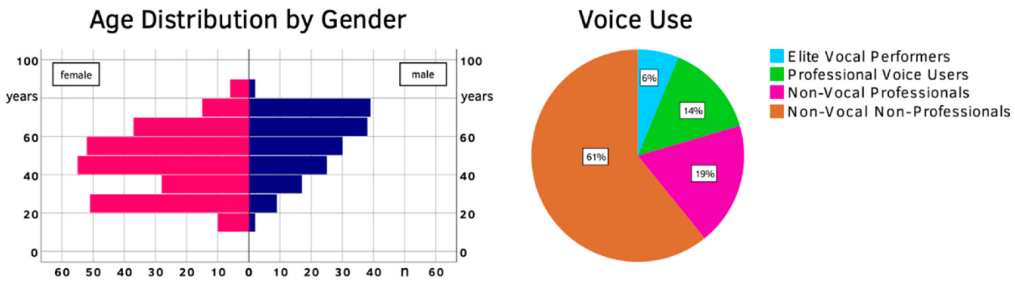

Figure 1. Overview of the test–retest population (age, gender, voice use classification).

Table 2. Study participant distribution and VHI-9i score differences between test and retest.

	Number *n* (%)	Mean Total Score Difference (±SD)
Male	162 (38.9%)	0.38 (±3.68)
Female	254 (61.1%)	0.17 (±3.42)
Voice Use Level 1	26 (6.3%)	0.75 (±3.45)
Voice Use Level 2	59 (14.2%)	0.82 (±3.48)
Voice Use Level 3	78 (18.7%)	0.40 (±2.91)
Voice Use Level 4	253 (60.8%)	0.02 (±3.70)
Age Group 0–24 years	46 (11.1%)	0.41 (±2.17)
Age Group 25–64 years	267 (64.2%)	0.45 (±3.47)
Age Group 65–99 years	103 (24.7%)	−0.33 (±4.06)

The median gap between test and retest was 2 days, with a mean of 3.3 days. The overall mean difference between VHI-9i scores (± SD) was very small at 0.25 (±3.52). Gender, voice use or age showed similarly minor differences (see Figure 2 and Table 2).

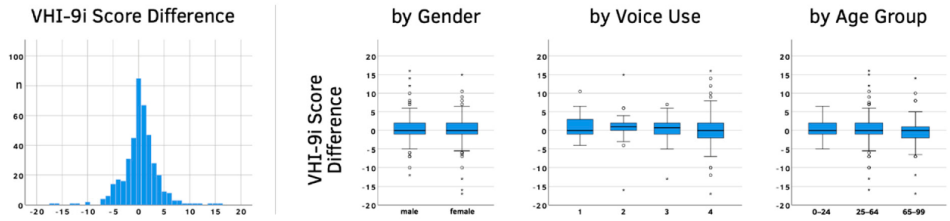

Figure 2. VHI-9i score difference between test and retest (total differences, by gender, by voice use, by age group). Age dependency was analyzed using discrete age values; age groups were only used in the diagram to improve the graphical representation. Circles (○) mark outliers (3rd quartile + 1.5*interquartile range; 1st quartile − 1.5*interquartile range) and asterisks (*) mark far outliers (3rd quartile + 3*interquartile range; 1st quartile − 3*interquartile range).

A paired-sample *t*-test between the VHI-9i total scores showed no significant differences ($p = 0.146$). Test and retest scores also correlated very well ($r = 0.919$, $p < 0.01$), indicating a highly reliable questionnaire. Only 5% of the population had a difference larger than 7 points. Gender had no impact on the reliability of the questionnaire. The independent sample *t*-test for the absolute VHI-9i score difference between males and females was not significant ($p = 0.589$). The level of voice use did also not affect reliability. The Kruskal–Wallis H-test showed no significance between the four voice use classifications ($p = 0.701$). The absolute score differences lightly depended on age. For every year of life, the difference rose by 0.016 points ($p = 0.028$).

4.2. Validation

Of the 3661 participants remaining for VHI-9i validation, 1456 were male (39.8%) and 2205 were female (60.2%). The mean age (±SD) was 48 (±17), with males being on average slightly older at 50 (±18) years compared to females at 47 (±17) years of age. Vocal impairment was caused by functional dysphonia in 40.8% of the study population. Patients with organic dysphonia (50.8%) showed various pathologies: mostly lesions of the lamina propria (e.g., vocal fold nodules, polyps, cysts, Reinke's edema), followed by benign and malignant changes of the epithelium (e.g., leukoplakia, papillomatosis, carcinoma), as well as neurogenic voice disorders (e.g., unilateral paralyses of the recurrent laryngeal nerve, spasmodic dysphonia). The remaining 8.4% were healthy subjects without dysphonia, mainly college applicants who presented to receive a vocal fitness examination, or prior to starting a profession associated with high vocal demands (e.g., teachers, singers, lecturers). The population pyramid and pathology classification are shown in Figure 3.

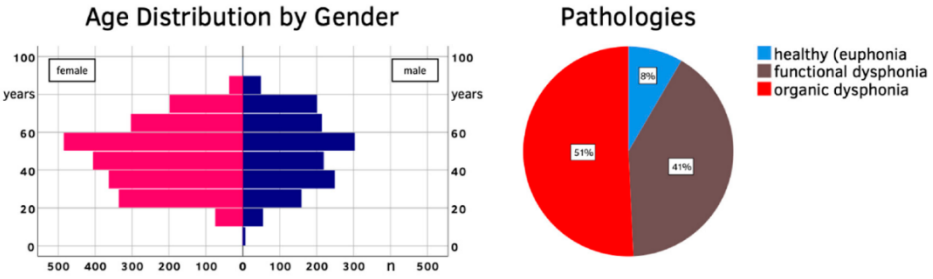

Figure 3. Overview of the validation population (age, gender, pathology classification).

As the test–retest examinations demonstrated, the reliability of VHI-9i scores is not affected by gender or voice use. Although statistically significant, the age dependency is so small that it can be neglected in clinical practice. Therefore, all further observations and calculations were conducted for the entire population of 3661 participants. Using the old VHI-9i classification scale based on percentiles [25], 15.5% of our participants had healthy voices (total score $0 \leq 5$), 25.7% mild dysphonia ($6 \leq 13$), 32.3% moderate ($14 \leq 22$) and 26.5% severe dysphonia ($23 \leq 36$). Applying the same method to the current database, 25% of patients had a score between 0 and 9, 50% up to 16, and 75% up to 22 points. The severity distribution for the other vocal parameters can be found in Table 3. Regarding VHIs, 63 cases had to be excluded ($n = 3598$ instead of 3661), because these test subjects had marked this question outside or in-between the provided options for the severity levels, rendering them invalid.

Table 3. Collected voice data by vocal parameter, classified according to the associated level of severity as shown in Table 1.

Vocal Parameter		Level of Severity			
		0: Healthy	1: Mild	2: Moderate	3: Severe
VHIs	number (%)	559 (15.5%)	1170 (32.5%)	1425 (39.6%)	444 (12.4%)
	mean VHI-9i score (±SD)	6.6 (±6.8)	12.8 (±7.2)	19.5 (±7.4)	23.9 (±7.8)
DSI	number (%)	879 (24.0%)	1210 (33.0%)	1244 (34.0%)	328 (9.0%)
	mean VHI-9i score (±SD)	11.9 (±8.1)	15.5 (±8.8)	17.7 (±8.9)	21.0 (±8.5)
VEM	number (%)	732 (20.0%)	673 (18.4%)	945 (25.8%)	1311 (35.8%)
	mean VHI-9i score (±SD)	11.1 (±8.0)	13.5 (±8.2)	15.7 (±8.3)	19.9 (±8.8)
G	number (%)	537 (14.7%)	1693 (46.2%)	1169 (31.9%)	262 (7.2%)
	mean VHI-9i score (±SD)	10.4 (±8.3)	14.2 (±8.4)	19.1 (±8.4)	23.3 (±7.8)
R	number (%)	602 (16.4%)	1864 (50.9%)	1031 (28.2%)	164 (4.5%)
	mean VHI-9i score (±SD)	11.7 (±8.9)	15.0 (±8.7)	18.9 (±8.4)	21.8 (±8.2)
B	number (%)	1865 (50.9%)	1205 (32.9%)	446 (12.2%)	145 (4.0%)
	mean VHI-9i score (±SD)	12.8 (±8.4)	17.3 (±8.4)	21.8 (±8.1)	25.6 (±6.7)

The size and mean of each severity category as well as the distribution of scores were notably different between parameters. The VHI-9i histogram shows a centered flat curve (skewness 0.063, kurtosis −0.90), the DSI is still centered but steeper (skewness −0.04, kurtosis 0.48) and the VEM is even steeper and skewed towards lower VEM values (skewness −1.08, kurtosis 1.94), with most patients falling into severity category 3 (Figure 4).

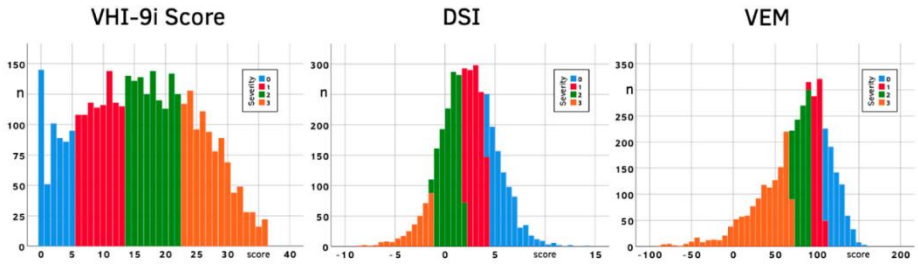

Figure 4. Observed VHI-9i, DSI and VEM scores with their associated severities.

The VHI-9i total scores correlated the most with the VHIs, even though ρ was only moderate (ρ = 0.592; see Table 4). All other parameters correlated notably weaker with the VHI-9i. The objective DSI and VEM were also moderately correlated to each other at ρ = 0.663. The distribution of subjects into G and R severity levels was rather similar, while B showed a different result with over 50% of all cases falling into the "healthy" category. G and R also had the strongest correlation among each other (ρ = 0.871), reinforcing clinical experience that G serves as the gold standard for hoarseness evaluations via the GRB scale.

Table 4. Results of correlation analysis between vocal parameters (Spearman's rho). All correlation coefficients were significant ($p < 0.001$).

	VHIs (0–3)	DSI (0–3)	VEM (0–3)	G (0–3)	R (0–3)	B (0–3)
VHI-9i	0.592	0.292	0.373	0.393	0.299	0.386
VHIs (0–3)		0.229	0.261	0.328	0.263	0.287
DSI (0–3)			0.663	0.525	0.454	0.494
VEM (0–3)				0.494	0.390	0.501
G (0–3)					0.871	0.665
R (0–3)						0.449

Figure 5 shows the distribution of VHI-9i total scores using the classifications for VHIs, DSI, VEM and G. The boxplots reveal a clear tendency: the higher the severity level, the higher the associated median. However, there is also a lot of overlap between the quartiles of different severity levels. This especially applies to DSI and VEM, which makes these parameters less suitable for VHI-9i classification.

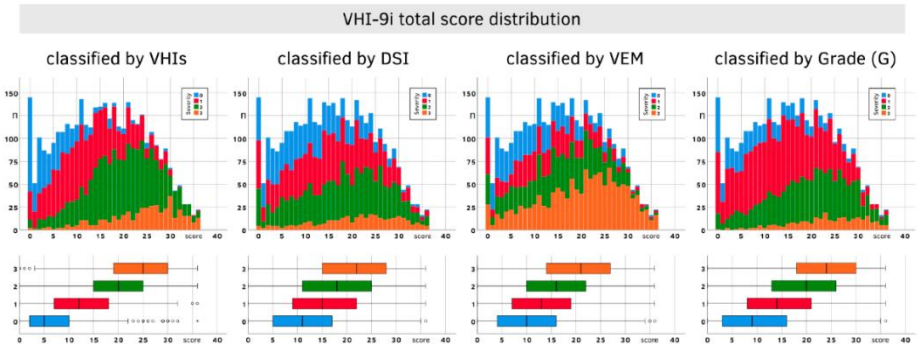

Figure 5. Distribution of VHI-9i total scores classified by VHIs, DSI, VEM and G severity levels. Upper row: stacked bar chart showing the number of subjects with their VHI-9i scores. Lower row: boxplots showing the percentiles of patients' VHI-9i scores by severity level. Circles (○) and asterisks (*) mark outliers and far outliers.

The ROC plots (Figure 6) also favor the VHIs as the best classifying index. DSI, VEM and G are visibly less suitable classifiers, because their curves are closer to the hypothetical diagonal through the ROC plot, signifying weaker discriminating performance.

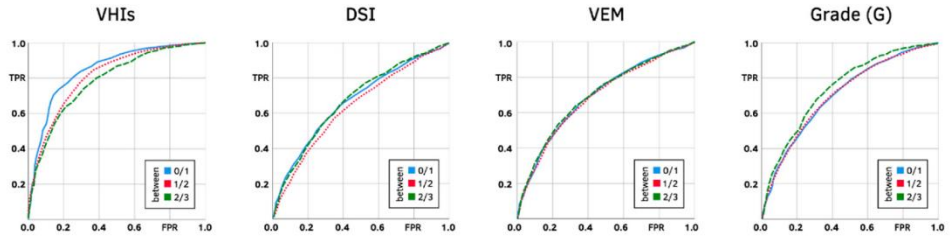

Figure 6. Combined ROC plots to determine cut-off points between severity categories 0 and 1 (blue), 1 and 2 (red), 2 and 3 (green).

The AUC results (Table 5) mirror the correlations of vocal parameters (compare Table 4). The best performance was achieved by the VHIs with excellent AUCs, followed by acceptable values for G. The parameters DSI and VEM turned out to be poor discriminators, with AUCs below 0.7.

As shown by our reliability analysis, severity categories must be at least 7 points in size to account for significant changes and minimize the possibility of retest artifacts. Neither optimizing for sensitivity and specificity (Max J) nor correctly classified cases (Max CCC) alone produced classes that were all wide enough (>7 points). Apart from the VHIs, Max CCC even produced cut-off recommendations that would eliminate the lowest (VEM) or lowest and highest (DSI, G) severity categories (highlighted in Table 5). Since both methods did not produce plausible cut-off values or categories of reasonable size, medians between the Max J and Max CCC measurements had to be calculated.

Table 5. ROC results for potential cut-offs between severity categories (0–1, 1–2, 2–3) using Max J, Max CCC and Median calculations. Yellow cells mark impossible cut-offs. Median calculations for every ROC parameter (TPR, FPR, J, CCC) resulted in slightly different class boundaries, which were specified by the ranges of cut-off values.

		VHIs			G		
		Cut 0–1	**Cut 1–2**	**Cut 2–3**	**Cut 0–1**	**Cut 1–2**	**Cut 2–3**
AUC		0.846	0.811	0.783	0.704	0.709	0.748
Max J	TPR	0.737	0.781	0.743	0.633	0.664	0.683
	FPR	0.174	0.298	0.316	0.33	0.352	0.311
	J	0.564	0.483	0.427	0.303	0.311	0.372
	CCC	2702	2674	2486	2336	2394	2521
	cut-off	11.5	14.75	19.5	13.5	16.75	20.5
Max CCC	TPR	0.966	0.818	0.115	1	0.464	0
	FPR	0.651	0.337	0.014	1	0.193	0
	J	0.315	0.481	0.101	0	0.271	0
	CCC	3132	2675	3162	3124	2464	3399
	cut-off	2.5	13.5	32.5	0	21.25	36
J–CCC–Median	TPR	0.86	0.78	0.43	0.83	0.59	0.32
	FPR	0.35	0.3	0.1	0.56	0.28	0.09
	J	0.51	0.48	0.33	0.27	0.31	0.23
	CCC	2988	2674	3026	2813	2443	3182
	cut-off	7–8	14–15	26–27	7–8	19	28

Table 5. *Cont.*

		DSI			VEM		
		Cut 0–1	**Cut 1–2**	**Cut 2–3**	**Cut 0–1**	**Cut 1–2**	**Cut 2–3**
AUC		0.667	0.64	0.674	0.692	0.689	0.699
Max J	TPR	0.651	0.569	0.683	0.648	0.585	0.639
	FPR	0.39	0.344	0.416	0.35	0.296	0.329
	J	0.26	0.226	0.267	0.298	0.289	0.31
	CCC	2346	2266	2170	2373	2309	2415
	cut-off	13.5	17.25	17.75	13.5	16.75	17.75
Max CCC	TPR	1	0.408	0	1	0.786	0.44
	FPR	1	0.216	0	1	0.537	0.167
	J	0	0.193	0	0	0.25	0.273
	CCC	2782	2280	3333	2929	2425	2535
	cut-off	0	21.75	36	0	10.75	22.5
J–CCC–Median	TPR	0.83	0.48	0.3	0.83	0.66	0.53
	FPR	0.66	0.28	0.13	0.61	0.38	0.23
	J	0.17	0.2	0.17	0.22	0.28	0.3
	CCC	2604	2266	3012	2718	2360	2512
	cut-off	7–8	18–20	26–27	7–8	14	20–21

However, both median calculations did not always return the exact same result, which is why the J–CCC–Median cut-off values are expressed as ranges in Table 5. In general, the difference between both medians was below 0.25 points most of the time and very rarely exceeded 0.5 points. The medians for all vocal parameters agreed on the first boundary (i.e., between severity levels 0 and 1) at 7 or 8. Between "mild" and "moderate" (severity levels 1 and 2), the median recommendations ranged from 14 to 20. Except for the VEM, the medians led to a cut-off point between 26 and 28 for the boundary distinguishing "moderate" from "severe" impairment (i.e., severity levels 2 and 3).

5. Discussion

The VHI-9i short scale has proven to be a valuable diagnostic tool in our clinical practice for well over a decade. The total number of 17,660 consecutively completed questionnaires documented in our database confirms its high acceptance among patients and medical staff. In our test–retest analysis, the VHI-9i questionnaire demonstrated very high reliability independent of gender or voice use. Age had a minor influence, which we do not consider clinically relevant: For every year of life, the absolute score difference between test and retest increased by 0.016. If we applied that difference to the entire age range of our study population, the VHI-9i total score of an adolescent compared to a senior person would differ by about 1. The reliability analysis also showed that the severity classes for the VHI-9i need to be at least 7 points in size (2*SD of paired sample *t*-test), since only differences of 7 points and above account for significant changes and minimize the possibility of retest artifacts. Our interpretation of the ROC analysis had to consider this requirement. Unfortunately, neither optimizing for Max J nor Max CCC resulted in categories that were all large enough. Calculating the median between them for each cut-off point, however, yielded satisfactory results for clinical use.

All classification ranges are listed in Table 6. The Median J method strikes a good balance between sensitivity, specificity and the minimum class width of 7 points. The new boundary of a score of 7 corresponds directly with the VHIs Median J result for healthy voices (class 0). Finding a reasonable upper boundary for severity level 1 is more difficult: using VHIs Median J (a score of 14) would result in a category that is too small. The median for the expert auditory–perceptual assessment (G) points towards an even higher boundary (a score of 19). Since we were trying to find a mid-point for our severity classes, we decided to use the upper boundary of the 50% quartile (a score of 16). The upper boundary for

severity level 2 (moderate impairment) can be taken once again from the VHIs Median J row, placing class 2 between 17 ≤ 26 and class 3 between 27 ≤ 36.

Table 6. Sizes of severity classes based on Max J, Max CCC and Median calculations. Green cells serve as the basis for our proposed new VHI-9i severity classification.

Classifying Method	Level of Severity			
	0: Healthy	1: Mild	2: Moderate	3: Severe
VHIs (Max J)	0 ≤ 12	13 ≤ 15	16 ≤ 20	21 ≤ 36
VHIs (Max CCC)	0 ≤ 3	4 ≤ 14	15 ≤ 33	34 ≤ 36
VHIs (Median J)	0 ≤ 7	8 ≤ 14	15 ≤ 26	27 ≤ 36
VHIs (Median CCC)	0 ≤ 8	9 ≤ 15	16 ≤ 27	28 ≤ 36
G (Max J)	0 ≤ 14	15 ≤ 17	18 ≤ 21	22 ≤ 36
G (Max CCC)	-	0 ≤ 21	22 ≤ 36	-
G (Median J)	0 ≤ 7	8 ≤ 19	20 ≤ 28	29 ≤ 36
G (Median CCC)	0 ≤ 8	9 ≤ 19	20 ≤ 28	29 ≤ 36
DSI (Max J)	0 ≤ 14	15 ≤ 17	18	19 ≤ 36
DSI (Max CCC)	-	0 ≤ 22	23 ≤ 36	-
DSI (Median J)	0 ≤ 8	9 ≤ 20	21 ≤ 27	28 ≤ 36
DSI (Median CCC)	0 ≤ 8	9 ≤ 18	19 ≤ 26	27 ≤ 36
VEM (Max J)	0 ≤ 14	15 ≤ 17	18	19 ≤ 36
VEM (Max CCC)	-	0 ≤ 11	12 ≤ 23	24 ≤ 36
VEM (Median J)	0 ≤ 8	9 ≤ 14	15 ≤ 20	21 ≤ 36
VEM (Median CCC)	0 ≤ 7	8 ≤ 14	15 ≤ 21	22 ≤ 36
VHI-9i quartiles	0 ≤ 9	10 ≤ 16	17 ≤ 22	23 ≤ 36
Proposed new classification	**0 ≤ 7**	**8 ≤ 16**	**17 ≤ 26**	**27 ≤ 36**

Compared to the old VHI-9i classification scale based on percentiles [25], the revised severity ranges classify more patients towards the lower categories. Severity level 3 is reduced by 4 points and is no longer the largest category. Level 1 and 2 start at higher class boundaries due to the size increase in level 0.

The best correlation was observed between VHI-9i and VHIs, making the overall self-assessed vocal impairment the best candidate for the validation process. However, the VHI-9i did not correlate well with the two objective parameters DSI and VEM, and had only slightly higher correlations with GRB. This supports recent studies that all these vocal parameters measure different aspects of a patient's voice and are neither mutually interchangeable nor redundant [31,36,41,43]. Due to the weak correlations, poor discriminating performance and sometimes impossible cut-off points, DSI, VEM and G ultimately had no part in our recommendation for the revised VHI-9i cut-off points. It is important to remember that the VHI-9i does not measure objective voice impairment (DSI) or vocal capacity (VEM), but personal suffering due to a subjectively perceived vocal handicap. None of the parameters allow conclusions to be drawn about the diagnoses or underlying causes of the voice disorder.

Study Limitations

Over 60% of our test–retest population were categorized as non-vocal non-professionals. Ideally, the study would have included more subjects with professional backgrounds in singing, acting or teaching, especially since establishing independence from voice use was one of our goals during the rest-retest analysis. A bigger population of elite vocal performers and professional voice users would have been preferrable, but does not represent the actual proportions of our clinic clientele.

Furthermore, males are underrepresented in our study, so there may be participation bias. Despite the limited number of male subjects, we concluded that the VHI-9i was independent of gender, but a more balanced gender involvement would have been more representative. However, our clinical experience shows that women are generally more likely to see a doctor for voice problems.

In addition, signal-to-noise ratio (SNR) analysis and signal typing are considered to be important for valid and reliable perturbation measurements [44–46]. Unfortunately, this functionality is not included in the DiVAS software, which was specified in our study design as the main tool for objective voice analysis. One of the fundamental limitations of the DSI is the inclusion of jitter without sufficient evaluation of the signal type. In general, only type 1 and 2 are considered viable for perturbation analysis. The 5% jitter cut-off applied in our study was established to exclude type 4 signals only [46]. However, the categorization of a small test sample (n = 40) revealed signal type 1 and 2 exclusively, even for patients with low DSI and high jitter values. Furthermore, the majority of SNR results were between 42 and 50 dB ("recommended"), with a smaller number between 30 and 42 dB ("acceptable") [45]. Therefore, we believe that our exclusion criteria were sufficient to eliminate voices which are not suitable for perturbation analysis. We recognize that this estimate cannot be taken as proof for the entire dataset and plan to include SNR and signal typing analyses in our future studies from the outset. It should also be noted that jitter was only used for DSI calculation, which proved to be irrelevant for the main goal of our study, i.e., a revised VHI-9i classification. Therefore, our recommendations regarding VHI-9i severity categories should not have been distorted.

Moreover, our initial ROC analysis produced boundary recommendations that were not feasible for diagnostic purposes. The resulting severity categories would have been either too small (<7 points) or would even not exist at all. Calculating the median between Max J and Max CCC is not a commonly used method for solving these problems. However, based on the frequent use of the VHI-9i in clinical investigations [18–22,31,36,41], it appears that the new classification will be a practical option for clinical settings.

In general, the auditory-perceptual assessment of voices via GRB was conducted only by two experienced examiners. Safer larger group judgments were not made. Due to the enormous number of cases (n = 17,660) and over a decade of diagnostic voice recordings, a retrospective blinded voice evaluation with 4-5 raters was not an option.

6. Conclusions

The VHI-9i is a reliable questionnaire which is independent of gender and professional voice use. Its dependency on age is negligible. Based on many years of clinical experience, it also has high acceptance among patients and medical staff, making it a valuable diagnostic tool.

The old cut-off values for the VHI-9i severity categories based on percentiles had to be adjusted. We recommend setting class 0 (healthy) between $0 \leq 7$, class 1 (mild impairment) between $8 \leq 16$, class 2 (moderate impairment) between $17 \leq 26$ and class 3 (severe impairment) between $27 \leq 36$.

The subjective VHI-9i does not correlate well with objective vocal parameters (DSI, VEM) or subjective auditory–perceptual assessment (GRB), reinforcing the notion that all these parameters measure different dimensions of a patient's voice and are neither mutually interchangeable nor redundant.

Author Contributions: Conceptualization, T.N., M.S. and P.P.C.; Methodology, F.C., T.N. and P.P.C.; Literature Review, F.C. and P.P.C.; Investigation, T.N., M.S. and P.P.C.; Data Analysis, F.C. and K.N.; Original Draft Writing, F.C. and P.P.C.; Draft Review and Editing, F.C., T.N., K.N. and P.P.C.; Visualization, F.C. and K.N.; Supervision, T.N. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Ethics Committee of Charité-Universitätsmedizin Berlin, Berlin, Germany (reference number: EA4/140/10).

Informed Consent Statement: Informed consent was obtained from all study participants.

Data Availability Statement: All data of the study are available in the Department of Audiology and Phoniatrics, Charité-Universitätsmedizin Berlin, Berlin, Germany.

Acknowledgments: The authors wish to thank Tatiana Ermakova for the statistical advice.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Table A1. VHI-9i questionnaire items (*German translation*) as used in the study.

Item Text	Score				
My voice makes it difficult for people to hear me. (*Man hört mich wegen meiner Stimme schlecht.*)	0	1	2	3	4
People have difficulty understanding me in a noisy room. (*Anderen fällt es schwer, mich in einem lauten Raum zu verstehen.*)	0	1	2	3	4
The sound of my voice varies throughout the day. (*Der Klang meiner Stimme ändert sich im Laufe des Tages.*)	0	1	2	3	4
My family has difficulty hearing me when I call them throughout the house. (*Meine Familie hört mich kaum, wenn ich zuhause nach ihnen rufe.*)	0	1	2	3	4
My voice difficulties restrict my personal and social life. (*Meine Stimmschwierigkeiten schränken mich in meinem Privatleben ein.*)	0	1	2	3	4
The clarity of my voice is unpredictable. (*Bevor ich spreche, weiß ich nicht, wie klar meine Stimme klingen wird.*)	0	1	2	3	4
My voice is worse in the evening. (*Abends ist meine Stimme schlechter.*)	0	1	2	3	4
I am less outgoing because of my voice problem. (*Ich bin weniger kontaktfreudig wegen meines Stimmproblems.*)	0	1	2	3	4
My voice makes me feel incompetent. (*Wegen meiner Stimme fühle ich mich unfähig.*)	0	1	2	3	4

Scoring: 0 = never (*nie*), 1 = almost never (*selten*), 2 = sometimes (*manchmal*), 3 = almost always (*oft*), 4 = always (immer).

Table A2. Global VHIs question added to the study questionnaire.

Question	Score			
How do you rate your voice today? (*Wie schätzen Sie Ihre Stimme heute ein?*)	0	1	2	3

Scoring: 0 = normal (*normal*), 1 = mildly (*leicht*), 2 = moderately (*mittelgradig*), 3 = severely disturbed (*hochgradig gestört*).

References

1. Dejonckere, P.H.; Bradley, P.; Clemente, P.; Cornut, G.; Crevier-Buchman, L.; Friedrich, G.; Van de Heyning, P.; Remacle, M.; Woisard, V. A basic protocol for functional assessment of voice pathology, especially for investigating the efficacy of (phonosurgical) treatments and evaluating new assessment techniques. Guideline elaborated by the Committee on Phoniatrics of the European Laryngological Society (ELS). *Eur. Arch. Oto-Rhino-Laryngol.* **2001**, *258*, 77–82. [CrossRef]
2. Carding, P.N.; Wilson, J.; MacKenzie, K.; Deary, I.J. Measuring voice outcomes: State of the science review. *J. Laryngol. Otol.* **2009**, *123*, 823–829. [CrossRef]
3. Sataloff, R.T. Professional voice users: The evaluation of voice disorders. *Occup. Med.* **2001**, *16*, 633–647.
4. Mori, M.C.; Francis, D.O.; Song, P.C. Identifying Occupations at Risk for Laryngeal Disorders Requiring Specialty Voice Care. *Otolaryngol. Neck Surg.* **2017**, *157*, 670–675. [CrossRef] [PubMed]
5. Jacobson, B.H.; Johnson, A.; Grywalski, C.; Silbergleit, A.; Jacobson, G.; Benninger, M.S.; Newman, C.W. The Voice Handicap Index (VHI): Development and Validation. *Am. J. Speech Lang. Pathol.* **1997**, *6*, 66–70. [CrossRef]
6. Frajkova, Z.; Krizekova, A.; Missikova, V.; Tedla, M. Translation, Cross-Cultural Validation of the Voice Handicap Index (VHI-30) in Slovak Language. *J. Voice* **2020**. [CrossRef]
7. Sakaguchi, Y.; Kanazawa, T.; Okui, A.; Hirosaki, M.; Konomi, U.; Sotome, T.; Tashiro, N.; Kurihara, M.; Omae, T.; Nakayama, Y.; et al. Assessment of Dysphonia Using the Japanese Version of the Voice Handicap Index and Determination of Cutoff Points for Screening. *J. Voice* **2020**. [CrossRef] [PubMed]

8. Sotirović, J.; Grgurević, A.; Mumović, G.; Grgurević, U.; Pavićević, L.; Perić, A.; Erdoglija, M.; Milojević, M. Adaptation and Validation of the Voice Handicap Index (VHI)-30 into Serbian. *J. Voice* **2016**, *30*, 758.e1–758.e6. [CrossRef] [PubMed]
9. Jaruchinda, P.; Suwanwarangkool, T. Cross-Cultural Adaptation and Validation of the Voice Handicap Index into Thai. *J. Med. Assoc. Thail.* **2015**, *98*, 1199–1208.
10. Trinite, B.; Sokolovs, J. Adaptation and Validation of the Voice Handicap Index in Latvian. *J. Voice* **2014**, *28*, 452–457. [CrossRef]
11. Nawka, T.; Wiesmann, U.; Gonnermann, U. Validation of the German version of the Voice Handicap Index (VHI). *Hno* **2003**, *51*, 921–929. [CrossRef] [PubMed]
12. Leeuw, I.V.-D.; Kuik, D.; De Bodt, M.; Guimarães, I.; Holmberg, E.; Nawka, T.; Rosen, C.; Schindler, A.; Whurr, R.; Woisard, V. Validation of the Voice Handicap Index by Assessing Equivalence of European Translations. *Folia Phoniatr. Logop.* **2008**, *60*, 173–178. [CrossRef] [PubMed]
13. Seifpanahi, S.; Jalaie, S.; Nikoo, M.R.; Sobhani-Rad, D. Translated Versions of Voice Handicap Index (VHI)-30 across Languages: A Systematic Review. *Iran. J. Public Health* **2015**, *44*, 458–469.
14. Nawka, T.; Leeuw, I.V.-D.; De Bodt, M.; Guimarães, I.; Holmberg, E.; Rosen, C.; Schindler, A.; Woisard, V.; Whurr, R.; Konerding, U. Item Reduction of the Voice Handicap Index Based on the Original Version and on European Translations. *Folia Phoniatr. Logop.* **2009**, *61*, 37–48. [CrossRef]
15. Hanschmann, H.; Lohmann, A.; Berger, R. Comparison of Subjective Assessment of Voice Disorders and Objective Voice Measurement. *Folia Phoniatr. Logop.* **2011**, *63*, 83–87. [CrossRef]
16. Gilbert, M.R.; Gartner-Schmidt, J.L.; Rosen, C.A. The VHI-10 and VHI Item Reduction Translations—Are we all Speaking the Same Language? *J. Voice* **2017**, *31*, 250.e1–250.e7. [CrossRef] [PubMed]
17. Rosen, C.A.; Lee, A.S.; Osborne, J.; Zullo, T.; Murry, T. Development and validation of the voice handicap index-10. *Laryngoscope* **2004**, *114*, 1549–1556. [CrossRef]
18. Song, W.; Caffier, F.; Nawka, T.; Ermakova, T.; Martin, A.; Mürbe, D.; Caffier, P. T1a Glottic Cancer: Advances in Vocal Outcome Assessment after Transoral CO_2-Laser Microsurgery Using the VEM. *J. Clin. Med.* **2021**, *10*, 1250. [CrossRef]
19. Langenfeld, A.; Bohlender, J.E.; Swanenburg, J.; Brockmann-Bauser, M. Cervical Spine Disability in Correlation with Subjective Voice Handicap in Patients with Voice Disorders: A Retrospective Analysis. *J. Voice* **2020**, *34*, 371–379. [CrossRef] [PubMed]
20. Reetz, S.; Bohlender, J.E.; Brockmann-Bauser, M. Do Standard Instrumental Acoustic, Perceptual, and Subjective Voice Outcomes Indicate Therapy Success in Patients With Functional Dysphonia? *J. Voice* **2019**, *33*, 317–324. [CrossRef]
21. Salmen, T.; Ermakova, T.; Schindler, A.; Ko, S.-R.; Göktas, O.; Gross, M.; Nawka, T.; Caffier, P. Efficacy of microsurgery in Reinke's oedema evaluated by traditional voice assessment integrated with the Vocal Extent Measure (VEM). *Acta Otorhinolaryngol. Ital.* **2018**, *38*, 194–203. [CrossRef]
22. Caffier, P.P.; Salmen, T.; Ermakova, T.; Forbes, E.; Ko, S.-R.; Song, W.; Gross, M.; Nawka, T. Phonomicrosurgery in Vocal Fold Nodules: Quantification of Outcomes in Professional and Non-Professional Voice Users. *Med Probl. Perform. Artist.* **2017**, *32*, 187–194. [CrossRef] [PubMed]
23. Koufman, J.A.; Isaacson, G. The spectrum of vocal dysfunction. *Otolaryngol. Clin. N. Am.* **1991**, *24*, 985–988. [CrossRef]
24. Titze, I.R.; Liang, H. Comparison of F o Extraction Methods for High-Precision Voice Perturbation Measurements. *J. Speech Lang. Heart Res.* **1993**, *36*, 1120–1133. [CrossRef] [PubMed]
25. Nawka, T.; Rosanowski, F.; Gross, M. How to render an expert opinion on dysphonia. *Laryngorhinootologie* **2014**, *93*, 591–598.
26. Hanschmann, H.; Berger, R. Perceptual and acoustic evaluation of hoarseness. *Laryngorhinootologie* **2011**, *90*, 68–70.
27. Ptok, M.; Schwemmle, C.; Iven, C.; Jessen, M.; Nawka, T. On the auditory evaluation of voice quality. *Hno* **2006**, *54*, 793–802. [CrossRef]
28. Schönweiler, R.; Wübbelt, P.; Hess, M.; Ptok, M. Psychoacoustic scaling of acoustic voice parameters by multicenter voice ratings. *Laryngorhinootologie* **2001**, *80*, 117–122. [CrossRef]
29. Wendler, J.; Rauhut, A.; Krüger, H. Classification of voice qualities. *J. Phon.* **1986**, *14*, 483–488. [CrossRef]
30. Anders, L.; Hollien, H.; Hurme, P.; Sonninen, A.; Wendler, J. Perception of Hoarseness by Several Classes of Listeners. *Folia Phoniatr. Logop.* **1988**, *40*, 91–100. [CrossRef] [PubMed]
31. Seipelt, M.; Möller, A.; Nawka, T.; Gonnermann, U.; Caffier, F.; Caffier, P.P. Monitoring the Outcome of Phonosurgery and Vocal Exercises with Established and New Diagnostic Tools. *BioMed Res. Int.* **2020**, *2020*, 4208189. [CrossRef]
32. Schutte, H.; Seidner, W. Recommendation by the Union of European Phoniatricians (UEP): Standardizing Voice Area Measurement/Phonetography. *Folia Phoniatr. Logop.* **1983**, *35*, 286–288. [CrossRef]
33. Ternström, S.; Pabon, P.; Södersten, M.; Peter, P.; Maria, S. The Voice Range Profile: Its Function, Applications, Pitfalls and Potential. *Acta Acust. United Acust.* **2016**, *102*, 268–283. [CrossRef]
34. Printz, T.; Godballe, C.; Grøntved, Å.M. The Dual-Microphone Voice Range Profile Assessment—Interrater Reliability. *J. Voice* **2020**. [CrossRef] [PubMed]
35. Caffier, P.P.; Möller, A.; Forbes, E.; Müller, C.; Freymann, M.-L.; Nawka, T. The Vocal Extent Measure: Development of a Novel Parameter in Voice Diagnostics and Initial Clinical Experience. *BioMed Res. Int.* **2018**, *2018*, 3836714. [CrossRef] [PubMed]
36. Müller, C.; Caffier, F.; Nawka, T.; Müller, M.; Caffier, P.P. Pathology-Related Influences on the VEM: Three Years' Experience since Implementation of a New Parameter in Phoniatric Voice Diagnostics. *BioMed Res. Int.* **2020**, *2020*, 5309508. [CrossRef] [PubMed]

37. Wuyts, F.L.; Bodt, M.S.D.; Molenberghs, G.; Remacle, M.; Heylen, L.; Millet, B.; Lierde, K.V.; Raes, J.; Heyning, P.H.V.D. The dysphonia severity index: An objective measure of vocal quality based on a multiparameter approach. *J. Speech Lang. Heart Res.* **2000**, *43*, 796–809. [CrossRef]
38. Gonnermann, U. *Quantifizierbare Verfahren zur Bewertung von Dysphonien [Quantifiable Techniques for Evaluation of Dysphonia]*; Peter Lang: Frankfurt/Main, Germany, 2007.
39. Möller, A. Vocal Extent Measure as a New Parameter in Instrumental Voice Diagnostics. Unpublished. Bachelor Thesis, Fachhochschule Stralsund—University of Applied Sciences, Stralsund, Germany, 2010.
40. Caffier, P.P.; Möller, A. Das Stimmumfangsmaß SUM als neuer Parameter in der objektiven Stimmdiagnostik. *Sprache · Stimme · Gehör* **2016**, *40*, 183–187. [CrossRef]
41. Salmen, T.; Ermakova, T.; Möller, A.; Seipelt, M.; Weikert, S.; Rummich, J.; Gross, M.; Nawka, T.; Caffier, P.P. The Value of Vocal Extent Measure (VEM) Assessing Phonomicrosurgical Outcomes in Vocal Fold Polyps. *J. Voice* **2017**, *31*, 114.e7–114.e15. [CrossRef]
42. Youden, W.J. Index for rating diagnostic tests. *Cancer* **1950**, *3*, 32–35. [CrossRef]
43. Woisard, V.; Bodin, S.; Yardeni, E.; Puech, M. The Voice Handicap Index: Correlation Between Subjective Patient Response and Quantitative Assessment of Voice. *J. Voice* **2007**, *21*, 623–631. [CrossRef] [PubMed]
44. Titze, I.R. *Workshop on Acoustic Voice Analysis: Summary Statement*; National Center for Voice and Speech: Iowa City, IA, USA, 1995.
45. Deliyski, D.D.; Shaw, H.S.; Evans, M.K. Adverse Effects of Environmental Noise on Acoustic Voice Quality Measurements. *J. Voice* **2005**, *19*, 15–28. [CrossRef] [PubMed]
46. Sprecher, A.; Olszewski, A.; Jiang, J.J.; Zhang, Y. Updating signal typing in voice: Addition of type 4 signals. *J. Acoust. Soc. Am.* **2010**, *127*, 3710–3716. [CrossRef] [PubMed]

Auszug aus der Journal Summary List: Publikation 2

Journal Data Filtered By: **Selected JCR Year: 2018** Selected Editions: SCIE,SSCI
Selected Categories: **"BIOTECHNOLOGY and APPLIED MICROBIOLOGY"**
Selected Category Scheme: WoS
Gesamtanzahl: 162 Journale

Rank	Full Journal Title	Total Cites	Journal Impact Factor	Eigenfactor Score
1	NATURE REVIEWS DRUG DISCOVERY	32,266	57.618	0.054890
2	NATURE BIOTECHNOLOGY	60,971	31.864	0.158140
3	GENOME BIOLOGY	38,920	14.028	0.133310
4	TRENDS IN BIOTECHNOLOGY	15,857	13.747	0.019230
5	BIOTECHNOLOGY ADVANCES	18,021	12.831	0.020270
6	GENOME RESEARCH	39,240	9.944	0.079580
7	BIOSENSORS & BIOELECTRONICS	57,168	9.518	0.077810
8	MOLECULAR THERAPY	16,991	8.402	0.030050
9	CURRENT OPINION IN BIOTECHNOLOGY	15,326	8.083	0.023020
10	METABOLIC ENGINEERING	7,967	7.808	0.016120
11	CRITICAL REVIEWS IN BIOTECHNOLOGY	3,785	7.054	0.005370
12	PLANT BIOTECHNOLOGY JOURNAL	7,866	6.840	0.016180
13	BIORESOURCE TECHNOLOGY	118,251	6.669	0.106420
14	Tissue Engineering Part B-Reviews	3,550	6.512	0.004970
15	npj Biofilms and Microbiomes	475	6.333	0.001800
16	MUTATION RESEARCH-REVIEWS IN MUTATION RESEARCH	3,566	6.081	0.004090
17	STEM CELLS	21,467	5.614	0.030220
18	Biotechnology for Biofuels	9,655	5.452	0.021070
19	JOURNAL OF NANOBIOTECHNOLOGY	3,534	5.345	0.005330
20	Annual Review of Animal Biosciences	709	5.200	0.003160

Rank	Full Journal Title	Total Cites	Journal Impact Factor	Eigenfactor Score
21	REVIEWS IN ENVIRONMENTAL SCIENCE AND BIO-TECHNOLOGY	2,582	4.938	0.003010
22	Microbial Biotechnology	3,576	4.857	0.006790
23	Nanomedicine	7,763	4.717	0.011840
24	CANCER GENE THERAPY	2,842	4.681	0.003200
25	BIOINFORMATICS	107,600	4.531	0.205400
26	Current Opinion in Chemical Engineering	1,656	4.463	0.004210
27	Artificial Cells Nanomedicine and Biotechnology	3,209	4.462	0.003030
28	Microbial Cell Factories	7,567	4.402	0.013950
29	CYTOTHERAPY	5,969	4.297	0.009690
30	BIOTECHNOLOGY AND BIOENGINEERING	26,077	4.260	0.019030
31	Biofuels Bioproducts & Biorefining-Biofpr	3,165	4.224	0.003440
32	FOOD MICROBIOLOGY	10,786	4.089	0.012130
33	APPLIED AND ENVIRONMENTAL MICROBIOLOGY	105,845	4.077	0.064410
34	Stem Cell Research	3,077	3.929	0.008410
35	HUMAN GENE THERAPY	5,639	3.855	0.007540
36	INTERNATIONAL BIODETERIORATION & BIODEGRADATION	10,708	3.824	0.011950
37	GENE THERAPY	7,223	3.749	0.007040
38	New Biotechnology	2,938	3.739	0.004640
39	Algal Research-Biomass Biofuels and Bioproducts	4,778	3.723	0.010660
40	Advances in Applied Microbiology	1,695	3.700	0.001850
41	APPLIED MICROBIOLOGY AND BIOTECHNOLOGY	43,697	3.670	0.047580

Rank	Full Journal Title	Total Cites	Journal Impact Factor	Eigenfactor Score
42	MOLECULAR PLANT-MICROBE INTERACTIONS	11,283	3.649	0.008960
43	Tissue Engineering Part A	9,754	3.616	0.012380
44	EXPERT OPINION ON BIOLOGICAL THERAPY	4,481	3.585	0.007720
45	ENZYME AND MICROBIAL TECHNOLOGY	11,801	3.553	0.005540
46	Biotechnology Journal	5,615	3.543	0.009790
47	BIOMASS & BIOENERGY	20,720	3.537	0.017380
48	BMC GENOMICS	39,876	3.501	0.095280
49	BIOCHEMICAL ENGINEERING JOURNAL	10,223	3.371	0.009680
50	GM Crops & Food-Biotechnology in Agriculture and the Food Chain	328	3.333	0.000830
51	FOOD AND BIOPRODUCTS PROCESSING	3,846	3.324	0.004930
52	Journal of Tissue Engineering and Regenerative Medicine	4,776	3.319	0.007680
53	JOURNAL OF BIOTECHNOLOGY	15,934	3.163	0.017610
54	GENOMICS	8,698	3.160	0.005530
55	Briefings in Functional Genomics	1,641	3.133	0.003270
56	OncoTargets and Therapy	7,177	3.046	0.018620
57	JOURNAL OF INDUSTRIAL MICROBIOLOGY & BIOTECHNOLOGY	7,889	2.993	0.007040
58	Probiotics and Antimicrobial Proteins	716	2.962	0.000940
59	PROCESS BIOCHEMISTRY	17,457	2.883	0.010630
60	BIOFOULING	4,157	2.847	0.004290
61	JOURNAL OF GENERAL VIROLOGY	18,927	2.809	0.018780
62	SYSTEMATIC AND APPLIED MICROBIOLOGY	5,118	2.808	0.004490

Rank	Full Journal Title	Total Cites	Journal Impact Factor	Eigenfactor Score
63	Environmental Technology & Innovation	545	2.800	0.000850
64	MARINE BIOTECHNOLOGY	3,176	2.798	0.002720
65	DISEASE MARKERS	3,479	2.761	0.006820
66	Pharmacogenetics and Genomics	3,177	2.693	0.003410
67	JOURNAL OF APPLIED MICROBIOLOGY	18,997	2.683	0.012760
68	Journal of Biological Engineering	1,076	2.667	0.002110
69	JOURNAL OF CHEMICAL TECHNOLOGY AND BIOTECHNOLOGY	11,188	2.659	0.009380
70	WORLD JOURNAL OF MICROBIOLOGY & BIOTECHNOLOGY	8,751	2.652	0.007780
71	Tissue Engineering Part C-Methods	3,268	2.638	0.005250
72	JOURNAL OF APPLIED PHYCOLOGY	9,070	2.635	0.008290
73	OMICS-A JOURNAL OF INTEGRATIVE BIOLOGY	2,021	2.610	0.002550
74	BIOLOGICAL CONTROL	7,567	2.607	0.006780
75	Human Vaccines & Immunotherapeutics	5,467	2.592	0.019690
76	MICROBES AND ENVIRONMENTS	1,732	2.575	0.002510
77	Biotechnology & Genetic Engineering Reviews	524	2.571	0.000220
78	BIODEGRADATION	2,918	2.534	0.001670
79	BMC BIOINFORMATICS	30,607	2.511	0.045720
79	MOLECULAR AND CELLULAR PROBES	1,950	2.511	0.002110
81	FEMS YEAST RESEARCH	4,334	2.458	0.005730
82	JOURNAL OF ANTIBIOTICS	8,442	2.446	0.005350
83	BIOTECHNOLOGY PROGRESS	8,426	2.406	0.005710
84	YEAST	4,355	2.395	0.002540

Rank	Full Journal Title	Total Cites	Journal Impact Factor	Eigenfactor Score
85	BIOPROCESS AND BIOSYSTEMS ENGINEERING	4,024	2.371	0.005990
86	MAMMALIAN GENOME	2,715	2.343	0.003410
87	Advances in Biochemical Engineering-Biotechnology	2,261	2.341	0.002540
88	BMC BIOTECHNOLOGY	3,504	2.303	0.004970
88	International Journal of Genomics	740	2.303	0.002170
90	JOURNAL OF BIOMOLECULAR SCREENING	2,934	2.297	0.004290
91	MUTATION RESEARCH-GENETIC TOXICOLOGY AND ENVIRONMENTAL MUTAGENESIS	6,318	2.256	0.003920
92	AMB Express	1,906	2.226	0.003920
93	PLANT CELL TISSUE AND ORGAN CULTURE	6,382	2.200	0.005350
94	Biomed Research International	36,776	2.197	0.109840
95	SLAS Discovery	347	2.192	0.001130
96	BIOTECHNOLOGY LETTERS	10,296	2.154	0.007690
97	GENOME	4,264	2.152	0.002730
98	APPLIED BIOCHEMISTRY AND BIOTECHNOLOGY	12,340	2.140	0.013010
99	Human Gene Therapy Methods	410	2.089	0.001220
100	ELECTRONIC JOURNAL OF BIOTECHNOLOGY	1,692	2.040	0.001820
101	JOURNAL OF BIOSCIENCE AND BIOENGINEERING	8,324	2.032	0.007120
102	MUTATION RESEARCH-FUNDAMENTAL AND MOLECULAR MECHANISMS OF MUTAGENESIS	7,025	2.011	0.003490
103	PROTEIN ENGINEERING DESIGN & SELECTION	4,865	1.980	0.003540
104	JOURNAL OF BIOACTIVE AND COMPATIBLE POLYMERS	1,057	1.976	0.000830

Publikation 2

Seipelt M, Möller A, Nawka T, Gonnermann U, **Caffier F**, Caffier PP.
Monitoring the Outcome of Phonosurgery and Vocal Exercises with Established and New Diagnostic Tools.
Biomed Res Int 2020 Jan 23;2020:4208189. https://doi.org/10.1155/2020/4208189

JCR Impact Factor 2020: 3.411 (Q2; 69/159 Biotechnology & applied Microbiology)

Hindawi
BioMed Research International
Volume 2020, Article ID 4208189, 10 pages
https://doi.org/10.1155/2020/4208189

Clinical Study

Monitoring the Outcome of Phonosurgery and Vocal Exercises with Established and New Diagnostic Tools

Matthias Seipelt,[1] Andreas Möller,[2] Tadeus Nawka,[1] Ute Gonnermann,[3] Felix Caffier,[1] and Philipp P. Caffier[1]

[1]*Department of Audiology and Phoniatrics, Charité-University Medicine Berlin, Campus Charité Mitte, Chariteplatz 1, D-10117 Berlin, Germany*
[2]*Max-Planck Institute for Plasma Physics, Wendelsteinstraße 1, D-17491 Greifswald, Germany*
[3]*ENT Department, University of Greifswald, Fleischmannstraße 8, D-17475 Greifswald, Germany*

Correspondence should be addressed to Philipp P. Caffier; philipp.caffier@charite.de

Received 14 August 2019; Accepted 20 December 2019; Published 23 January 2020

Academic Editor: Jan Plzak

Instrument-assisted measuring procedures expand the options within phoniatric diagnostics by quantifying the condition of the voice. The aim of this study was to examine objective treatment-associated changes of the recently developed vocal extent measure (VEM) and the established dysphonia severity index (DSI) in relation to subjective tools, i.e., self-evaluation via voice handicap index (VHI-12) and external evaluation via auditory-perceptual assessment of hoarseness (*H*). The findings for *H* (3 raters' group assessment), VHI-12, DSI, and VEM in 152 patients of both sexes (age range 16–75 years), taken before and 3 months after phonosurgery or vocal exercises, were compared and correlated. Posttherapeutically, all of the recorded parameters improved ($p < 0.001$). The degree of *H* reduced on average by 0.5, the VHI-12 score sank by 5 points, while DSI and VEM rose by 1.5 and 19, respectively. The correlations of these changes were significant but showed gradual differences between *H* and VHI-12 ($r = 0.3$), *H* and DSI ($r = -0.3$), and *H* and VEM ($r = -0.4$). We conclude that all investigated parameters are adequate to verify therapeutic outcomes but represent different dimensions of the voice. However, changes in the degree of *H* as gold standard were best recognized with the new VEM.

1. Introduction

The human voice is a very complex phenomenon that is difficult to quantify [1–3]. According to the basic protocol of the European Laryngological Society, a comprehensive assessment of vocal function can be gained using a multidimensional diagnostic approach [4]. Several measurements are recommended for voice evaluation, comprising subjective procedures such as self-assessment of the voice and external auditory-perceptual judgment, as well as objective procedures such as voice range profile (VRP) measurements, acoustic-aerodynamic analysis, and videolaryngostroboscopy (VLS).

In order to quantify the self-experienced extent of a vocal problem, the subjective impairment can be assessed using standardized questionnaires [5, 6]. The original voice handicap index (VHI) consists of 30 questions addressing functional, physical, or emotional aspects in the context of dysphonia [7]. Shorter VHI versions were designed because many patients perceive the answering of 30 questions tedious and partly redundant [8]. The 9-item VHI-9i and the 12-item VHI-12 had been created after original item reduction based on factor analysis and test-retest validation. They represent reliable, commonly used questionnaires with improved acceptance and practicability in clinical routine [9]. Regardless of the subjective self-evaluation, the examiner's auditory perception of the patient's voice is considered in many medical studies as the gold standard for voice assessment [10–12]. Different evaluation systems were developed, assessing various parameters including the

Common pillars in voice diagnostics:
DSI: dysphonia severity index
MPT: maximum phonation time
RBH: roughness (*R*), breathiness (*B*), and (overall grade of) hoarseness (*H*)
VEM: vocal extent measure
VHI: voice handicap index
VLS: videolaryngostroboscopy
VRP: voice range profile

Figure 1: List of common abbreviations in voice diagnostics.

perceived roughness (*R*), breathiness (*B*), and the overall grade of hoarseness (*H*). The application of the RBH scale is considered to be reliable, particularly when group assessments are used for further analysis [13–15].

The inclusion of instrument-assisted measurement procedures can support and usefully expand the diagnostic investigation by objectively quantifying the current condition of the voice [16, 17]. The established dysphonia severity index (DSI) is calculated as a weighted combination of the highest possible fundamental frequency (*F*0_max), the lowest phonation intensity (*I*_min), maximum phonation time (MPT), and jitter [18]. Since the DSI quantifies dysphonia as a negative criterion and involves the risk of inaccurate results due to its multidimensional acquisition, we recently developed the one-dimensional vocal extent measure (VEM) for objective VRP evaluation [19]. The VEM quantifies the subject's dynamic performance and frequency range and is calculated as a relation of area and perimeter of the VRP. The VEM describes the vocal abilities and enables a classification of voice performance as a positive criterion [20]. A list of common abbreviations in voice diagnostics is given in Figure 1.

For comprehensive documentation of vocal status and treatment, it is necessary to ensure that changes in voice quality can be adequately identified. For this purpose, the measurement data used must be sensitive to the slightest changes in voice quality, and the registration equipment must be able to detect them. In order to investigate the suitability of objective and subjective parameters for the assessment of vocal improvement after phonosurgery and vocal exercises, the changes in DSI and VEM values should be monitored and compared with those of the subjective auditory perception via RBH and VHI-12.

2. Materials and Methods

A total of 152 patients with various voice problems underwent therapy in a clinical prospective study. The trial was conducted in accordance with the Declaration of Helsinki and on approval by the local ethical review board. All data were collected at the pretherapeutic visit and three months after the intervention. According to the diagnosed clinical pathology and the previous course of the disease, the patients received either surgical treatment or conservative vocal exercises. Logopedic voice therapy was conducted by qualified speech therapists and included 20 sessions (2 times per week, for 45 minutes). Phonomicrosurgery was performed by 3 experienced senior phonosurgeons via direct microlaryngoscopy in general anesthesia (TIVA with propofol/remifentanil).

Various established examination instruments were applied to evaluate the treatment outcome. Digital VLS was performed using a high-resolution rigid video laryngoscope (10 mm; 70°) with an integrated microphone (XION medical, Berlin, Germany) [21]. Laryngoscopy served to discriminate between organic dysphonia and functional dysphonia. Stroboscopy visualized the vocal fold vibrations during phonation and indicated impairment by showing reduced/absent mucosal wave propagation or reduced/eliminated phonatory vibration.

The LingWAVES program (WEVOSYS, Forchheim, Germany) was used for standardized registration of the VRP and acoustic-aerodynamic analysis. Several acoustic and aerodynamic parameters were recorded, such as I_min, F0_max, MPT, and jitter. Based on the defined combination of these parameters, the DSI was calculated to classify the voice into nondysphonic (≥ 4.2) versus mildly (<4.2 to ≥ 1.8), moderately (<1.8 to ≥ -1.2), or severely (< -1.2) dysphonic [22]. In addition, the VEM as a recently introduced new diagnostic tool for the objective assessment of vocal capacity was computed [19, 23]. VEM calculation was done after VRP measurement by a proprietary software program (AVA) which can extract various other parameters from the VRP, thereby enabling VRP comparisons [24]. The VEM multiplies the VRP area by the quotient of the theoretical perimeter of a circle with the VRP surface area and the actual VRP circumference. The mathematical derivation of the equation of this measure is explained elsewhere [19]. The VEM quantifies the dynamic performance and frequency range of the voice by a one-dimensional, interval-scaled value without unit, typically between 0 and 120. These limits may be exceeded at both ends ($\text{VEM}_{\text{min}} \geq -150$; $\text{VEM}_{\text{max}} \leq 150$), describing a small vocal capacity by a small VEM and a large VRP by a high VEM.

The VHI-12 was applied for the patient's subjective self-assessment of the own voice [9]. Study participants rated all 12 questions on a scale from 0 to 4 (0: never, 1: almost never, 2: sometimes, 3: almost always, 4: always), followed by one question concerning the overall voice impairment at the present time (VHIs) on a scale from 0 to 3 (0: normal, 1: mild, 2: moderate, 3: severe). The VHI-12 total score allowed an impairment-related severity classification (0–7: no dysphonia, 8–14: mild dysphonia, 15–22: moderate dysphonia, 23–48: severe dysphonia).

External auditory-perceptual voice evaluation was assessed using the RBH system when the patients were reading the standardized text "The north wind and the sun" (German version). The perceived roughness (*R*), breathiness (*B*), and overall grade of hoarseness (*H*) of the patients voices were scored on a scale from 0 to 3 (0: not existing, 1: mild, 2: moderate, 3: severe) by three experienced examiners (one phoniatric physician, one clinical linguist, and one biomedical engineer). To enhance the evaluation objectivity, all audio recordings were rated independently in one session after being shuffled and blinded regarding the patient assignment and pre-/posttherapeutic status. The degree of *H* served as a gold standard to provide an indication of the therapy success.

The outcome analysis was based on pre- and posttherapeutic voice function diagnostics and VLS. The

parameters H, VHI-12, VHIs, DSI, and VEM were compared with each other before and three months after the intervention, as well as their changes. It was tested whether the therapy resulted in a significant difference in the parameters measured. In addition, the measurement data were correlated with each other before and after therapy, as were the respective changes. Statistical methods applied were the calculation of Spearman's rank-order correlation coefficients (r), as well as the paired t-test. The level of significance was set at $\alpha = 0.05$.

3. Results

Altogether, 152 patients were examined before and after therapeutic treatment: 102 females (17–70 years, median 48) and 50 males (16–75 years, median 42). A total of 304 data sets were collected. Subjects of both sexes were comparable in terms of age, sociodemographic characteristics, hoarseness level (H), and underlying pathologies. Sixty-six individuals (43%) used their voice in a nonprofessional manner (e.g., clerks, IT-specialists, and laborers), whereas 86 patients (57%) had a high vocal strain in their profession (e.g., teachers, actors, and singers). Pretherapeutically, the patients exhibited various clinical disorders. VLS revealed in 101 subjects (66%) organic diseases at vocal fold level. Classification of the resulting organic dysphonia according to the underlying pathology showed in 41 patients (27%) diseases of the lamina propria (e.g., nodules, polyps, cysts, and edema), in 24 patients (16%) movement disorders (vocal fold paralysis, spasmodic dysphonia), in 19 patients (12%) diseases of the epithelium (e.g., leukoplakia, hyperkeratosis, carcinoma, and papillomatosis), and in 5 patients (3%) arytenoid pathologies (granuloma). Fifty-one participants (34%) had normal laryngeal anatomy but suffered from a vocal load-induced functional dysphonia. Altogether, 46 subjects (30%) had initially no hoarseness (H0), including 29 patients with functional dysphonia, 13 patients with small glottal findings (marginal edema, nodules, and leukoplakia), and 4 patients with pathologies distant from the vocal fold level (arytenoid granuloma). A summary of relevant pretherapeutic patient characteristics, pathology classification according to Rosen and Murry [25], and a listing of all diagnoses are shown in Table 1.

Posttherapeutically, all investigated vocal parameters had improved. Regarding subjective evaluation, the mean RBH status exposed less roughness, breathiness, and overall grade of hoarseness ($p < 0.001$). H decreased from 1.2 to 0.7, changing in most patients from H1 to H0 ($n = 29$; i.e., 19%) and from H2 to H1 ($n = 22$; i.e., 14%). The VHI-12 reduced on average from 15 to 10, corresponding to a self-assessed improvement from moderately to mildly impaired ($p < 0.001$). Respectively, the overall VHIs score sank from 1.4 to 0.7, changing most often from mild voice impairment to normal ($n = 44$; i.e., 29%) and from moderate to mild voice impairment ($n = 35$; i.e., 23%). Figure 2 summarizes the mean pre- and posttherapeutic data in all patients for the investigated subjective vocal parameters using column diagrams.

A comparison of objective parameters revealed for the DSI a mean increase from 2.2 to 3.7, showing significant improvement ($p < 0.001$) that remained at the level of mild dysphonia. The VEM rose from 60 to 79, reflecting voice improvement with significantly enhanced vocal capacity ($p < 0.001$). Figure 3 illustrates the pre- and posttherapeutic data for both objective voice parameters using boxplots. It indicates additionally the different distribution of DSI and VEM in relation to the degree of H.

A comparison of treatment groups revealed that phonosurgery had the largest impact on voice function with higher numerical improvement of subjective and objective parameters. A patient example demonstrating phonosurgery-induced changes of laryngeal and vocal findings is presented in Figure 4.

To evaluate the extent of treatment-related benefits, Table 2 shows the mean differences between pre- and posttherapeutic values and the 95% confidence intervals for them. The numeric outcome of the values after conservative logopedic therapy was much smaller, but the vocal capabilities improved in most patients, too. Furthermore, Table 2 displays the pre- and posttherapeutic comparison concerning both dysphonia groups (functional/organic) and all pathology classification subgroups. In general, age and gender had no significant influence on the treatment outcome.

The correlation of the parameters H, VHI-12, VHIs, DSI, and VEM with each other showed a significant ($p < 0.01$) but weak to moderate linear relationship. The strength of the relationship changed only slightly due to the therapy. However, the DSI proved an exception in this regard. While the DSI data before treatment showed a moderate negative relationship with H ($r = -0.4$) and moderate positive relationship with VEM ($r = 0.6$), these correlations decreased considerably after therapy, revealing weaker relationships for H ($r = -0.3$) and VEM ($r = 0.3$). The weak relationship between DSI and VHI-12 as well as DSI and VHIs did not show relevant changes posttherapeutically. In contrast, the VEM correlated with the VHI-12 at $r = -0.4$ and with H at $r = -0.7$, revealing moderate and strong negative relationships, irrespective of the therapy status. Furthermore, H and VHI-12 correlated after therapy at $r = 0.4$, and H and VHIs at $r = 0.5$.

The investigation of therapy-induced changes (Δ) in the individual measurement data indicated that correlations of these changes resulted in rather small coefficients for all parameters. The relationship between ΔDSI and ΔH was $r = -0.3$ ($p < 0.01$). The ΔDSI showed no significant relationship to ΔVHI-12 ($r = -0.04$) and ΔVHIs ($r = -0.09$). The relationship between ΔVEM and ΔH was $r = -0.4$ ($p < 0.001$). In contrast to DSI, ΔVEM revealed also a significant relationship to ΔVHI-12 and ΔVHIs ($r = -0.2$ each, $p < 0.01$). Besides, ΔH and ΔVHI-12 correlated at $r = 0.3$, and ΔH and ΔVHIs at $r = 0.4$ ($p < 0.01$). Finally, the relationship between ΔDSI and ΔVEM was moderate at $r = 0.5$ ($p < 0.01$). A summary of all correlation results can be found in Table 3.

4. Discussion

It was possible to show that all parameters under investigation reacted to the therapy and improved on average, thus

Table 1: Pretherapeutic patient characteristics.

Characteristics	No. of all patients	% of total group ($n = 152$)	No. of male patients	% of male group ($n = 50$)	No. of female patients	% of female group ($n = 102$)
Gender						
Male	50	33	—	—	—	—
Female	102	67	—	—	—	—
Age						
Years (mean ± SD)	45 ± 16	—	43 ± 17	—	45 ± 15	
Main voice use						
Nonprofessional	66	43	22	44	44	43
Professional	86	57	28	56	58	57
Sociodemographic						
Scholar	10	7	4	8	6	6
Student/apprentice	14	9	7	14	7	7
Employed	97	64	29	58	68	66
Unemployed	8	5	2	4	6	6
Pensioner	23	15	8	16	15	15
Overall hoarseness level (H)						
*H*0 (not existing)	46	30	12	24	34	33
*H*1 (mild)	6	40	21	42	39	38
*H*2 (moderate)	27	18	10	20	17	17
*H*3 (severe)	19	12	7	14	12	12
*Pathology classification**						
Functional dysphonia	51	34	15	30	36	35
Organic dysphonia	101	66	35	70	66	65
Rosen I (epithelium)	19	12	10	20	9	9
Rosen II (lamina propria)	41	27	7	14	34	33
Rosen III (arytenoid)	5	3	3	6	2	2
Rosen IV (other)	36	24	15	30	21	21
Organic diagnosis						
Vocal fold paralysis	18	11.8	9	18.0	9	8.8
Vocal fold nodules	13	8.6	0	—	13	12.7
Vocal fold polyp	9	5.9	4	8.0	5	4.9
Reinke's edema	9	5.9	0	—	9	8.8
Laryngeal papillomatosis	8	5.3	3	6.0	5	4.9
Marginal edema	6	3.9	2	4.0	4	3.9
Spasmodic dysphonia	6	3.9	0	—	6	5.9
Contact granuloma	5	3.2	3	6.0	2	2.0
Vocal fold atrophy	5	3.2	4	8.0	1	1.0
Hyperkeratosis	4	2.6	3	6.0	1	1.0
Leukoplakia	4	2.6	3	6.0	1	1.0
Sulcus vocalis	3	2.0	2	4.0	1	1.0
Glottal carcinoma (pT1a)	3	2.0	1	2.0	2	2.0
Vocal fold cyst	3	2.0	1	2.0	2	2.0
Varix cordis	1	0.7	0	—	1	1.0
Laryngotracheal stenosis	1	0.7	0	—	1	1.0
Glottal web	1	0.7	0	—	1	1.0
Traumatic laryngeal fracture	1	0.7	0	—	1	1.0
Bamboo nodes	1	0.7	0	—	1	1.0

Unless otherwise specified, data expressed as number of patients and percentage of group. *Pathology classification according to Rosen and Murry [25], i.e., I: epithelium, e.g., leukoplakia, hyperkeratosis, CIS (=carcinoma in situ), carcinoma, and papillomatosis. II: lamina propria, e.g., Reinke's edema, polyps, cysts, scars, and vascular malformation. III: arytenoid, e.g., granuloma and infection. IV: other, including movement disorders, hypo-/atrophy, and malformation as e.g., sulcus or glottal web.

presenting their general suitability for documentation of the therapeutic process. However, due to the individual construction and intention, each objective and subjective parameter performed differently. The established DSI represents a weighted sum of I_min, F0_max, MPT, and jitter [18] and therefore integrates parameters of VRP, aerodynamic, and acoustic measures. As assumed, most of our patients showed DSI values ranging from −5 to 5, whereby −5 corresponded to a very dysphonic voice and 5 to a perceptual normal voice. Due to the special structure of our patient cohort including a considerable number of elite vocal performers and subjects with extremely dysphonic voices, more study participants than expected (34%) had initial DSI values which exceeded these boundaries at both ends. After

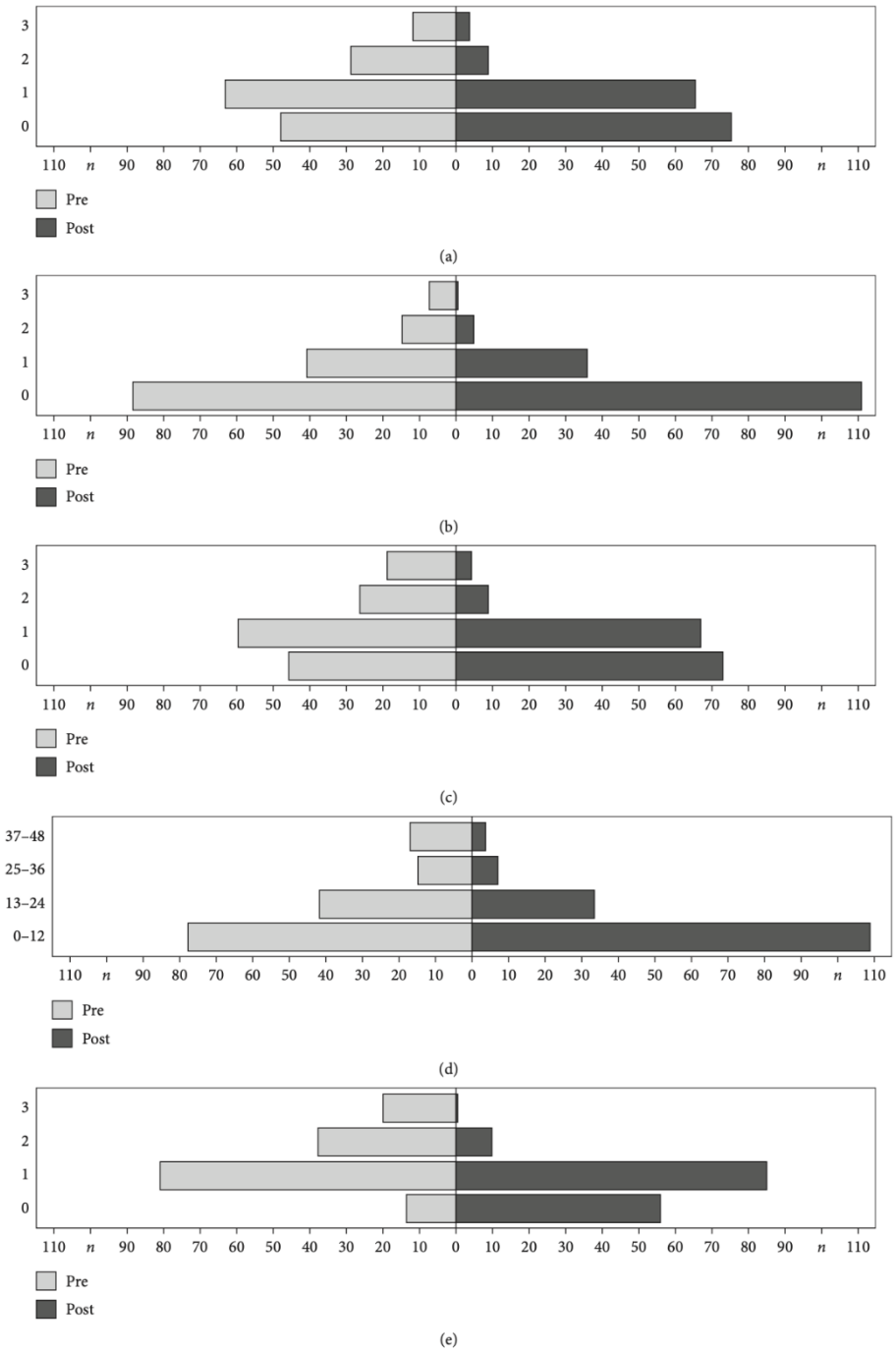

FIGURE 2: Subjective vocal parameters: (a) *R*, (b) *B*, (c) *H*, (d) VHI-12, and (e) VHIs before treatment (light grey, left columns) and after treatment (dark grey, right columns). The abscissae show the number of patients (*n*), and the ordinates represent the scales of the RBH system and VHIs system (0–3) as well as the VHI-12 score.

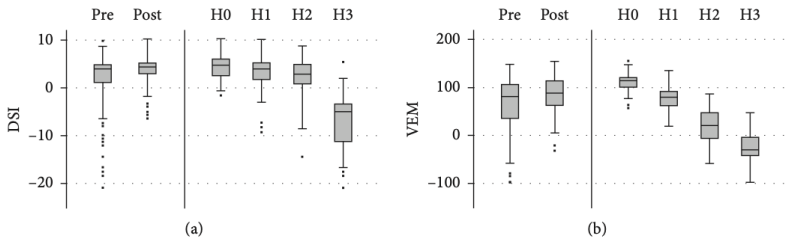

Figure 3: Objective vocal parameters: (a) DSI and (b) VEM before and after treatment, as well as their distribution according to the degree of *H*. The boxplots display the median, quartiles, range of values covered by the data, and any outliers (single spots).

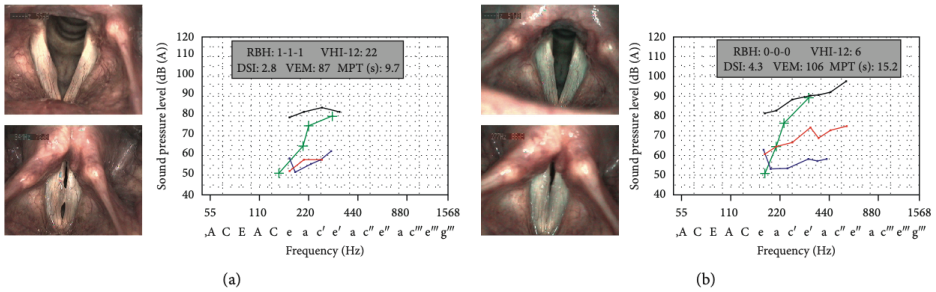

Figure 4: Example of phonosurgery-induced changes of laryngeal and vocal findings in a 49-year-old female dental assistant suffering from persisting dysphonia and dysodia. (a) Preoperative VLS shows a marginal edema of the right vocal fold with a glottal gap during phonation, asynchronous oscillations, and impaired mucosal wave propagation. The preoperative VRP pattern displays envelope curves for the loudest (black lines) and softest (blue lines) singing voice and for the speaking voice at different vocal intensity levels (green lines) with little dynamic and frequency ranges. The singer's formant levels (red lines) are low, characterizing the impaired concentration of acoustic energy by resonator amplification of certain frequency ranges in the vocal tract. The values of all objective voice parameters (DSI: dysphonia severity index; VEM: vocal extent measure; MPT: maximum phonation time) and subjective voice parameters (RBH: roughness, breathiness, overall grade of hoarseness; VHI-12: twelve-item voice handicap index) are reduced. (b) Three months after phonomicrosurgical removal of the edema, the treated vocal fold shows a straight margin. The glottal closure is complete, and the oscillations have normalized (mucosal wave propagation regular and symmetric). The patient reveals higher dynamic and frequency ranges of speaking and singing voice with considerably improved objective and subjective parameters.

Table 2: Changes in vocal measures after treatment for all patients and separated for both intervention groups (logopedics/phonosurgery), both dysphonia groups (functional/organic), and all pathology classification subgroups according to Rosen and Murry [25].

	H	VHI-12	VHIs	DSI	VEM
Total group of patients ($n = 152$)	−0.5 (−0.6; −0.4)	−4.8 (−6.2; −3.3)	−0.7 (−0.8; −0.6)	1.5 (1.0; 2.0)	18.7 (13.4; 24.1)
Logopedic treatment group ($n = 79$)	−0.3 (−0.4; −0.2)	−1.4 (−2.6; −0.1)	−0.5 (−0.7; −0.4)	0.2 (−0.1; 0.5)	6.5 (1.6; 11.4)
Phonosurgery group ($n = 73$)	−0.8 (−1.0; −0.6)	−8.5 (−10.9; −6.0)	−0.9 (−1.1; −0.7)	2.8 (1.9; 3.7)	32.0 (22.9; 41.0)
Functional dysphonia group ($n = 51$)	−0.3 (−0.4; −0.2)	−0.3 (−1.4; 0.7)	−0.5 (−0.7; −0.4)	0.1 (−0.2; 0.4)	3.6 (0.1; 7.1)
Organic dysphonia group ($n = 101$)	−0.7 (−0.8; −0.5)	−7.0 (−9.0; −5.0)	−0.8 (−1.0; −0.7)	2.2 (1.5; 2.9)	26.4 (18.9; 33.9)
Rosen I subgroup (epithelium) ($n = 19$)	−0.7 (−1.2; −0.3)	−5.1 (−9.6; −0.6)	−1.0 (−1.3; −0.7)	2.0 (0.5; 3.6)	30.2 (10.1; 50.3)
Rosen II subgroup (lamina propria) ($n = 41$)	−0.6 (−0.7; −0.4)	−4.6 (−6.8; −2.4)	−0.7 (−0.9; −0.5)	1.2 (0.2; 2.1)	15.7 (7.3; 24.1)
Rosen III subgroup (arytenoid) ($n = 5$)	−0.1 (−0.4; 0.3)	−1.6 (−6.0; 2.8)	−0.3 (−0.9; 0.3)	0.2 (−0.8; 0.9)	2.5 (−11.8; 16.7)
Rosen IV subgroup (other) ($n = 36$)	−0.8 (−1.1; −0.5)	−11.5 (−15.7; −7.4)	−1.0 (−1.2; −0.7)	3.9 (2.6; 5.3)	39.9 (24.7; 55.1)

Data expressed as mean differences of preoperative and postoperative values (upper line), with 95% confidence intervals (lower line, in brackets). VEM = vocal extent measure; DSI = dysphonia severity index; VHI-12 = voice handicap index; VHIs = self-perceived impairment of voice at the present time.

therapy, the mean DSI measurement data did not improve by an average of one degree of severity [22] but still showed a significant increase. This confirms previous studies, which describe the DSI as a useful parameter to measure the severity of dysphonia and the improvement after therapy [26–29]. However, various studies could show that the DSI is influenced by differences in measurements of the registration programs as well as by age or gender [19, 30–32]. Therefore, we developed and investigated the VEM as a new objective vocal parameter unimpaired by these interacting factors.

Table 3: Results of correlation analysis.

		DSI	VEM	H	VHI-12	VHIs
Pretherapeutic	DSI	1	0.6	−0.4	−0.2	−0.3
	VEM	0.6	1	−0.7	−0.4	−0.3
	H	−0.4	−0.7	1	0.4	0.4
	VHI-12	−0.2	−0.4	0.4	1	0.6
	VHIs	−0.3	−0.3	0.4	0.6	1
Posttherapeutic	DSI	1	0.3	−0.3	−0.3	−0.3
	VEM	0.3	1	−0.7	−0.3	−0.4
	H	−0.3	−0.7	1	0.4	0.5
	VHI-12	−0.3	−0.3	0.4	1	0.7
	VHIs	−0.3	−0.4	0.5	0.7	1
Therapy-induced changes (Δ)	DSI	1	0.5	−0.3	−0.04 (ns)	−0.09 (ns)
	VEM	0.5	1	−0.4	−0.2	−0.2
	H	−0.3	−0.4	1	0.3	0.4
	VHI-12	−0.04 (ns)	−0.2	0.3	1	0.6
	VHIs	−0.09 (ns)	−0.2	0.4	0.6	1

All correlation coefficients were significant ($p < 0.01$), unless otherwise specified (ns = not significant).

The VEM calculation is based on the size and shape of the VRP, instead of gathering data from a combination of different objective parameters [19, 24]. Most of our patients showed VEM values between 0 and 120. As expected, these limits were exceeded: (1) at the upper end in functionally impaired singers with professionally trained "great" voices and very large VRP, and (2) at the lower end in extremely dysphonic and nearly aphonic voices with very small VRP. These findings support the underlying idea, during the construction of this vocal parameter, that the ideal VRP should not show abrupt differences in the dynamic range of notes produced by the patients along with their frequency range [24]. A well-balanced dynamic extent approximated the optimal VRP shape to a circle where the area is the biggest for a given perimeter compared to other geometric figures [19]. Our results confirmed that larger and "smoother" VRP without relevant "jumps" in intensity achieved higher values. Vocal capacity quantified in this way, i.e., as a relation of area and perimeter of the VRP, showed a very distinct increase during the course of therapy in our study participants. This is in line with the results of the very few phonosurgical studies assessing VEM values in patients with Reinke's edema [20], vocal fold polyps [23], and nodules [33]. All of these investigations observed significantly increased VEM values after treatment.

Concerning subjective parameters, the VHI-12 successfully quantified the self-experienced extent of the vocal problem in our patients. Due to therapy, most of them rated an improvement from moderately to mildly impaired. These results correspond to other studies using VHI questionnaires for the investigation of surgical and conservative treatment success in organic dysphonia and functional dysphonia [20, 23, 33–36]. Our overall impression supports the general acknowledgment that short-form VHI versions represent reliable instruments with excellent acceptance and practicability in clinical routine [8, 9]. Compared to previous investigations, the examiners' auditory perception was the main indicator in our study for the assessment of therapy success [10–13]. We consider the RBH system to be reliable, particularly in case of evaluation by group assessments [14, 15, 23]. The mean RBH status of our patients' voices revealed significantly less roughness, breathiness, and overall grade of hoarseness. These results also confirmed the outcomes of former studies [23, 33–35]. Additionally, our analysis of the degree of H in relation to the objective parameters DSI and VEM revealed a better representation and graphical distinction of the auditory-perceptual assessment via the VEM. This is a new and important study finding which was confirmed in our investigations of correlation.

Correlations between the changes in individual parameters are able to show how well the improvement in one measurement value is reproduced by another measuring procedure. The generally weak correlations in our results can be explained by the different approaches to the individual parameters and are likewise a manifestation of the additional information content of the respective measurement methods. A relatively high correlation between parameters confirms the success of the therapy from the different aspects of these parameters. Regarding DSI, only a weak negative correlation with H and no significant relationship to VHI-12 could be found. Overall, this implies that although the DSI seems suitable for indicating the success of therapy, the increase of DSI has very little to do with the improvement in the degree of hoarseness and with the patient's perception of the vocal problem. This is also seen in the decreasing correlation of the values for DSI and H after therapy. The changes in the VHI-12, on the other hand, had a weak relationship to the changes of the VEM. Moreover, changes in the VEM demonstrated a moderately negative relationship to the changes of H, which means that the VEM increases as hoarseness decreases. Thus, in addition to the quantification of vocal capacity, the VEM is validated by the auditory assessment. The novel numeric description of the VRP by means of the interval-scaled VEM provides the researcher with a diagnostic parameter which is suitable for monitoring the course of treatment.

While interpreting these results, some limitations of our study should be considered. First, the number of patients was too small and the cohort was too heterogeneous to examine comparably sized groups of H levels, pathology

classification, or diagnosis-related subgroups. Therefore, there could be participation bias. Second, individual treatment recommendations depended on phoniatric indication and were based on comprehensive counseling related to clinical signs, symptoms, individual vocal requirements, abilities, and medical history. Nevertheless, at the end, the patients decided about the kind of intervention; thus, there may be selection bias. Third, our posttherapeutic follow-up of three months was too short to allow statements about the long-term outcome. Fourth, the investigated treatment modalities are often used in a combined mode to accelerate and optimize vocal improvement. We were not able to control whether patients after phonosurgery received hidden other therapies. Additional logopedic treatment or singing lessons are easily accessible and could influence the results especially in the recovery of operated patients. Therefore, there may be also performance bias. Finally, some well-known factors influencing the VRP registration have to be taken into account, such as the routine of the examiner, musicality and motivation of the patients, and the absence of generally accepted specifications regarding the number of registered tones. However, all VRPs were recorded by one experienced examiner under practically equal conditions, so that most of the mentioned factors can be ignored in this study.

Overall, our specific therapeutic outcomes confirmed the results of other studies investigating treatment effects in patients with various voice problems [33–39]. As expected, phonosurgery had the largest numeric impact on the improvement of vocal function. Conservative therapy provided smaller quantitative enhancements but often also qualitative vocal restoration with recovered artistic capabilities, particularly in singers with functional dysphonia. Logopedic training goals typically included reducing extrinsic laryngeal tension, using a relaxed laryngeal posture, and effective abdominal-diaphragmatic support for all phonation events [40]. Specific attention was given to the balance of respiratory forces, laryngeal coordination, and optimal filtering of the source signal via resonance and articulatory awareness [41, 42]. With this approach, also some of our patients with organic findings gained substantial voice improvement. As known from the literature, mainly younger patients with short duration of dysphonia and small benign pathologies of the lamina propria (e.g., vocal fold polyps, marginal edema) due to overuse benefitted from voice therapy [43–45].

5. Conclusions

The investigated parameters DSI, VEM, VHI, and RBH are all suitable for monitoring the course of voice treatment and adequate to quantify the outcomes of phonosurgery and logopedic vocal exercises. Correlation analysis confirms the clinical impression that DSI, VEM, VHI, and RBH represent different dimensions of the voice and are complementing objective or subjective measurements either for the evaluation of voice quality, vocal performance, or perceived vocal handicap. The VEM proves to be a comprehensible and easy-to-use parameter for objective VRP evaluation. Changes in the degree of hoarseness as gold standard were best recognized with the new VEM. Thus, in addition to the quantification of vocal capacity, the VEM is supported and validated by the auditory findings and provides an interval-scaled parameter for documentation.

Data Availability

The datasets generated and analyzed for this study are not publicly available, as they were obtained from a proprietary database via a licensing agreement.

Conflicts of Interest

The authors have no funding, financial relationships, or conflicts of interest to disclose.

Acknowledgments

The authors wish to thank Professor Eleanor Forbes for manuscript proofreading and language editing.

References

[1] B. Liu, E. Polce, H. Raj, and J. Jiang, "Quantification of voice type components present in human phonation using a modified diffusive chaos technique," *Annals of Otology, Rhinology & Laryngology*, vol. 128, no. 10, pp. 921–931, 2019.

[2] B. Barsties and M. De Bodt, "Assessment of voice quality: current state-of-the-art," *Auris Nasus Larynx*, vol. 42, no. 3, pp. 183–188, 2015.

[3] N. Roy, J. Barkmeier-Kraemer, T. Eadie et al., "Evidence-based clinical voice assessment: a systematic review," *American Journal of Speech-Language Pathology*, vol. 22, no. 2, pp. 212–226, 2013.

[4] P. H. Dejonckere, P. Bradley, P. Clemente et al., "A basic protocol for functional assessment of voice pathology, especially for investigating the efficacy of (phonosurgical) treatments and evaluating new assessment techniques," *European Archives of Oto-Rhino-Laryngology*, vol. 258, no. 2, pp. 77–82, 2001.

[5] M. Behlau, G. Madazio, F. Moreti et al., "Efficiency and cutoff values of self-assessment instruments on the impact of a voice problem," *Journal of Voice*, vol. 30, no. 4, pp. 506.e9–506.e18, 2016.

[6] M. Behlau, F. Zambon, F. Moreti, G. Oliveira, and E. de Barros Couto Jr., "Voice self-assessment protocols: different trends among organic and behavioral dysphonias," *Journal of Voice*, vol. 31, no. 1, pp. 112.e13–112.e27, 2017.

[7] B. H. Jacobson, A. Johnson, C. Grywalski et al., "The voice handicap index (VHI)," *American Journal of Speech-Language Pathology*, vol. 6, no. 3, pp. 66–70, 1997.

[8] C. A. Rosen, A. S. Lee, J. Osborne, T. Zullo, and T. Murry, "Development and validation of the voice handicap index-10," *The Laryngoscope*, vol. 114, no. 9, pp. 1549–1556, 2004.

[9] T. Nawka, I. M. Verdonck-de Leeuw, M. de Bodt et al., "Item reduction of the voice handicap index based on the original version and on European translations," *Folia Phoniatrica et Logopaedica*, vol. 61, no. 1, pp. 37–48, 2009.

[10] Y. Lee, G. Kim, and S. Kwon, "The usefulness of auditory perceptual assessment and acoustic analysis for classifying the voice severity," *Journal of Voice*, no. 19, pp. 30087–30096, 2019, In press.

[11] G. B. Kempster, B. R. Gerratt, K. Verdolini Abbott, J. Barkmeier-Kraemer, and R. E. Hillman, "Consensus auditory-perceptual evaluation of voice: development of a standardized clinical protocol," *American Journal of Speech-Language Pathology*, vol. 18, no. 2, pp. 124–132, 2009.

[12] R. Schönweiler, P. Wübbelt, M. Hess, and M. Ptok, "Psychoakustische skalierung akustischer stimmparameter durch multizentrisch validierte RBH-bewertung1," *Laryngo-Rhino-Otologie*, vol. 80, no. 3, pp. 117–122, 2001.

[13] M. Ptok, C. Schwemmle, C. Iven, M. Jessen, and T. Nawka, "Zur auditiven bewertung der stimmqualität," *HNO*, vol. 54, no. 10, pp. 793–802, 2006.

[14] H. Hanschmann and R. Berger, "Perceptual and acoustic evaluation of hoarseness," *Laryngorhinootologie*, vol. 90, no. 2, pp. 68–70, 2011.

[15] R. Schönweiler, M. Hess, P. Wübbelt, and M. Ptok, "Novel approach to acoustical voice analysis using artificial neural networks," *Journal of the Association for Research in Otolaryngology: JARO*, vol. 1, no. 4, pp. 270–282, 2000.

[16] P. H. Dejonckere, "Assessment of voice and respiratory function," in *Surgery of Larynx and Trachea*, M. Remacle and H. E. Eckel, Eds., Springer, Berlin, Germany, pp. 11–26, 2010.

[17] D. D. Mehta and R. E. Hillman, "Voice assessment: updates on perceptual, acoustic, aerodynamic, and endoscopic imaging methods," *Current Opinion in Otolaryngology & Head and Neck Surgery*, vol. 16, no. 3, pp. 211–215, 2008.

[18] F. L. Wuyts, M. S. D. Bodt, G. Molenberghs et al., "The dysphonia severity index," *Journal of Speech, Language, and Hearing Research*, vol. 43, no. 3, pp. 796–809, 2000.

[19] P. P. Caffier, A. Möller, E. Forbes, C. Müller, M. L. Freymann, and T. Nawka, "The vocal extent measure: development of a novel parameter in voice diagnostics and initial clinical experience," *BioMed Research International*, vol. 2018, Article ID 3836714, 10 pages, 2018.

[20] T. Salmen, T. Ermakova, A. Schindler et al., "Efficacy of microsurgery in Reinke's oedema evaluated by traditional voice assessment integrated with the vocal extent measure (VEM)," *Acta Otorhinolaryngologica Italica: Organo Ufficiale Della Societa Italiana di Otorinolaringologia E Chirurgia Cervico-Facciale*, vol. 38, no. 38, pp. 194–203, 2018.

[21] P. P. Caffier, B. Schmidt, M. Gross et al., "A comparison of white light laryngostroboscopy versus autofluorescence endoscopy in the evaluation of vocal fold pathology," *The Laryngoscope*, vol. 123, no. 7, pp. 1729–1734, 2013.

[22] U. Gonnermann, *Quantifizierbare Verfahren Zur Bewertung Von Dysphonien*, Peter Lang, Frankfurt/Main, Germany, 2007.

[23] T. Salmen, T. Ermakova, A. Möller et al., "The value of vocal extent measure (VEM) assessing phonomicrosurgical outcomes in vocal fold polyps," *Journal of Voice*, vol. 31, no. 1, pp. 114.e7–114.e15, 2017.

[24] P. Caffier and A. Möller, "Das stimmumfangsmaß SUM als neuer parameter in der objektiven stimmdiagnostik," *Sprache · Stimme · Gehör*, vol. 40, no. 4, pp. 183–187, 2016.

[25] C. A. Rosen and T. Murry, "Nomenclature of voice disorders and vocal pathology," *Otolaryngologic Clinics of North America*, vol. 33, no. 5, pp. 1035–1045, 2000.

[26] M. M. Hakkesteegt, M. P. Brocaar, and M. H. Wieringa, "The applicability of the dysphonia severity index and the voice handicap index in evaluating effects of voice therapy and phonosurgery," *Journal of Voice*, vol. 24, no. 2, pp. 199–205, 2010.

[27] K. Nemr, M. Simões-Zenari, G. S. de Souza, A. Hachiya, and D. H. Tsuji, "Correlation of the dysphonia severity index (DSI), consensus auditory-perceptual evaluation of voice (CAPE-V), and gender in Brazilians with and without voice disorders," *Journal of Voice*, vol. 30, no. 6, pp. 765.e7–765.e11, 2016.

[28] V. Uloza, B. B. V. Latoszek, N. Ulozaite-Staniene, T. Petrauskas, and Y. Maryn, "A comparison of dysphonia severity index and acoustic voice quality index measures in differentiating normal and dysphonic voices," *European Archives of Oto-Rhino-Laryngology*, vol. 275, no. 4, pp. 949–958, 2018.

[29] M. Brockmann-Bauser, B. Balandat, and J. E. Bohlender, "Immediate lip trill effects on the standard diagnostic measures voice range profile, jitter, maximum phonation time, and dysphonia severity index," *Journal of Voice*, no. 18, p. 30573, 2019, In press.

[30] M. M. Hakkesteegt, M. P. Brocaar, M. H. Wieringa, and L. Feenstra, "Influence of age and gender on the dysphonia severity index," *Folia Phoniatrica et Logopaedica*, vol. 58, no. 4, pp. 264–273, 2006.

[31] P. Aichinger, F. Feichter, B. Aichstill, W. Bigenzahn, and B. Schneider-Stickler, "Inter-device reliability of DSI measurement," *Logopedics Phoniatrics Vocology*, vol. 37, no. 4, pp. 167–173, 2012.

[32] B. B. V. Latoszek, N. Ulozaitė-Stanienė, Y. Maryn, T. Petrauskas, and V. Uloza, "The influence of gender and age on the acoustic voice quality index and dysphonia severity index: a normative study," *Journal of Voice*, vol. 33, no. 3, pp. 340–345, 2019.

[33] P. Caffier, T. Salmen, T. Ermakova et al., "Phonomicrosurgery in vocal fold nodules: quantification of outcomes in professional and non-professional voice users," *Medical Problems of Performing Artists*, vol. 32, no. 4, pp. 187–194, 2017.

[34] M. D. M. Ropero Rendón, T. Ermakova, M.-L. Freymann, A. Ruschin, T. Nawka, and P. P. Caffier, "Efficacy of phonosurgery, logopedic voice treatment and vocal pedagogy in common voice problems of singers," *Advances in Therapy*, vol. 35, no. 7, pp. 1069–1086, 2018.

[35] S. Reetz, J. E. Bohlender, and M. Brockmann-Bauser, "Do standard instrumental acoustic, perceptual, and subjective voice outcomes indicate therapy success in patients with functional dysphonia?," *Journal of Voice*, vol. 33, no. 3, pp. 317–324, 2019.

[36] T. Yılmaz, "Surgical treatment of glottic web using butterfly mucosal flap technique: experience on 12 patients," *The Laryngoscope*, vol. 129, no. 6, pp. 1423–1427, 2019.

[37] J. Ruotsalainen, J. Sellman, L. Lehto, and J. Verbeek, "Systematic review of the treatment of functional dysphonia and prevention of voice disorders," *Otolaryngology—head and Neck Surgery*, vol. 138, no. 138, pp. 557–565, 2008.

[38] E. Sielska-Badurek, E. Osuch-Wójcikiewicz, M. Sobol, E. Kazanecka, A. Rzepakowska, and K. Niemczyk, "Combined functional voice therapy in singers with muscle tension dysphonia in singing," *Journal of Voice*, vol. 31, no. 4, pp. 509.e23–509.e31, 2017.

[39] S. M. Zeitels, R. E. Hillman, M. Mauri, R. Desloge, and P. B. Doyle, "Phonomicrosurgery in singers and performing artists: treatment outcomes, management theories, and future directions," *Annals of Otology, Rhinology & Laryngology*, vol. 111, no. 12, pp. 21–40, 2002.

[40] J. C. Goffi-Fynn and L. M. Carroll, "Collaboration and conquest: MTD as viewed by voice teacher (singing voice specialist) and speech-language pathologist," *Journal of Voice*, vol. 27, no. 3, pp. 391.e9–391.e14, 2013.

[41] K. Pietsch, T. Lyon, and V. K. Dhillon, "Speech language pathology rehabilitation," *Medical Clinics of North America*, vol. 102, no. 6, pp. 1121–1134, 2018.

[42] D. E. Hazlett, O. M. Duffy, and S. A. Moorhead, "Review of the impact of voice training on the vocal quality of professional voice users: implications for vocal health and recommendations for further research," *Journal of Voice*, vol. 25, no. 2, pp. 181–191, 2011.

[43] W. D. LeBorgne and E. N. Donahue, "Voice therapy as primary treatment of vocal fold pathology," *Otolaryngologic Clinics of North America*, vol. 52, no. 4, pp. 649–656, 2019.

[44] M. Sahin, S. Gode, M. Dogan, T. Kirazli, and F. Ogut, "Effect of voice therapy on vocal fold polyp treatment," *European Archives of Oto-Rhino-Laryngology*, vol. 275, no. 6, pp. 1533–1540, 2018.

[45] J. Agarwal, A. Wong, W. Karle, M. Naunheim, M. Mori, and M. Courey, "Comparing short-term outcomes of surgery and voice therapy for patients with vocal fold polyps," *The Laryngoscope*, vol. 129, no. 5, pp. 1067–1070, 2019.

Auszug aus der Journal Summary List: Publikation 3

Journal Data Filtered By: **Selected JCR Year: 2019** Selected Editions: SCIE,SSCI
Selected Categories: **"MEDICINE, GENERAL and INTERNAL"**
Selected Category Scheme: WoS
Gesamtanzahl: 165 Journale

Rank	Full Journal Title	Total Cites	Journal Impact Factor	Eigenfactor Score
1	NEW ENGLAND JOURNAL OF MEDICINE	347,451	74.699	0.660800
2	LANCET	256,199	60.392	0.437300
3	JAMA-JOURNAL OF THE AMERICAN MEDICAL ASSOCIATION	158,632	45.540	0.290050
4	Nature Reviews Disease Primers	7,567	40.689	0.032310
5	BMJ-British Medical Journal	118,586	30.223	0.145170
6	ANNALS OF INTERNAL MEDICINE	58,033	21.317	0.091210
7	JAMA Internal Medicine	17,260	18.652	0.086180
8	PLOS MEDICINE	32,312	10.500	0.065990
9	Journal of Cachexia Sarcopenia and Muscle	3,553	9.802	0.007860
10	Cochrane Database of Systematic Reviews	67,763	7.890	0.134360
11	CANADIAN MEDICAL ASSOCIATION JOURNAL	15,212	7.744	0.016160
12	JOURNAL OF TRAVEL MEDICINE	2,659	7.089	0.006360
13	MAYO CLINIC PROCEEDINGS	15,627	6.942	0.024990
14	JOURNAL OF INTERNAL MEDICINE	10,912	6.871	0.014180
15	BMC Medicine	15,204	6.782	0.042500
16	MEDICAL JOURNAL OF AUSTRALIA	11,075	6.112	0.011070
17	Translational Research	4,043	5.411	0.008350
18	JOURNAL OF THE ROYAL SOCIETY OF MEDICINE	4,214	5.238	0.002580
19	JAMA Network Open	2,239	5.032	0.007660

Rank	Full Journal Title	Total Cites	Journal Impact Factor	Eigenfactor Score
20	Deutsches Arzteblatt International	4,817	4.796	0.007380
21	ANNALS OF FAMILY MEDICINE	5,567	4.686	0.010880
22	JOURNAL OF GENERAL INTERNAL MEDICINE	20,229	4.597	0.026960
23	AMERICAN JOURNAL OF MEDICINE	24,975	4.529	0.024230
24	Journal of Personalized Medicine	617	4.433	0.001950
25	AMERICAN JOURNAL OF PREVENTIVE MEDICINE	23,547	4.420	0.040180
26	European Journal of Internal Medicine	4,933	4.329	0.010280
27	AMYLOID-JOURNAL OF PROTEIN FOLDING DISORDERS	1,486	4.323	0.002920
28	BRITISH JOURNAL OF GENERAL PRACTICE	6,669	4.190	0.008670
29	Frontiers in Medicine	3,034	3.900	0.009870
30	PREVENTIVE MEDICINE	17,316	3.788	0.030080
31	PALLIATIVE MEDICINE	5,413	3.739	0.008460
32	AMERICAN JOURNAL OF CHINESE MEDICINE	3,531	3.682	0.002970
33	MEDICAL CLINICS OF NORTH AMERICA	3,161	3.529	0.004080
34	EUROPEAN JOURNAL OF CLINICAL INVESTIGATION	6,344	3.481	0.006590
35	PANMINERVA MEDICA	806	3.467	0.000660
36	Journal of Clinical Medicine	5,214	3.303	0.010940
37	ANNALS OF MEDICINE	4,510	3.243	0.005190
38	CANADIAN FAMILY PHYSICIAN	3,833	3.112	0.005150

Publikation 3

Song W, **Caffier F**, Nawka T, Ermakova T, Martin A, Mürbe D, Caffier PP.

T1a Glottic Cancer: Advances in Vocal Outcome Assessment after Transoral CO_2-Laser Microsurgery Using the VEM.

J Clin Med 2021; 10(6):1250. https://doi.org/10.3390/jcm10061250

JCR Impact Factor 2020: 4.242 (Q1; 39/168 Medicine, General & Internal)

Journal of
Clinical Medicine

MDPI

Article

T1a Glottic Cancer: Advances in Vocal Outcome Assessment after Transoral CO_2-Laser Microsurgery Using the VEM

Wen Song [1], Felix Caffier [1], Tadeus Nawka [1], Tatiana Ermakova [2], Alexios Martin [3], Dirk Mürbe [1] and Philipp P. Caffier [1,*]

[1] Department of Audiology and Phoniatrics, Charité–Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Campus Charité Mitte, Charitéplatz 1, D-10117 Berlin, Germany; wensonwen@hotmail.com (W.S.); felix.caffier@charite.de (F.C.); tadeus.nawka@charite.de (T.N.); dirk.muerbe@charite.de (D.M.)
[2] Fraunhofer Institute for Open Communication Systems, Kaiserin-Augusta-Allee 31, D-10589 Berlin, Germany; tatiana.ermakova@fokus.fraunhofer.de
[3] Klinikum Mutterhaus der Borromäerinnen, Academic Teaching Hospital of Johannes Gutenberg-Universität Mainz, Feldstraße 16, D-54290 Trier, Germany; alexios.martin@mutterhaus.de
* Correspondence: philipp.caffier@charite.de

Abstract: Patients with unilateral vocal fold cancer (T1a) have a favorable prognosis. In addition to the oncological results of CO_2 transoral laser microsurgery (TOLMS), voice function is among the outcome measures. Previous early glottic cancer studies have reported voice function in patients grouped into combined T stages (Tis, T1, T2) and merged cordectomy types (lesser- vs. larger-extent cordectomies). Some authors have questioned the value of objective vocal parameters. Therefore, the purpose of this exploratory prospective study was to investigate TOLMS-associated oncological and vocal outcomes in 60 T1a patients, applying the ELS protocols for cordectomy classification and voice assessment. Pre- and postoperative voice function analysis included: Vocal Extent Measure (VEM), Dysphonia Severity Index (DSI), auditory-perceptual assessment (GRB), and 9-item Voice Handicap Index (VHI-9i). Altogether, 51 subjects (43 male, eight female, mean age 65 years) completed the study. The 5-year recurrence-free, overall, and disease-specific survival rates (Kaplan–Meier method) were 71.4%, 94.4%, and 100.0%. Voice function was preserved; the objective parameter VEM (64 $\pm$ 33 vs. 83 $\pm$ 31; mean $\pm$ SD) and subjective vocal measures (G: 1.9 $\pm$ 0.7 vs. 1.3 $\pm$ 0.7; VHI-9i: 18 $\pm$ 8 vs. 9 $\pm$ 9) even improved significantly ($p < 0.001$). The VEM best reflected self-perceived voice impairment. It represents a sensitive measure of voice function for quantification of vocal performance.

Keywords: T1a glottic carcinoma; transoral laser microsurgery; treatment outcome; vocal function; objective voice diagnostics; vocal extent measure (VEM)

Citation: Song, W.; Caffier, F.; Nawka, T.; Ermakova, T.; Martin, A.; Mürbe, D.; Caffier, P.P. T1a Glottic Cancer: Advances in Vocal Outcome Assessment after Transoral CO_2-Laser Microsurgery Using the VEM. *J. Clin. Med.* **2021**, *10*, 1250. https://doi.org/10.3390/jcm10061250

Academic Editor: Renee Speyer

Received: 15 February 2021
Accepted: 15 March 2021
Published: 17 March 2021

1. Introduction

Laryngeal cancer is the most frequent malignant tumor in the head and neck area and one of the most common tumors of the respiratory tract [1–3]. GLOBOCAN estimates that more than 177,000 people worldwide developed laryngeal cancer in 2018, with men being affected significantly more often than women (155,000 vs. 22,000) [4]. The prognosis depends mainly on the localization, the TNM classification and the R-status, but also the differentiation and the presence of lymphangiosis carcinomatosa are relevant predictors [5–7]. In the glottis, squamous cell carcinomas are the most frequent type (60 to 80%) compared to other tumor sites within the larynx [8–10]. In early glottic cancer, carcinoma in situ (Tis) must be differentiated from T1 and T2 laryngeal cancer. Invasive T1 glottic cancer is limited to one (T1a) or both (T1b) vocal folds (VF) with normal respiratory but impaired phonatory VF mobility.

T1 and early T2 glottic carcinomas have a very good prognosis due to the early symptom of hoarseness, which usually leads to a quick diagnosis and prompt initiation of

therapy. In addition, metastasis rates are low [11–13]. In the literature, the 5-year overall survival after therapy of early glottic cancer is reported to be in the 74–100% range [14,15]. Involvement of the anterior commissure is more likely to have higher local recurrence, lower laryngeal preservation, but no statistical difference in 5-year overall survival [16,17]. In Steiner's landmark study of 240 patients with laryngeal cancer, early-stage carcinomas had an overall 5-year survival rate of 86.5% (disease-specific 100%), 6% local recurrences, with 99.4% larynx preservation [18]. Ledda and Puxeddu evaluated the oncologic efficacy in 103 patients with early glottic carcinoma, reporting for T1 a 5-year recurrence-free rate of 96% (local control 98%, larynx preservation 100%) [19]. Canis et al. showed in 404 pT1a patients the following 5-year Kaplan-Meier estimates: local control 86.8%, overall survival 87.8%, disease-specific survival 98.0%, recurrence-free survival 76.1%, and larynx preservation 97.3% [20]. Batra et al. presented in 53 patients with Tis and T1 comparable results: local control 86.7%, ultimate local control (with CO_2-laser alone) 90.5%, 3-year overall survival 92.4%, 3-year disease-specific survival and larynx preservation 98.1% [21]. An analysis of 2436 transorally treated T1/T2 carcinomas showed a 5-year overall survival of 82% [22]. For disease-specific survival after T1 and T2 transoral resection, 5-year survival rates of 89–100% are reported in the literature [23]. Meta-analyses on laryngeal preservation after transoral laser resection of T1 and T2 report rates of 83–100% [24].

Early detection of laryngeal cancer can minimize surgical trauma, improve therapeutic outcome and reduce mortality [25]. It is a general consensus that the larynx should be examined laryngoscopically in all patients with hoarseness lasting more than 3 to 4 weeks [26,27]. Videolaryngostroboscopy (VLS) can indicate invasive tissue growth by eliminated mucosal wave propagation and reduced or absent phonatory VF mobility [28,29]. Electronic chromoendoscopy can improve the recognition of tumor margins [30]. A recording of connected speech to document the impaired vocal function is considered a minimum requirement for functional assessment [31]. Small glottal findings suspected of malignancy such as precursor lesions, Tis, and T1a carcinomas, can be completely removed during diagnostic microlaryngoscopy to confirm the diagnosis by excision biopsy [32,33]. Apart from the health status, the quality of life in patients with T1 glottic cancer depends mainly on the voice quality and thus on the extent of the resected VF tissue [34–36]. Surgical therapy is preferred [37,38]; primary radiotherapy, however, can also be used as a conservative VF preserving procedure [39,40].

Transoral CO_2-laser microsurgery (TOLMS) was introduced by Strong and Jako for the therapy of early laryngeal cancer in the 1970s [41], and Steiner gave further impetus in the propagation of this technique [18,42]. Today, TOLMS is established for the treatment of early glottic carcinoma with highly satisfying oncological and functional outcomes (e.g., [20,43,44]). However, many studies predominately focus on oncological results and not on functional outcomes. As the vocal outcome depends on the amount of removed tissue, the consistent classification of endoscopic cordectomies of the European Laryngological Society (ELS) allows interpretation of postoperative results with regard to the surgical strategy and comparison between different surgical centers [45]. The main objective of this exploratory study was to examine in detail the vocal outcome in patients with T1a glottic cancer. The hypothesis was that voice function can be preserved after TOLMS. Therefore, we planned to explore the pre- and postoperative vocal function using specific subjective and objective parameters including the vocal extent measure (VEM) based on the voice range profile (VRP) [46].

2. Materials and Methods

2.1. Study Design and Patients

Patients diagnosed with suspected T1a glottic carcinoma underwent direct microlaryngoscopy in general anaesthesia with TOLMS in a prospective study. Clinical examination and data acquisition took place at the initial pre-therapeutic visit, during operation, and at regular follow-ups postoperatively. The voice was examined the day before TOLMS and 3 months after in-sano resection and completed wound healing. Study participants

were patients presenting with hoarseness at the Department of Audiology and Phoniatrics, Charité–University Medicine Berlin, Germany. Altogether, 60 consecutive patients were recruited between June 2009 and October 2019. Selection criteria comprised histologically confirmed pT1a cN0 cM0 glottic carcinoma, complete treatment documentation, and informed consent. Patients with Tis, T1b and T2 glottic cancer were not included in this investigation.

2.2. Surgical Procedure and Postoperative Regimen

Microlaryngoscopy was conducted via the operating microscope type OPMI Sensera (Zeiss, Jena, Germany) and the Kleinsasser laryngoscope suspension system (Storz, Tuttlingen, Germany). TOLMS was performed with the AcuPulse 30W/40 ST CO_2-laser system (Lumenis, Yokneam, Israel) using the following parameters: output power 2 to 5 watt, super pulse mode, continuous wave, spot size 200 µm, focal length 400 mm. Conventional intraoperative safety precautions were respected (patient covering with moist cloths, safety goggles, laser-resistant endotracheal tube, ventilation with oxygen concentration below 40%). After inspection and palpation under the microscope, saline containing epinephrine (1 mg/mL; 10 gtt. in 10 mL NaCl) was injected into the VF. As a result, stretching the epithelium allowed to assess the fixation of the lesion to deeper structures. The saline also protected the healthy surrounding VF tissue from thermal damage. Laser incisions were made at the site where the suspicious lesions could be distinguished from normal epithelium, considering a safety margin of at least 1 mm. Depending on the pre- and intraoperative findings, cordectomy was conducted. After having removed the suspicious cancerous tissue, the surgeon classified the resection type according to the cordectomy types of the ELS [45]. Lesions within the epithelial level without fixation or signs of infiltration were superficially removed en bloc. Marginal resections were taken if the complete tumor removal was uncertain. All excision biopsies were sent for histopathological examination. The guidelines of the American Joint Committee on Cancer (AJCC) were used for tumor staging [47]. Patients with histopathologically confirmed R1 status were rescheduled for follow-up resection. All TOLMS operations were performed by 5 experienced laryngologists. After surgery, patients were monitored on the ward for 1–2 nights. Before discharge, all treated patients received vocal hygiene counseling. In the event of recurring voice impairment, they were asked to present again between regular follow-up intervals. Postoperative voice rest was not recommended.

2.3. Examination Instruments and Criteria

The analysis of treatment outcome was based on postoperative histopathological findings, pre- and postoperative VLS, and voice function diagnostics. Digital 2D or 3D VLS was carried out via rigid transoral or flexible transnasal endoscopes with integrated microphones (XION GmbH, Berlin, Germany) [28,48]. According to the ELS protocol, voice function diagnostics consisted of established subjective (i.e., auditory-perceptual assessment, self-evaluation of voice) and objective procedures (i.e., VRP measurement, acoustic-aerodynamic analysis) [49–51]. Objective procedures quantify the investigated aspects of vocal function in an apparatus-based and neutral manner. Subjective tests describe the individual self-perceived vocal impairment from the examined person's point of view as well as auditory-perceptual assessments from the examiner's viewpoint.

Auditory-perceptual assessment of the recorded voice samples was conducted using the GRB system [31]. The perceived overall grade of hoarseness (G), roughness (R), and breathiness (B) were independently rated on a scale from 0 to 3 (0 = not existing, 1 = mild, 2 = moderate, 3 = severe) by two senior phoniatricians. From each audio recording the mean score of both GRB evaluations served for further analysis.

Subjective self-assessment of voice was obtained using the 9-item Voice Handicap Index (VHI-9i) including 9 questions rated on a scale from 0 to 4 (0 = never, 1 = almost never, 2 = sometimes, 3 = almost always, 4 = always) [52]. The VHI-9i reflects the functional, physical and emotional impact of the voice disorder on the patient's quality of life. Addi-

tionally, an estimation of the self-perceived overall vocal impairment (VHIs) at the time of questioning was scored between 0 and 3 (0 = normal, 1 = mild, 2 = moderate, 3 = severe).

VRP measurements and acoustic-aerodynamic analyses were performed with the DiVAS software (XION GmbH) to obtain objective quantitative data of the speaking and singing voice. The following parameters were collected: soft phonation threshold, highest and lowest pitch, maximum phonation time (MPT), jitter, dysphonia severity index (DSI) [53], and VEM [46]. The VEM is the logarithmised product of the area of the VRP (A_{VRP}) and the quotient of the circumference of a circle with the same area and the actual VRP circumference (P_{VRP}), supplemented by the addition of a coefficient (50) and an offset (−200). The mathematical formula is:

$$VEM = 50 \ln \left(A_{VRP} \frac{2\pi \sqrt{\frac{A_{VRP}}{\pi}}}{P_{VRP}} \right) - 200 \quad (1)$$

The VEM quantifies the patient's dynamic performance and the frequency range as documented in the VRP. It expresses the vocal capacity as an interval-scaled value, mostly between 0 and 120. A high vocal capacity is characterized by a high VEM; conversely, a small VRP results in a small VEM.

3. Data Analysis

Descriptive statistics were used to describe the quantitative features of all pre- and postoperative parameters and their changes. As graphical techniques to display the data, we chose histograms and violin plots, i.e., box plots with kernel density plots rotated and surrounding them on each side. Being suitable for both continuous and ordinal variables, Spearman's rank-order correlation (r_s) was used to investigate the strength and direction of association between the pre- und postoperatively measured characteristics and their differences. Wilcoxon signed-rank test was used to test whether vocal function parameters significantly improved as the result of TOLMS. Mean values and 95% confidence intervals for these changes were calculated. The impact of patient-related, tumor-related, and treatment-related factors on disease control and survival was analyzed using the Kaplan–Meier method. All statistical tests and graphics were done using R version 4.0.1 (GNU project, Free Software Foundation, Boston, MA, USA). The level of significance was set at $\alpha = 0.05$. Due to the exploratory nature of the study no adjustment for multiple testing was performed. To show different significance levels, the following abbreviations were used: * = 5%; ** = 1%; *** = 0.1%.

4. Results

4.1. Sample Description and Preoperative Assessment

From 60 patients initially recruited with histopathologically confirmed diagnosis of pT1a, six subjects (10.0%) were lost to follow-up and three subjects (5.0%) had to be excluded due to incomplete treatment documentation. In the remaining 51 patients, all diagnostic tests and therapeutic procedures were carried out as planned. The total sample consisted of 43 men and 8 woman, with a mean age of 65 years (range 31–84). At the time of intervention, women were on average 16 years younger than men (52 ± 14 vs. 68 ± 10, mean ± SD, $p < 0.01$). Regarding medical history, 39 subjects (76.5%) gave information about current or past tobacco abuse, with 12 subjects (23.5%) having smoked rarely or not at all. While 15.7% of the patients (8/51) never drank alcohol, 62.7% (32/51) reported regular and 21.6% (11/51) daily consumption of alcohol. Relevant preoperative patient characteristics within the examined cohort are shown in Table 1 (left side).

VLS revealed an almost equal distribution of tumor growth on both VF (28 right, 23 left). The lesions appeared flat and hyperkeratotic in 20/51 (39.2%), exophytic in 29/51 (56.9%), and ulcerating in 2/51 (3.9%) subjects. Concerning macroscopic assessment of tumor size at initial presentation, 51.0% of the patients (26/51) showed involvement of the entire VF, while in 27.4% (14/51) two-thirds and in 21.6% (11/51) one-third of the VF were

affected. During phonation, phonatory VF mobility was reduced or absent on the affected tumor side in all subjects. Additionally, patients with bulged VF due to exophytic tissue growth displayed highly impaired glottal closure.

Subjective auditory-perceptual evaluation of patient's voices was categorized preoperatively with a mean of G2 R2 B1 (range 0–3). The VHI-9i had an average score of 18 ± 8, corresponding to moderate self-assessed patient complaints. The objective acoustic and aerodynamic parameters also indicated moderate impairment (e.g., VEM 64 ± 33; DSI 1.2 ± 2.4; MPT 13 ± 6 s). Correlation analysis performed on preoperative values showed that both VEM and DSI correlated with VHI-9i ($r_s = -0.62$*** and $r_s = -0.29$*, respectively), G ($r_s = -0.42$** and $r_s = -0.34$*), R ($r_s = -0.41$** and $r_s = -0.37$**), B ($r_s = -0.47$*** and $r_s = -0.30$*), and with each other ($r_s = 0.51$***).

Table 1. Patient characteristics ($n = 51$) before TOLMS (left) and after TOLMS (right). Unless otherwise specified, data expressed as number of patients and percentage of group.

	Number	%		Number	%
Gender			Initial cordectomy (via TOLMS)		
male	43	84.3%	type I (subepithelial)	24	47.1%
female	8	15.7%	type II (subligamental)	18	35.3%
			type III (transmuscular)	9	17.6%
Age (in years; mean ± SD)	65 ± 12	-	Grading of pT1a		
			G1 (well differentiated)	15	29.4%
			G2 (moderately differentiated)	34	66.7%
			G3 (poorly differentiated)	2	3.9%
Occurrence of pT1a			Follow-up (in months; mean ± SD)	45 ± 26	-
left vocal fold	23	45.1%			
right vocal fold	28	54.9%			
Vocal fold involvement			Treatment response		
anterior third	3	5.9%	local disease control	41	80.4%
middle third	7	13.7%	local disease recurrence	10	19.6%
posterior third	1	2.0%	contralateral secondary pT1a	2	3.9%
anterior and middle third	7	13.7%	ultimate local disease control	49	96.1%
middle and posterior third	7	13.7%	with TOLMS alone)		
entire length	26	51.0%	larynx preservation	50	98.0%
Appearance of pT1a			Survival		
hyperkeratotic	20	39.2%	disease-specific	51	100.0%
exophytic	29	56.9%	overall	49	96.1%
ulcerating	2	3.9%	recurrence-free	39	76.5%

4.2. Postoperative Assessment

Via TOLMS, 24 patients received subepithelial cordectomy (type I; 47.1%), 18 patients subligamental cordectomy (type II; 35.3%), and nine patients transmuscular cordectomy (type III; 17.6%). According to histopathology, the diagnosis confirmed in all subjects squamous cell carcinoma limited to one VF (pT1a). The grading classification revealed in most patients moderately differentiated tissue (G2; 66.7%), less frequent well differentiated (G1; 29.4%) and seldom poorly differentiated tissue (G3; 3.9%). Through primary operation, the pT1a was completely excised (R0 status) in 29 patients (56.9%). Following the piecemeal strategy, a second excision was necessary in 22 subjects (43.1%), as a residuum could not be ruled out (close tumor margin vs. R1 status). Of these 22 subjects with suspicious findings, 17 patients (77.3%) had no visual or histopathological malignant residue in the scheduled control TOLMS. Among the remaining five patients, the follow-up resections revealed residual invasive tumor in three patients (13.7%), Tis in one patient (4.5%), and a precursor lesion (squamous intraepithelial neoplasia SIN III) in the other patient (4.5%). All these lesions were completely excised during the second TOLMS.

The operative procedures were conducted without complications. Postoperatively, no patient complained about swallowing dysfunction. VLS check-ups showed fibrin formation on the wound surfaces followed by formation of scar tissue during healing. While extensive tumor growth was associated with larger glottal defects after removal, in smaller superficial findings treated via type I cordectomy a stable epithelium regenerated on the preserved lamina propria without relevant defects or scarring. In some patients, the scarred VF developed after about 6 months a restored phonatory mobility. Figure 1 gives an impression of pre- and postoperative VLS findings with videostrobokymographic illustration of VF oscillations.

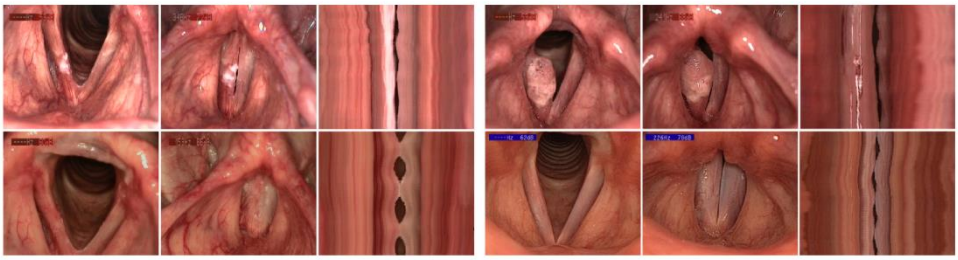

Figure 1. Videolaryngostroboscopic pictures and videostrobokymographic illustration of vocal fold anatomy and function, preoperative (**upper row**) vs. postoperative (**lower row**). Example A (**left side**): 45-year-old male professional theater actor with a flat hyperkeratotic lesion of the right vocal fold. Example B (**right side**): 32-year-old female medical doctor with an exophytic tumor of the right vocal fold. Findings three months postoperatively show: pT1a completely removed, healing process finished, vocal folds with straight margin, complete glottal closure, and restored phonatory mobility (A: normalized, regular and symmetric oscillations; B: oscillations with scarring-related reduced amplitude and phase shift).

Within the mean postoperative observation period of 45 ± 26 months (median: 41 months), 10 patients (19.6%) suffered from a local recurrence (1× Tis, 7× rpT1a, 1× rpT1b, 1× cT3) with an average tumor-free interval of 15 months (median 10 months). Eight of these subjects had only one recurrence within the follow-up period. Among the remaining two, further recurrences occurred: one patient with the initial diagnosis of pT1a (G3) suffered from two recurrences of rpT1a after 17 and 80 months. The other subject with the initial diagnosis of pT1a (G2) had altogether four recurrences; after 13 (rpT1a), 27 (rpT2), 44 (rT3), and 92 months (rpT4a). During follow-up, a secondary glottic pT1a on the contralateral VF was detected in two patients after an interval of 1 and 3 years after removal of the primary tumor, respectively. All recurrent and secondary laryngeal carcinomas were successfully treated: Tis, T1 and T2 via secondary TOLMS, both T3 recurrences via radio-chemotherapy, and the T4 recurrence via total laryngectomy. One subject died due to a secondary pancreas carcinoma, another one died intercurrently. The 5-year recurrence-free, overall, and disease-specific survival rates (Kaplan–Meier method) were 71.4%, 94.4%, and 100.0% (Figure 2). Relevant postoperative and oncological patient characteristics are shown in Table 1 (right side).

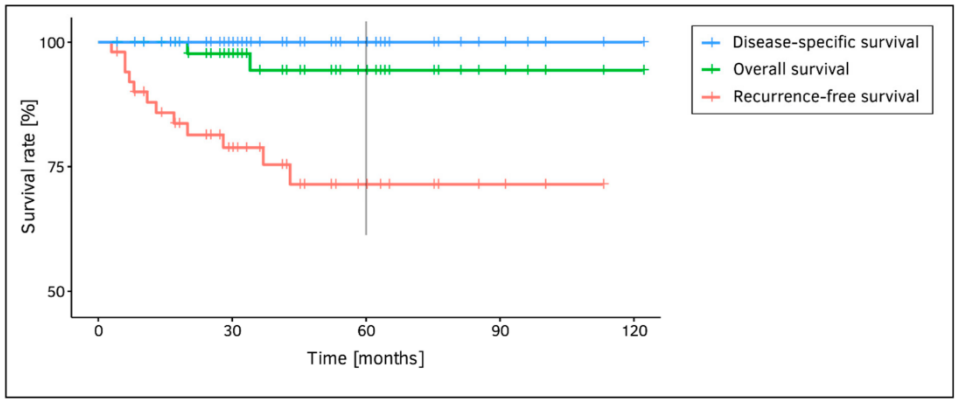

Figure 2. Five-year Kaplan–Meier estimates for recurrence-free survival, overall survival, and disease-specific survival.

Three months after TOLMS, vocal function improved considerably compared to the preoperative measurements (Table 2). With respect to auditory-perceptual GRB evaluation, the pre- vs. post-therapeutical comparison revealed that the voices were less hoarse (1.9 ± 0.7 vs. 1.3 ± 0.7), rough (1.8 ± 0.7 vs. 1.2 ± 0.7), and breathy (1.0 ± 0.6 vs. 0.6 ± 0.6). The subjective vocal self-assessment via VHI-9i questionnaire demonstrated a mean reduction from 18 ± 8 to 9 ± 9 points. The VHIs criterion indicated a change from moderately (2 ± 1) to mildly disturbed voices (1 ± 1). The improvements regarding all these subjective parameters were found significant at the 0.1% level ($p < 0.001$). The subjective vocal parameters both pre- and postoperatively are displayed by histograms in Figure 3.

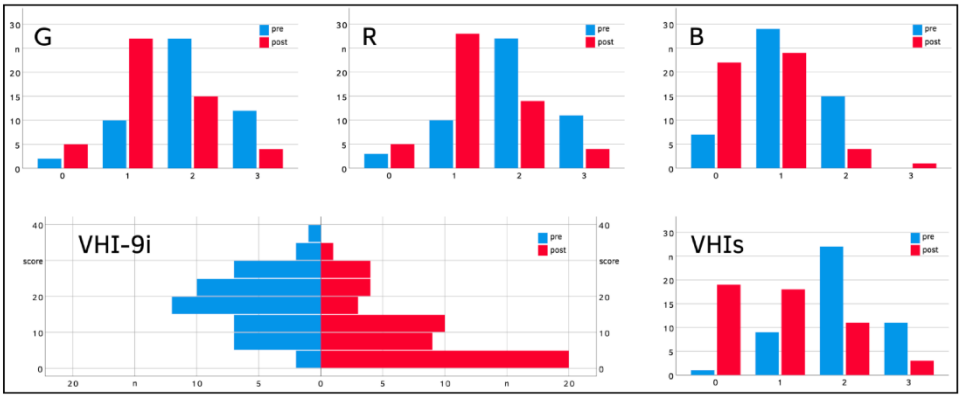

Figure 3. Subjective vocal parameters before and after pT1a removal. Upper row: Comparison of pre- and postoperative voice parameters according to the GRB-classification. Lower row: Comparison of pre- and postoperative VHI-9i and VHIs scores.

Table 2. Pre- and posttherapeutic parameters of vocal function in all patients and all cordectomy types (mean ± SD), their mean therapeutic differences (Diff) and 95% confidence intervals (CI) for changes in vocal measures three months after pT1a removal.

Vocal Measure		Total Group (n = 51)	Type I Cordectomy (n = 24)	Type II Cordectomy (n = 18)	Type III Cordectomy (n = 9)
VEM	Pre	64.4 ± 32.7	65.4 ± 36.9	70.3 ± 31.7	51.0 ± 18.4
	Post	82.8 ± 30.5	86.7 ± 33.5	81.9 ± 25.4	74.1 ± 33.2
	Diff (CI)	18.4 (9.0; 29.8) ***	21.3 (5.1; 37.6) *	11.6 (−3.2; 32.6) *	23.1 (−5.7; 52.0) *
DSI	Pre	1.2 ± 2.4	1.5 ± 2.4	1.4 ± 2.3	−0.2 ± 2.6
	Post	1.5 ± 2.3	1.8 ± 2.6	1.0 ± 2.1	1.8 ± 1.8
	Diff (CI)	0.3 (−0.2; 1.3)	0.3 (−0.5; 1.9)	−0.4 (−1.4; 0.6)	2.0 (0.1; 3.9) *
Jitter (%)	Pre	0.9 ± 1.1	0.8 ± 1.1	0.7 ± 0.9	1.5 ± 1.6
	Post	0.6 ± 0.4	0.6 ± 0.3	0.6 ± 0.5	0.5 ± 0.3
	Diff (CI)	−0.3 (−0.7; −0.02)	−0.2 (−0.7; 0.2)	−0.1 (−0.7; 0.3)	−1.0 (−2.0; 0.1) *
MPT (s)	Pre	13.3 ± 5.6	14.1 ± 5.2	12.3 ± 6.6	13.3 ± 4.5
	Post	13.3 ± 6.0	14.7 ± 6.3	10.9 ± 5.7	14.6 ± 4.5
	Diff (CI)	−0.01 (−1.9; 1.9)	0.6 (−2.4; 3.6)	−1.4 (−4.6; 1.7)	1.3 (−3.6; 6.0)
VHI−9i	Pre	17.7 ± 8.1	16.6 ± 8.3	17.1 ± 7.1	22.1 ± 9.1
	Post	9.3 ± 8.8	10.5 ± 9.0	7.7 ± 8.7	9.2 ± 8.8
	Diff (CI)	−8.4 (−10.9; −5.6) ***	−6.1 (−10.5; −2.1) **	−9.4 (−13.1; −4.9) **	−12.9 (−20.4; −4.3) *
VHIs	Pre	2.0 ± 0.7	1.9 ± 0.9	1.9 ± 0.6	2.4 ± 0.5
	Post	1.0 ± 0.9	1.0 ± 1.0	0.8 ± 0.9	1.0 ± 0.9
	Diff (CI)	−1.0 (−1.4; −0.8) ***	−0.9 (−1.3; −0.6) ***	−1.1 (−1.7; −0.7) ***	−1.4 (−2.2; −0.6) *
G	Pre	1.9 ± 0.7	1.5 ± 0.8	2.2 ± 0.4	2.2 ± 0.7
	Post	1.3 ± 0.7	1.0 ± 0.8	1.5 ± 0.6	1.4 ± 0.6
	Diff (CI)	−0.6 (−0.8; −0.4) ***	−0.5 (−0.8; −0.2) **	−0.7 (−0.9; −0.4) **	−0.8 (−1.2; −0.2) *
R	Pre	1.8 ± 0.7	1.5 ± 0.8	2.1 ± 0.5	2.0 ± 0.8
	Post	1.2 ± 0.7	1.0 ± 0.8	1.5 ± 0.6	1.3 ± 0.6
	Diff (CI)	−0.6 (−0.8; −0.4) ***	−0.5 (−0.8; −0.2) **	−0.6 (−0.9; −0.3) **	−0.7 (−1.2; −0.1) *
B	Pre	1.0 ± 0.6	0.8 ± 0.7	1.2 ± 0.4	1.4 ± 0.4
	Post	0.6 ± 0.6	0.4 ± 0.6	0.9 ± 0.5	0.9 ± 0.7
	Diff (CI)	−0.4 (−0.6; −0.2) ***	−0.4 (−0.7; −0.1) **	−0.3 (−0.6; −0.1) **	−0.5 (−1.1; 0.1) *

B: breathiness; DSI: dysphonia severity index; G: (overall) grade of hoarseness; MPT: maximum phonation time; R: roughness; VEM: vocal extent measure; VHI-9i: 9-item voice handicap index, VHIs: self-perceived overall vocal impairment. The level of significance is indicated as follows: * significant at $p < 0.05$; ** significant at $p < 0.01$; *** significant at $p < 0.001$ (Wilcoxon signed-rank test).

Regarding objective measures, the VEM improved significantly in the total cohort (from 64 ± 33 to 83 ± 31; $p < 0.001$), in both genders (males $p < 0.01$; females $p < 0.05$) and all cordectomy types ($p < 0.05$). In contrast, the decrease of jitter (0.9 ± 1.1 to 0.6 ± 0.4) and the increase of DSI (1.2 ± 2.4 to 1.5 ± 2.3) did not reach the level of significance in the total group, only in females ($p < 0.05$) and cordectomy type III ($p < 0.05$). VEM and DSI correlated significantly with each other also postoperatively ($r_s = 0.62$***). The VEM showed a significant negative correlation with VHI-9i ($r_s = -0.29$*) but not with age ($r_s = -0.18$), while the DSI correlated significantly with age ($r_s = -0.39$**) but not with VHI−9i ($r_s = -0.11$). Selected objective parameters before and after pT1a removal are graphically displayed via boxplots in Figure 4 with regard to the total cohort and cordectomy type.

To provide insights into the magnitude of changes induced by TOLMS, Table 2 also presents the mean differences (and 95% confidence intervals) between pre- and post-therapeutic values. As a result, the numeric outcome of all subjective and objective parameters was larger in women compared to men. Similarly, the improvement of these parameters in cordectomy type III was higher compared to the other cordectomy types.

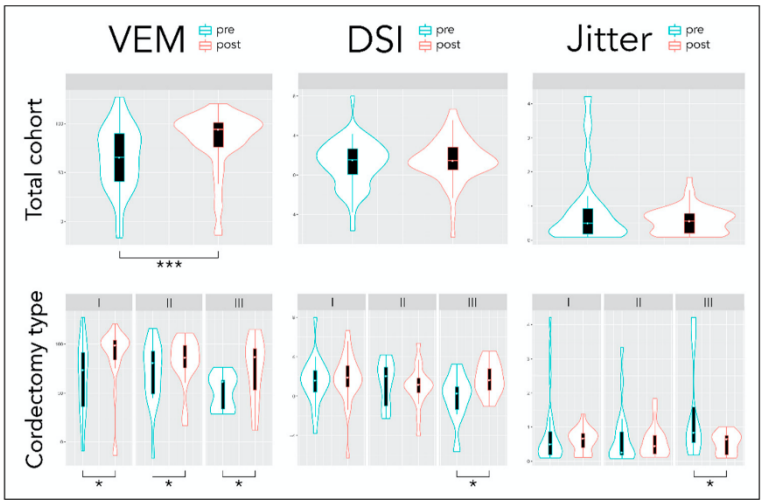

Figure 4. Objective acoustic parameters VEM, DSI, and jitter before and after pT1a removal concerning the total cohort and cordectomy types. Data are compared pre- vs. postoperatively via violin plots, i.e., box plots with kernel density plots rotated and surrounding them on each side. The boxplots display the median, quartiles, and the range of values covered by the data. The density curves display the full distribution of the data including any outliers. The level of significance is indicated as follows: * significant at $p < 0.05$; ** significant at $p < 0.01$; *** significant at $p < 0.001$ (Wilcoxon signed-rank test).

5. Discussion

Given the established favorable oncological results of CO_2-TOLMS in T1a glottic carcinoma, functional aspects should be another treatment objective. We successfully examined the oncological and functional outcomes after TOLMS in pT1a patients, focusing on the evaluation of voice with subjective and objective parameters. Our T1a cohort is consistent with the literature in terms of patient characteristics, treatment methods, and oncological results (see Table 1, Figure 2). Therefore, a closer look at our vocal outcomes is warranted compared to the results of previous investigations.

Many studies were conducted to compare TOLMS with radiotherapy in patients with early glottic cancer [54–56]. The vocal outcomes were either superior in radiotherapy [57,58] or in TOLMS [59,60], or they did not show relevant differences between both treatment groups [61–64]. In general, pre-therapeutic voice data was often not collected [57–59,61,63–69]. In these investigations, it is impossible to relate the postoperative voice function to the pretherapeutic baseline. Some studies evaluated vocal function before and after TOLMS according to the cordectomy type [70–74]. Mainly, voice quality differed depending on the amount of tissue resected: vocal outcomes after lesser-extent cordectomies (ELS type I, II) were superior compared to larger-extent cordectomies. However, a multidimensional, detailed pre- and post-therapeutic documentation and evaluation of voice was only carried out in a few studies [62,70,71,74,75]. To compare the vocal outcomes after TOLMS, Table 3 summarizes the main results of previous investigations including the number of T1a patients treated and the parameters used for evaluation.

Table 3. Published vocal outcomes for T1a glottic cancer treated with TOLMS, taken from representative studies (last 14 years, $n > 10$ T1a patients operated via TOLMS).

Study	Numbers	Parameters for Evaluation of Vocal Function			Vocal Outcome after Transoral Lasermicrosurgery (TOLMS)
		Clinician-Rated Assessment (Subjective)	Patient's Self-Assessment (Subjective)	Acoustic-Aerodynamic Evaluation (Objective)	
Hamzany et al. (2021) [70]	27 T1a	GRB	VHI	F0, jitter, shimmer, NHR, MPT	significant subjective improvement, no objective improvement
Strieth et al. (2019) [76]	14 T1a	–	VHI	–	improved voice preservation by KTP-TOLMS (lower VHI scores) compared to CO_2-TOLMS (higher VHI scores)
Gandhi et al. (2018) [59]	40 T1a + b (N/S)	GRBAS	VHI	F0, jitter, shimmer, SPI, NHR	excellent vocal outcome (G 0.63, VHI 13); no pretherapeutic data
Hong et al. (2018) [61]	14 T1a + b (N/S)	GRBAS	–	F0, jitter, shimmer, NHR	GRB with mild dysphonia, Jitter 2.37%; no pretherapeutic data
Lee et al. (2016) [71]	50 T1a	GRBAS	VHI	F0, jitter, shimmer, NHR, voice intensity, MPT	G significantly improved; voice quality improved over time in limited ELS resections (I-II) but not in extended cordectomies (III-V)
Fink et al. (2016) [72]	38 T1a	VAS (0–100)	VHI	–	similar or improved voice in limited ELS resections (I-III), VHI improved significantly (VAS n.s.); poorer outcomes in extended resections
Kono et al. (2016) [62]	64 T1a	GRBAS	VHI, V-RQOL	F0, jitter, shimmer, NHR, MPT	mild to moderate impairment (GRB, VHI, jitter), better improvement over time in focused excision compared to defocused vaporization
Berania et al. (2015) [65]	18 T1a	PSS-H&N	VHI-10	–	favorable functional outcomes (40% mild voice handicap, VHI-10 > 11); no pretherapeutic data
Bertino et al. (2015) [66]	135 T1a	degree of dysphonia (acc. Ricci Maccarini)	–	F0, HNR	mild to slight dysphonia in limited ELS resections (I-II), moderate to severe dysphonia in extended resections (III-V); no pretherapeutic data
Laoufi et al. (2014) [57]	44 T1a	–	VHI, EORTC QLQ-HN35	–	VHI score mild to moderate impaired (mean 29); no pretherapeutic data
Friedman et al. (2013) [77]	57 T1a	–	V-RQOL	F0, jitter, shimmer, NHR, max. SPL range, max. F0 range, SPL divided by subglottic pressure	significant improvement of subjective (V-RQOL) and most objective (acoustic, aerodynamic) measures
Tomifuji et al. (2013) [73]	33 T1a	GRBAS	VHI	jitter, shimmer, HNR, MPT, MFR	voice quality differs according to the type of cordectomy; no pretherapeutic data
van Gogh et al. (2012) [60]	67 T1a	–	–	F0, jitter, shimmer, NNE	quick voice outcome recovery apart from F0 (remains higher pitched), no significant long-term voice changes
Bajaj et al. (2011) [67]	14 T1a + b (N/S)	GRBAS	VoiSS, UW-QoL	F0, F0 irregularity, CQ range, CQ irregularity	preservation of acceptable vocal function (GRB mild to moderate impaired, low VoiSS score); no pretherapeutic data
Keilmann et al. (2011) [68]	11 T1a	RBH	VHI-12	F0, jitter, shimmer, MPT, GHD, VRP	discrepancy over time (VHI deteriorated; RBH and objective measures improved); no pretherapeutic data

Table 3. *Cont.*

Study	Numbers	Parameters for Evaluation of Vocal Function			Vocal Outcome after Transoral Lasermicrosurgery (TOLMS)
		Clinician-Rated Assessment (Subjective)	Patient's Self-Assessment (Subjective)	Acoustic-Aerodynamic Evaluation (Objective)	
Lester et al. (2011) [78]	19 T1a + b (N/S)	–	ordinal scale (1–5)	F0, jitter, shimmer, MPT	objective acoustic measures showed no significant changes; deterioration of MPT (13s to 12s) and subjective rating score (3 to 2)
Motta et al. (2008) [69]	49 T1a	–	–	MPT HNR, average voice intensity	outcomes vary in relation to the main site of the pseudo-glottis, vocal compensation without normal voice quality; no pretherapeutic data
Núñez Batalla et al. (2008) [63]	19 T1a	GRBAS	VHI	F0, jitter, shimmer, NNE, MPT	mild to moderate impairment (GRBAS, VHI); no pretherapeutic data
Sjögren et al. (2008) [64]	18 T1a	GRBAS	VHI	F0, jitter, shimmer, intensity, MPT, VC, phonation quotient	mild to moderate voice dysfunction (G, B, VHI) in ca. half of patients; no pretherapeutic data
Vilaseca et al. (2008) [79]	35 T1a	GRBAS	ordinal scale (1–3)	F0, jitter, shimmer, NHR, vocal range, MPT	self-assessed improvement; compared with healthy controls: increase of F0, jitter, shimmer (MPT decrease in extended resections); no pretherapeutic data
Roh et al. (2007) [75]	50 T1a	GRBAS	VHI, EORTC QLQ-HN35	F0, jitter, shimmer, HNR, MPT, average airflow	improved vocal outcomes, significant in type I and II cordectomies (VHI, G, jitter, shimmer, HNR)

Legend: CQ—closed quotient, EORTC QLQ-HN35—European Organization for Research and Treatment of Cancer Head and Neck Quality of Life questionnaire; F0—fundamental frequency; GHD—Goettinger Hoarseness Diagram; GRBAS—overall Grade, Roughness, Breathyness, Asthenia, Strain; NHR—harmonics-to-noise ratio; KTP—Potassium titanyl phosphate; MFR—mean flow rate; MPT—maximum phonation time; NHR—noise-to-harmonic ratio; NNE—normalized noise energy; N/S—not specified; PSS-H&N—performance status scale for head & neck cancer patients; RBH—Roughness, Breathyness, (overall grade of) Hoarseness; SNR—signal to-noise ratio; SPI—soft phonation index; SPL—sound pressure level; UW-QoL—University of Washington Quality of Life questionnaire; VAS—visual analogue scale; VC—vital capacity; VHI—voice handicap index; VHI-10—10-item VHI; VHI-12—12-item VHI; VoiSS—voice symptom scale; VRP—voice range profile; V-RQOL—Voice-Related Quality-of-Life survey.

The comparability of published studies is limited due to the lack of standardization regarding (1) vocal outcome assessment (different parameters, follow up), (2) patient selection (e.g., all early glottic cancer patients, low number of T1a), as well as (3) inclusion and treatment criteria (e.g., combined T stages and cordectomy types).

The usefulness of objective acoustic measures has been questioned. Some studies indicated that TOLMS results in an increase of F0, jitter, shimmer, and a moderate decrease of MPT in extended cordectomies when compared with healthy controls (e.g., [79]). Other studies found either a TOLMS-associated improvement [74,75,77], or no relevant changes throughout the postoperative course [70,78]. In our investigation, the patients revealed in all objective and subjective parameters postoperative changes. Similar to the literature, subjective parameters improved significantly [71,72,77,79]: GRB, VHI-9i and VHIs substantially improved in our total cohort, both genders, and in each cordectomy group. Among objective measures, the MPT showed non-specific, undirected changes without any significance. This is in concordance with the results of Hamzany et al., confirming that aerodynamic parameters seem to be less suitable for outcome assessment in T1a glottic carcinoma [70]. Regarding acoustic parameters, VEM seems to be very well suited to assess the resulting voice function after T1a excision compared to other objective acoustic parameters, as only this measure responded significantly in the total cohort and in all subgroups. Among cordectomy types, the larger the resections, the greater the postoperative subjective numerical benefit (Table 2). Similarly, the improvement of acoustic parameters in cordectomy type III was bigger compared to the other cordectomy types. This is related to the fact that larger tumors are associated with more severe voice impairment preoperatively. In contrast, better voice function in smaller tumors results in less postoperative numerical benefit, even if the final voice outcome is better. The relevant differences in the cordectomy groups (types I–III) suggest that pooling these types, as in previous studies of the literature, does not seem appropriate. Although all subjective and objective improvements were larger in women than men, we cannot draw general conclusions due to our limited number of female patients.

While the VEM is not yet widely applied in voice diagnostics, the multidimensional DSI represents an established parameter of instrumental voice evaluation based on a weighted combination of highest possible frequency, lowest intensity, MPT and jitter [53]. Former investigations showed that the DSI might be influenced by using different registration programs, as well as by age or gender [80,81]. These age and gender effects were also confirmed in our study. The DSI appears susceptible to extreme measures (e.g., highest frequency, lowest intensity), which are likely to be influenced by age or gender. In contrast, the VEM, calculated from area and shape of the VRP, is less affected by the above-mentioned extreme measures. Since VEM correlated highly significantly with DSI, both measurements can be seen as related and comparable parameters. Part of their shared variance could be accountable to age, although the linear relationship with age is considerably weaker for the VEM compared to the DSI. However, the VEM as a positive criterion characterizes the vocal abilities and enables a classification of voice performance, while the DSI as a negative criterion particularly describes the severity of dysphonia [80,82]. Among both parameters, the VEM better reflected the subjective vocal impairments. However, DSI, VEM, VHI, and GRB represent different aspects of the voice: They are complementary in objective and subjective evaluation of voice quality, vocal performance, or perceived vocal handicap.

Depending on preoperative T1a tumor characteristics, individual postoperative voice function might be better, similar, or slightly reduced. In general, objective and subjective voice quality improved during long-term postoperative follow-up. This is in line with the results of previous investigations [70,83]. Although voice diagnostics according to ELS protocol is more time-consuming, we consider this effort justified for evidence-based therapy and necessary for documentation of voice preservation. To preserve voice function, the intraoperative laser power should be selected as low as possible to avoid thermal damage in the surrounding healthy tissue. In addition, focused excision achieves better vocal outcomes than defocused vaporization [62]. The application of the KTP laser may be

able to offer improved voice preservation with similar oncological control compared to CO_2-TOLMS [76,77]. The focus on voice preservation may increase the number of interventions in cases with histologically questionable tumor margins [84,85]. Our experience confirms the literature, that re-operation can sometimes be avoided by close monitoring of local control using VLS [44,66].

Study Strengths and Limitations

Our study is characterized by the application of multidimensional voice evaluation, extended by the objective VEM. Further strengths comprise cohort homogeneity restricted to T1a instead of all early glottic cancer patients, and evaluation of specific cordectomy types in a sufficient number of patients rather than generalization or grouping into lesser- vs. larger-extent cordectomies. Applying the ELS protocols both for cordectomy classification and multidimensional voice evaluation enables a systematic comparison of our results with the outcomes of future studies.

Some limitations must be considered before drawing general conclusions. First, our results are investigations of a mono-centre study. To prevent centre bias, multicentre trials with a larger number of subjects are needed. Second, females are underrepresented in our study; thus, there may be participation bias. With a limited number of female patients, general gender-specific conclusions cannot be drawn. Our study sample reflects the well-known prevalence of laryngeal cancer in male patients, though. Third, a more precise preoperative assessment of the exact extent of the pathology would be useful. The importance of tumor size and shape should not be underestimated regarding voice function. The histopathologically determined tumor extent does not replace this information, because resections via TOLMS are not always performed en bloc and may lead to thermal tissue artefacts (e.g., shrinkage, coagulation, vaporization). Fourth, there were differences regarding the individual amount of interventions as well as rehabilitation strategies. Voice therapy could influence the vocal outcome in operated patients. Having neglected this may also result in a performance bias. Lastly, some factors influencing the VRP registration have to be considered. One limitation is the fact that in aphonic patients no perimeter of the VRP can be measured. However, in our study no T1a patient suffered from aphonia. Other factors comprise the routine of the examiner, motivation of the patients, and varying quantities of registered tones. Most of these influential factors are of minor importance in our investigation because all VRPs were recorded by one experienced examiner under practically equal conditions. Since precise VEM calculation is based on the actual VRP shape and circumference, future multicenter studies should be standardized by defining the number of registered tones per interval.

6. Conclusions

TOLMS has been proven to be an established and safe standard oncologic therapy for T1a glottic carcinoma with satisfactory preservation of vocal function both subjectively and objectively. Among objective voice parameters, the VEM seems to best reflect self-perceived subjective voice impairment showing significant changes after T1a treatment that incorporates phonosurgical principles. It represents a sensitive, positive measure of voice function, as well as an understandable and easy-to-use parameter for quantifying vocal performance as documented in the VRP. Therefore, it is reasonable to include the VEM as a diagnostic addition to the established voice measures of the ELS protocol.

Author Contributions: Conceptualization, T.N. and P.P.C.; Methodology, W.S., A.M. and P.P.C.; Literature Review, W.S., T.N. and P.P.C.; Investigation, T.N., A.M. and P.P.C.; Data Analysis, W.S. and T.E.; Original Draft Writing, W.S., F.C. and P.P.C.; Draft Review & Editing, W.S., T.N., D.M. and P.P.C.; Visualization, F.C. and T.E.; Supervision, D.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Ethics Committee of Charité–Universitätsmedizin Berlin, Berlin, Germany (reference number: EA4/140/10).

Informed Consent Statement: Informed consent was obtained from all study participants.

Data Availability Statement: All data of the study are available in the Department of Audiology and Phoniatrics, Charité–Universitätsmedizin Berlin, Berlin, Germany.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Siegel, R.L.; Miller, K.D.; Jemal, A. Cancer Statistics, 2020. *CA Cancer J. Clin.* **2020**, *70*, 7–30. [CrossRef]
2. Nocini, R.; Molteni, G.; Mattiuzzi, C.; Lippi, G. Updates on Larynx Cancer Epidemiology. *Chin. J. Cancer Res.* **2020**, *32*, 18–25. [CrossRef]
3. Steuer, C.E.; El-Deiry, M.; Parks, J.R.; Higgins, K.A.; Saba, N.F. An Update on Larynx Cancer. *CA Cancer J. Clin.* **2017**, *67*, 31–50. [CrossRef]
4. Ferlay, J.; Colombet, M.; Soerjomataram, I.; Mathers, C.; Parkin, D.M.; Pineros, M.; Znaor, A.; Bray, F. Estimating the Global Cancer Incidence and Mortality in 2018: GLOBOCAN Sources and Methods. *Int. J. Cancer.* **2019**, *144*, 1941–1953. [CrossRef]
5. El-Naggar, A.K.; Chan, J.K.C.; Grandis, J.R.; Takata, T.; Slootweg, P.J. *WHO Classification of Head and Neck Tumours*, 4th ed.; WHO: Lyon, France, 2017.
6. Brierley, J.D.; Gospodarowicz, M.K.; Wittekind, C. *TNM Classification of Malignant Tumours*, 8th ed.; Wiley: Chichester, UK, 2016.
7. Agaimy, A.; Weichert, W. Grading of Head and Neck Neoplasms. *Pathologe* **2016**, *37*, 285–292. [CrossRef]
8. Williamson, A.J.; Bondje, S. *Glottic Cancer*; StatPearls Publishing: Treasure Island, FL, USA, 2021.
9. Markou, K.; Christoforidou, A.; Karasmanis, I.; Tsiropoulos, G.; Triaridis, S.; Constantinidis, I.; Vital, V.; Nikolaou, A. Laryngeal Cancer: Epidemiological Data from Nuorthern Greece and Review of the Literature. *Hippokratia* **2013**, *17*, 313–318. [PubMed]
10. Pantel, M.; Guntinas-Lichius, O. Laryngeal Carcinoma: Epidemiology, Risk Factors and Survival. *HNO* **2012**, *60*, 32–40. [CrossRef]
11. Nahavandipour, A.; Jakobsen, K.K.; Gronhoj, C.; Hebbelstrup Jensen, D.; Kim Schmidt Karnov, K.; Klitmoller Agander, T.; Specht, L.; von Buchwald, C. Incidence and Survival of Laryngeal Cancer in Denmark: A Nation-wide Study from 1980 to 2014. *Acta Oncol.* **2019**, *58*, 977–982. [CrossRef] [PubMed]
12. Brandstorp-Boesen, J.; Sorum Falk, R.; Boysen, M.; Brondbo, K. Impact of Stage, Management and Recurrence on Survival Rates in Laryngeal Cancer. *PLoS ONE* **2017**, *12*, e0179371. [CrossRef] [PubMed]
13. Wiegand, S. Evidence-Based Review of Laryngeal Cancer Surgery. *Laryngorhinootologie* **2016**, *95* (Suppl. 1), S192–S216. [CrossRef]
14. Forner, D.; Rigby, M.H.; Corsten, M.; Trites, J.R.; Pyne, J.; Taylor, S.M. Oncological and Functional Outcomes after Repeat Transoral Laser Microsurgery for the Treatment of Recurrent Early Glottic Cancer. *J. Laryngol. Otol.* **2020**, 1–5. [CrossRef]
15. Luscher, M.S.; Pedersen, U.; Johansen, L.V. Treatment Outcome after Laser Excision of Early Glottic Squamous Cell Carcinoma–A Literature Survey. *Acta Oncol.* **2001**, *40*, 796–800. [CrossRef]
16. Zhou, J.; Wen, Q.; Wang, H.; Li, B.; Liu, J.; Hu, J.; Liu, S.; Zou, J. Prognostic Comparison of Transoral Laser Microsurgery for Early Glottic Cancer with or without Anterior Commissure Involvement: A Meta-analysis. *Am. J. Otolaryngol.* **2021**, *42*, 102787. [CrossRef] [PubMed]
17. Hendriksma, M.; Sjogren, E.V. Involvement of the Anterior Commissure in Early Glottic Cancer (Tis-T2): A Review of the Literature. *Cancers* **2019**, *11*, 1234. [CrossRef] [PubMed]
18. Steiner, W. Results of Curative Laser Microsurgery of Laryngeal Carcinomas. *Am. J. Otolaryngol.* **1993**, *14*, 116–121. [CrossRef]
19. Ledda, G.P.; Puxeddu, R. Carbon Dioxide Laser Microsurgery for Early Glottic Carcinoma. *Otolaryngol. Head Neck Surg.* **2006**, *134*, 911–915. [CrossRef]
20. Canis, M.; Ihler, F.; Martin, A.; Matthias, C.; Steiner, W. Transoral Laser Microsurgery for T1a Glottic Cancer: Review of 404 Cases. *Head Neck* **2015**, *37*, 889–895. [CrossRef] [PubMed]
21. Batra, A.; Goyal, A.; Goyal, M.; Goel, S. Oncological Outcomes Following Transoral CO2 Laser Microsurgery for T1 Glottic Cancer. *Indian J. Otolaryngol. Head Neck Surg.* **2019**, *71* (Suppl. 1), 542–547. [CrossRef] [PubMed]
22. Arens, C. Transoral Treatment Strategies for Head and Neck Tumors. *GMS Curr. Top. Otorhinolaryngol. Head Neck Surg.* **2012**, *11*. [CrossRef]
23. Wiegand, S. Evidence and Evidence Gaps of Laryngeal Cancer Surgery. *GMS Curr. Top. Otorhinolaryngol. Head Neck Surg.* **2016**, *15*. [CrossRef]
24. Ambrosch, P.; Fazel, A. Functional Organ Preservation in Laryngeal and Hypopharyngeal Cancer. *GMS Curr. Top. Otorhinolaryngol. Head Neck Surg.* **2011**, *10*. [CrossRef]
25. Chatenoud, L.; Garavello, W.; Pagan, E.; Bertuccio, P.; Gallus, S.; La Vecchia, C.; Negri, E.; Bosetti, C. Laryngeal Cancer Mortality Trends in European Countries. *Int. J. Cancer* **2016**, *138*, 833–842. [CrossRef]
26. Stachler, R.J.; Francis, D.O.; Schwartz, S.R.; Damask, C.C.; Digoy, G.P.; Krouse, H.J.; McCoy, S.J.; Ouellette, D.R.; Patel, R.R.; Reavis, C.C.W.; et al. Clinical Practice Guideline: Hoarseness (Dysphonia) (Update). *Otolaryngol. Head Neck Surg.* **2018**, *158* (Suppl. 1), S1–S42. [CrossRef] [PubMed]

27. Tikka, T.; Pracy, P.; Paleri, V. Refining the Head and Neck Cancer Referral Guidelines: A Two Centre Analysis of 4715 Referrals. *Br. J. Oral Maxillofac. Surg.* **2016**, *54*, 141–150. [CrossRef]
28. Caffier, P.P.; Schmidt, B.; Gross, M.; Karnetzky, K.; Nawka, T.; Rotter, A.; Seipelt, M.; Sedlmaier, B. A Comparison of White Light Laryngostroboscopy versus Autofluorescence Endoscopy in the Evaluation of Vocal Fold Pathology. *Laryngoscope* **2013**, *123*, 1729–1734. [CrossRef] [PubMed]
29. Whited, C.W.; Dailey, S.H. Evaluation of the Dysphonic Patient (in: Function Preservation in Laryngeal Cancer). *Otolaryngol. Clin. N. Am.* **2015**, *48*, 547–564. [CrossRef] [PubMed]
30. Piazza, C.; Cocco, D.; De Benedetto, L.; Del Bon, F.; Nicolai, P.; Peretti, G. Narrow Band Imaging and High Definition Television in the Assessment of Laryngeal Cancer: A Prospective Study on 279 Patients. *Eur. Arch. Otorhinolaryngol.* **2010**, *267*, 409–414. [CrossRef] [PubMed]
31. Ptok, M.; Schwemmle, C.; Iven, C.; Jessen, M.; Nawka, T. On the Auditory Evaluation of Voice Quality. *HNO* **2006**, *54*, 793–802. [CrossRef] [PubMed]
32. Ali, S.A.; Smith, J.D.; Hogikyan, N.D. The White Lesion, Hyperkeratosis, and Dysplasia. *Otolaryngol. Clin N. Am.* **2019**, *52*, 703–712. [CrossRef] [PubMed]
33. Nawka, T.; Martin, A.; Caffier, P.P. Microlaryngoscopy and Phonomicrosurgery. *HNO* **2013**, *61*, 108–116. [CrossRef] [PubMed]
34. Hartl, D.M.; Laoufi, S.; Brasnu, D.F. Voice Outcomes of Transoral Laser Microsurgery of the Larynx. *Otolaryngol. Clin. N. Am.* **2015**, *48*, 627–637. [CrossRef] [PubMed]
35. Mau, T.; Palaparthi, A.; Riede, T.; Titze, I.R. Effect of Resection Depth of Early Glottic Cancer on Vocal Outcome: An Optimized Finite Element Simulation. *Laryngoscope* **2015**, *125*, 1892–1899. [CrossRef] [PubMed]
36. Peeters, A.J.; van Gogh, C.D.; Goor, K.M.; Verdonck-de Leeuw, I.M.; Langendijk, J.A.; Mahieu, H.F. Health Status and Voice Outcome after Treatment for T1a Glottic Carcinoma. *Eur. Arch. Otorhinolaryngol.* **2004**, *261*, 534–540. [CrossRef] [PubMed]
37. Bozec, A.; Culie, D.; Poissonnet, G.; Dassonville, O. Current Role of Primary Surgical Treatment in Patients with Head and Neck Squamous Cell Carcinoma. *Curr. Opin. Oncol.* **2019**, *31*, 138–145. [CrossRef]
38. Hartl, D.M.; Brasnu, D.F. Contemporary Surgical Management of Early Glottic Cancer. *Otolaryngol. Clin. N. Am.* **2015**, *48*, 611–625. [CrossRef] [PubMed]
39. Baird, B.J.; Sung, C.K.; Beadle, B.M.; Divi, V. Treatment of Early-stage Laryngeal Cancer: A Comparison of Treatment Options. *Oral Oncol.* **2018**, *87*, 8–16. [CrossRef]
40. Huang, G.; Luo, M.; Zhang, J.; Liu, H. Laser Surgery versus Radiotherapy for T1a Glottic Carcinoma: A Meta-analysis of Oncologic Outcomes. *Acta Otolaryngol.* **2017**, *137*, 1204–1209. [CrossRef]
41. Strong, M.S.; Jako, G.J. Laser Surgery in the Larynx. Early Clinical Experience with Continuous CO_2 Laser. *Ann. Otol. Rhinol. Laryngol.* **1972**, *81*, 791–798. [CrossRef]
42. Harris, A.T.; Tanyi, A.; Hart, R.D.; Trites, J.; Rigby, M.H.; Lancaster, J.; Nicolaides, A.; Taylor, S.M. Transoral Laser Surgery for Laryngeal Carcinoma: Has Steiner Achieved a Genuine Paradigm Shift in Oncological Surgery? *Ann. R. Coll. Surg. Engl.* **2018**, *100*, 2–5. [CrossRef]
43. Sjogren, E.V. Transoral Laser Microsurgery in Early Glottic Lesions. *Curr. Otorhinolaryngol. Rep.* **2017**, *5*, 56–68. [CrossRef]
44. Peretti, G.; Piazza, C.; Cocco, D.; De Benedetto, L.; Del Bon, F.; Redaelli De Zinis, L.O.; Nicolai, P. Transoral CO(2) Laser Treatment for T(is)-T(3) Glottic Cancer: The University of Brescia Experience on 595 Patients. *Head Neck* **2010**, *32*, 977–983. [CrossRef]
45. Remacle, M.; Van Haverbeke, C.; Eckel, H.; Bradley, P.; Chevalier, D.; Djukic, V.; de Vicentiis, M.; Friedrich, G.; Olofsson, J.; Peretti, G.; et al. Proposal for Revision of the European Laryngological Society Classification of Endoscopic Cordectomies. *Eur. Arch. Otorhinolaryngol.* **2007**, *264*, 499–504. [CrossRef] [PubMed]
46. Caffier, P.P.; Moller, A.; Forbes, E.; Muller, C.; Freymann, M.L.; Nawka, T. The Vocal Extent Measure: Development of a Novel Parameter in Voice Diagnostics and Initial Clinical Experience. *BioMed Res. Int.* **2018**, *2018*, 3836714. [CrossRef]
47. Amin, M.B.; Edge, S.; Greene, F.; Byrd, D.R.; Brookland, R.K.; Washington, M.K.; Gershenwald, J.E.; Compton, C.C.; Hess, K.R.; Sullivan, D.C.; et al. *AJCC Cancer Staging Manual*, 8th ed.; Springer: New York, NY, USA, 2017.
48. Caffier, P.P.; Nawka, T.; Ibrahim-Nasr, A.; Thomas, B.; Muller, H.; Ko, S.R.; Song, W.; Gross, M.; Weikert, S. Development of Three-dimensional Laryngostroboscopy for Office-based Laryngeal Diagnostics and Phonosurgical Therapy. *Laryngoscope* **2018**, *128*, 2823–2831. [CrossRef]
49. Patel, R.R.; Awan, S.N.; Barkmeier-Kraemer, J.; Courey, M.; Deliyski, D.; Eadie, T.; Paul, D.; Švec, J.G.; Hillman, R. Recommended Protocols for Instrumental Assessment of Voice: American Speech-Language-Hearing Association Expert Panel to Develop a Protocol for Instrumental Assessment of Vocal Function. *Am. J. Speech Lang. Pathol.* **2018**, *27*, 887–905. [CrossRef] [PubMed]
50. Dejonckere, P.H.; Bradley, P.; Clemente, P.; Cornut, G.; Crevier-Buchman, L.; Friedrich, G.; Van De Heyning, P.; Remacle, M.; Woisard, V.; Committee on Phoniatrics of the European Laryngological Society (ELS). A Basic Protocol for Functional Assessment of Voice Pathology, Especially for Investigating the Efficacy of (Phonosurgical) Treatments and Evaluating New Assessment Techniques. Guideline Elaborated by the Committee on Phoniatrics of the European Laryngological Society (ELS). *Eur. Arch. Otorhinolaryngol.* **2001**, *258*, 77–82. [CrossRef]
51. Ternström, S.; Pabon, P.; Södersten, M. The Voice Range Profile: Its Function, Applications, Pitfalls and Potential. *Acta Acust. United Acust.* **2016**, *102*, 268–283. [CrossRef]

52. Nawka, T.; Verdonck-de Leeuw, I.M.; De Bodt, M.; Guimaraes, I.; Holmberg, E.B.; Rosen, C.A.; Schindler, A.; Woisard, V.; Whurr, R.; Konerding, U. Item reduction of the voice handicap index based on the original version and on European translations. *Folia Phoniatr. Logop.* **2009**, *61*, 37–48. [CrossRef] [PubMed]
53. Wuyts, F.L.; De Bodt, M.S.; Molenberghs, G.; Remacle, M.; Heylen, L.; Millet, B.; Van Lierde, K.; Raes, J.; Van de Heyning, P.H. The Dysphonia Severity Index: An Objective Measure of Vocal Quality Based on a Multiparameter Approach. *J. Speech Lang. Hear Res.* **2000**, *43*, 796–809. [CrossRef]
54. Greulich, M.T.; Parker, N.P.; Lee, P.; Merati, A.L.; Misono, S. Voice Outcomes Following Radiation versus Laser Microsurgery for T1 Glottic Carcinoma: Systematic Review and Meta-analysis. *Otolaryngol. Head Neck Surg.* **2015**, *152*, 811–819. [CrossRef] [PubMed]
55. Spielmann, P.M.; Majumdar, S.; Morton, R.P. Quality of Life and Functional Outcomes in the Management of Early Glottic Carcinoma: A Systematic Review of Studies Comparing Radiotherapy and Transoral Laser Microsurgery. *Clin. Otolaryngol.* **2010**, *35*, 373–382. [CrossRef]
56. Cohen, S.M.; Garrett, C.G.; Dupont, W.D.; Ossoff, R.H.; Courey, M.S. Voice-related Quality of Life in T1 Glottic Cancer: Irradiation versus Endoscopic Excision. *Ann. Otol. Rhinol. Laryngol.* **2006**, *115*, 581–586. [CrossRef]
57. Laoufi, S.; Mirghani, H.; Janot, F.; Hartl, D.M. Voice Quality after Treatment of T1a Glottic Cancer. *Laryngoscope* **2014**, *124*, 1398–1401. [CrossRef]
58. Krengli, M.; Policarpo, M.; Manfredda, I.; Aluffi, P.; Gambaro, G.; Panella, M.; Pia, F. Voice Quality after Treatment for T1a Glottic Carcinoma—Radiotherapy versus Laser Cordectomy. *Acta Oncol.* **2004**, *43*, 284–289. [CrossRef] [PubMed]
59. Gandhi, S.; Gupta, S.; Rajopadhye, G. A Comparison of Phonatory outcome Between Trans-oral CO_2 Laser Cordectomy and Radiotherapy in T1 Glottic Cancer. *Eur. Arch. Otorhinolaryngol.* **2018**, *275*, 2783–2786. [CrossRef]
60. van Gogh, C.D.; Verdonck-de Leeuw, I.M.; Wedler-Peeters, J.; Langendijk, J.A.; Mahieu, H.F. Prospective Evaluation of Voice Outcome during the First Two Years in Male Patients Treated by Radiotherapy or Laser Surgery for T1a Glottic Carcinoma. *Eur. Arch. Otorhinolaryngol.* **2012**, *269*, 1647–1652. [CrossRef] [PubMed]
61. Hong, Y.T.; Park, M.J.; Hong, K.H. Characteristics of Speech Production in Patients with T1 Glottic Cancer who Underwent Laser Cordectomy or Radiotherapy. *Logoped Phoniatr. Vocol.* **2018**, *43*, 120–128. [CrossRef] [PubMed]
62. Kono, T.; Saito, K.; Yabe, H.; Uno, K.; Ogawa, K. Comparative Multidimensional Assessment of Laryngeal Function and Quality of Life after Radiotherapy and Laser Surgery for Early Glottic Cancer. *Head Neck* **2016**, *38*, 1085–1090. [CrossRef] [PubMed]
63. Nunez Batalla, F.; Caminero Cueva, M.J.; Senaris Gonzalez, B.; Llorente Pendas, J.L.; Gorriz Gil, C.; Lopez Llames, A.; Alonso Pantiga, R.; Suárez Nieto, C. Voice Quality after Endoscopic Laser Surgery and Radiotherapy for Early Glottic Cancer: Objective Measurements Emphasizing the Voice Handicap Index. *Eur. Arch. Otorhinolaryngol.* **2008**, *265*, 543–548. [CrossRef]
64. Sjogren, E.V.; van Rossum, M.A.; Langeveld, T.P.; Voerman, M.S.; van de Kamp, V.A.; Friebel, M.O.; Wolterbeek, R.; Baatenburg de Jong, R.J. Voice Outcome in T1a Midcord Glottic Carcinoma: Laser Surgery vs Radiotherapy. *Arch. Otolaryngol. Head Neck Surg.* **2008**, *134*, 965–972. [CrossRef]
65. Berania, I.; Dagenais, C.; Moubayed, S.P.; Ayad, T.; Olivier, M.-J.; Guertin, L.; Bissada, E.; Tabet, J.C.; Christopoulos, A. Voice and Functional Outcomes of Transoral Laser Microsurgery for Early Glottic Cancer: Ventricular Fold Resection as a Surrogate. *J. Clin. Med. Res.* **2015**, *7*, 632–636. [CrossRef]
66. Bertino, G.; Degiorgi, G.; Tinelli, C.; Cacciola, S.; Occhini, A.; Benazzo, M. CO_2 Laser Cordectomy for T1-T2 Glottic Cancer: Oncological and Functional Long-term Results. *Eur. Arch. Otorhinolaryngol.* **2015**, *272*, 2389–2395. [CrossRef]
67. Bajaj, Y.; Uppal, S.; Sharma, R.K.; Grace, A.R.; Howard, D.M.; Nicolaides, A.R.; Coatesworth, A.P. Evaluation of Voice and Quality of Life after Transoral Endoscopic Laser Resection of Early Glottic Carcinoma. *J. Laryngol. Otol.* **2011**, *125*, 706–713. [CrossRef]
68. Keilmann, A.; Napiontek, U.; Engel, C.; Nakarat, T.; Schneider, A.; Mann, W. Long-term Functional Outcome after Unilateral Cordectomy. *ORL J. Otorhinolaryngol. Relat. Spec.* **2011**, *73*, 38–46. [CrossRef]
69. Motta, S.; Cesari, U.; Mesolella, M.; Motta, G. Functional Vocal Results after CO2 Laser Endoscopic Surgery for Glottic Tumours. *J. Laryngol. Otol.* **2008**, *122*, 948–951. [CrossRef]
70. Hamzany, Y.; Crevier-Buchman, L.; Lechien, J.R.; Bachar, G.; Brasnu, D.; Hans, S. Multidimensional Voice Quality Evaluation After Transoral CO2 Laser Cordectomy: A Prospective Study. *Ear Nose Throat J.* **2021**, *100* (Suppl. 1), 27S–32S. [CrossRef] [PubMed]
71. Lee, H.S.; Kim, J.S.; Kim, S.W.; Noh, W.J.; Kim, Y.J.; Oh, D.; Hong, J.C.; Lee, K.D. Voice Outcome According to Surgical Extent of Transoral Laser Microsurgery for T1 Glottic Carcinoma. *Laryngoscope* **2016**, *126*, 2051–2056. [CrossRef]
72. Fink, D.S.; Sibley, H.; Kunduk, M.; Schexnaildre, M.; Kakade, A.; Sutton, C.; McWhorter, A.J. Subjective and Objective Voice Outcomes After Transoral Laser Microsurgery for Early Glottic Cancer. *Laryngoscope* **2016**, *126*, 405–407. [CrossRef]
73. Tomifuji, M.; Araki, K.; Niwa, K.; Miyagawa, Y.; Mizokami, D.; Kitagawa, Y.; Yamashita, T.; Matsunobu, T.; Shiotani, A. Comparison of Voice Quality After Laser Cordectomy with That After Radiotherapy or Chemoradiotherapy for Early Glottic Carcinoma. *ORL J. Otorhinolaryngol. Relat. Spec.* **2013**, *75*, 18–26. [CrossRef] [PubMed]
74. Peretti, G.; Piazza, C.; Balzanelli, C.; Mensi, M.C.; Rossini, M.; Antonelli, A.R. Preoperative and Postoperative Voice in Tis-T1 Glottic Cancer Treated by Endoscopic Cordectomy: An Additional Issue for Patient Counseling. *Ann. Otol. Rhinol. Laryngol.* **2003**, *112*, 759–763. [CrossRef] [PubMed]
75. Roh, J.L.; Kim, D.H.; Kim, S.Y.; Park, C.I. Quality of Life and Voice in Patients After Laser Cordectomy for Tis and T1 Glottic Carcinomas. *Head Neck* **2007**, *29*, 1010–1016. [CrossRef]

76. Strieth, S.; Ernst, B.P.; Both, I.; Hirth, D.; Pfisterer, L.N.; Kunzel, J.; Eder, K. Randomized Controlled Single-blinded Clinical Trial of Functional Voice Outcome after Vascular Targeting KTP laser Microsurgery of Early Laryngeal Cancer. *Head Neck* **2019**, *41*, 899–907. [CrossRef]
77. Friedman, A.D.; Hillman, R.E.; Landau-Zemer, T.; Burns, J.A.; Zeitels, S.M. Voice Outcomes for Photoangiolytic KTP Laser Treatment of Early Glottic Cancer. *Ann. Otol. Rhinol. Laryngol.* **2013**, *122*, 151–158. [CrossRef] [PubMed]
78. Lester, S.E.; Rigby, M.H.; MacLean, M.; Taylor, S.M. 'How Does That sound?': Objective and Subjective Voice Outcomes Following CO(2) Laser Resection for Early Glottic Cancer. *J. Laryngol. Otol.* **2011**, *125*, 1251–1255. [CrossRef] [PubMed]
79. Vilaseca, I.; Huerta, P.; Blanch, J.L.; Fernandez-Planas, A.M.; Jimenez, C.; Bernal-Sprekelsen, M. Voice Quality after CO2 Laser Cordectomy–What Can We Really Expect? *Head Neck* **2008**, *30*, 43–49. [CrossRef]
80. Aichinger, P.; Feichter, F.; Aichstill, B.; Bigenzahn, W.; Schneider-Stickler, B. Inter-device Reliability of DSI Measurement. *Logoped. Phoniatr. Vocol.* **2012**, *37*, 167–173. [CrossRef] [PubMed]
81. Hakkesteegt, M.M.; Brocaar, M.P.; Wieringa, M.H.; Feenstra, L. Influence of Age and Gender on the Dysphonia Severity Index. A Study of Normative Values. *Folia Phoniatr. Logop.* **2006**, *58*, 264–273. [CrossRef]
82. Hakkesteegt, M.M.; Brocaar, M.P.; Wieringa, M.H. The Applicability of the Dysphonia Severity Index and the Voice Handicap Index in Evaluating Effects of Voice Therapy and Phonosurgery. *J. Voice* **2010**, *24*, 199–205. [CrossRef]
83. Chu, P.-Y.; Hsu, Y.-B.; Lee, T.-L.; Fu, S.; Wang, L.-M.; Kao, Y.-C. Longitudinal Analysis of Voice Quality in Patients with Early Glottic Cancer after Transoral Laser Microsurgery. *Head Neck* **2012**, *34*, 1294–1298. [CrossRef] [PubMed]
84. Burns, J.A.; Har-El, G.; Shapshay, S.; Maune, S.; Zeitels, S.M. Endoscopic Laser Resection of Laryngeal Cancer: Is it Oncologically Safe? Position Statement from the American Broncho-Esophagological Association. *Ann. Otol. Rhinol. Laryngol.* **2009**, *118*, 399–404. [CrossRef] [PubMed]
85. Aluffi Valletti, P.; Taranto, F.; Chiesa, A.; Pia, F.; Valente, G. Impact of Resection Margin Status on Oncological Outcomes after CO2 Laser Cordectomy. *Acta Otorhinolaryngol. Ital.* **2018**, *38*, 24–30. [PubMed]

Lebenslauf

Mein Lebenslauf wird aus datenschutzrechtlichen Gründen in der elektronischen Version meiner Arbeit nicht veröffentlicht.

Publikationsliste

Caffier F, Nawka T, Neumann K, Seipelt M, Caffier PP. *Validation and Classification of the 9-Item Voice Handicap Index (VHI-9i).* J Clin Med 2021 Jul 28;10(15):3325. https://doi.org/10.3390/jcm10153325

Song W, **Caffier F**, Nawka T, Ermakova T, Martin A, Mürbe D, Caffier PP. *T1a Glottic Cancer: Advances in Vocal Outcome Assessment after Transoral CO_2-Laser Microsurgery Using the VEM.* J Clin Med 2021 Mar 17;10(6):1250. https://doi.org/10.3390/jcm10061250

Müller C, **Caffier F**, Nawka T, Müller M, Caffier PP. *Pathology-Related Influences on the VEM: Three Years' Experience since Implementation of a New Parameter in Phoniatric Voice Diagnostics.* Biomed Res Int 2020 Dec 21;2020:5309508. https://doi.org/10.1155/2020/5309508

Seipelt M, Möller A, Nawka T, Gonnermann U, **Caffier F**, Caffier PP. *Monitoring the Outcome of Phonosurgery and Vocal Exercises with Established and New Diagnostic Tools.* Biomed Res Int 2020 Jan 23;2020:4208189. https://doi.org/10.1155/2020/4208189

Caffier F. *Der innere Dirigent. Betrachtungen über das Lernen, Üben, Hören und Musizieren.* Hochschule für Musik “Hanns Eisler” Berlin, Theoretische Arbeit zum Diplom im Ergänzungsstudium Musikpädagogik, 76 Seiten, 2012.

Danksagung

An dieser Stelle möchte ich allen danken, die mich bei der Anfertigung meiner Dissertation in diesen schwierigen Pandemiezeiten unterstützt haben. Besonderer Dank gilt hierbei Prof. Dr. med. Tadeus Nawka, meinem Erstbetreuer, für die Überlassung des Promotionsthemas. Gemeinsam mit Prof. Dr. med. Philipp Caffier haben beide durch ihre fachliche Unterstützung und nützlichen Hinweise die Arbeit maßgeblich beeinflusst und verbessert. Außerdem war ihre Motivation in den zahllosen Nächten, in denen Quellen recherchiert, Statistiken berechnet und an Formulierungen gefeilt wurde, entscheidend für die zeitgerechte Fertigstellung der Arbeit.

Weiterhin danke ich dem Personal der Klinik für Audiologie und Phoniatrie der Charité – Universitätsmedizin Berlin für die akribische Behandlungsdokumentation, die diese kumulative Dissertation erst ermöglicht hat. Für die statistische Analyse waren die vielen tausend Datensätze, die über mehr als ein Jahrzehnt gesammelt wurden, eine wahre Schatztruhe.

Meiner Familie und Freunden danke ich für ihre Geduld über die letzten Jahre und ihr Vertrauen in mich auf meinem nicht immer geradlinigen Weg.