

Analyse und Optimierung der Inspektionsreliabilität sublingualer Varizen

Inaugural-Dissertation

zur Erlangung des Doktorgrades

der Hohen Medizinischen Fakultät

der Rheinischen Friedrich-Wilhelms-Universität

Bonn

Christian Reinhard Klein

aus Reutlingen

2024

Angefertigt mit der Genehmigung
der Medizinischen Fakultät der Universität Bonn

1. Gutachter: Prof. Dr. Helmut Stark
2. Gutachterin: PD Dr. Svenja Beisel-Memmert

Tag der Mündlichen Prüfung: 06.05.2024

Aus der Klinik und Poliklinik für Zahnärztliche Prothetik, Propädeutik und
Werkstoffwissenschaften
Direktor: Prof. Dr. Helmut Stark

Inhaltsverzeichnis

	Abkürzungsverzeichnis	4
1.	Deutsche Zusammenfassung	5
1.1	Einleitung	5
1.2	Material und Methoden	6
1.3	Ergebnisse	10
1.4	Diskussion	12
1.5	Zusammenfassung	14
1.6	Literaturverzeichnis der deutschen Zusammenfassung	15
2.	Veröffentlichung	20
	Abstract	20
	Introduction	21
	Materials and Methods	21
	Results	23
	Discussion	25
	References	27
3.	Danksagung	28

Abkürzungsverzeichnis

A	Fläche der sichtbaren sublingualen Venen (px ²)
AUC	Area Under the Curve
GIMP	GNU Image Manipulation Program
L	Zungenlänge (px)
Px	Pixel
r	Korrelationskoeffizient (Bravais-Pearson-Korrelation)
R	Reliabilität R
RA-Score	Relative Area-Score
ROC	Receiver-Operating-Characteristics
SV	Sublinguale Varizen
SV _P	SV-Prävalenz

1. Deutsche Zusammenfassung

1.1 Einleitung

Sublinguale Varizen (SV) sind ein viel untersuchtes Objekt der Oralmedizin. Als sublingualen Varizen bezeichnet man dilatative Erweiterungen der sublingualen Venen (*Vv. linguae profundae*). Ein molekularer Pathomechanismus ihrer Entstehung ist bisher nicht bekannt, allerdings wurde in vorangehenden Arbeiten eine Reihe klinischer Parameter mit dem Vorliegen oder der Entwicklung von SV assoziiert: Bisher untersuchte Parameter sind Alter (Bergh, Kastberg, u. a. 2022; Lazos u. a. 2020; Accardo u. a. 2020; Ettinger und Manderson 1974), arterielle Hypertonie (Hedström, Albrektsson, und Bergh 2015; Akkaya, Ölmez, und Özkan 2019; Lazos u. a. 2020; Hedström und Bergh 2010; Accardo u. a. 2021), Varikosis der unteren Extremität (Ettinger und Manderson 1974), Rauchen (Hedström und Bergh 2010; Al-Shayyab und Baqain 2015; Barzideh, Alaei, und Azizi 2021), Diabetes mellitus Typ 2 (Bergh, Albrektsson, u. a. 2022; Accardo u. a. 2021), Dyslipidämie (Bergh, Albrektsson, u. a. 2022; Accardo u. a. 2021), Leberzirrhose (Duarte u. a. 2020; Gonzalez Ballerga u. a. 2018) und Allergien (González-Álvarez und García-Pola 2022). Zudem bestimmten mehrere Studien eine SV-Prävalenz in verschiedenen Populationen (Triantos 2005; Rabiei u. a. 2010; Corrêa u. a. 2007; do Egito Vasconcelos u. a. 2008; Espinoza u. a. 2003; Mozafari u. a. 2012; Bhatnagar u. a. 2013; Kovač-Kavčič und Skalerič 2000).

Unter den genannten klinischen Parametern, die mit SV assoziiert werden, ist die Diagnose einer arterielle Hypertonie besonders hervorzuheben. Der epidemiologische Zusammenhang dieser Befunde wurde in den letzten Jahren in mehreren diagnostischen Studien (Hedström, Albrektsson, und Bergh 2015; Accardo u. a. 2021; Ahadian u. a. 2023) nachgewiesen und in einer 2024 erschienen Meta-Analyse bestätigt (Eslami u. a. 2024). Weiterführend wurde auch eine Korrelation von SV mit dem Ereignis-Risiko der Koronaren Herzerkrankung gefunden (Bergh u. a. 2023); in dieser Arbeit wird auch auf die in dieser Zusammenfassung behandelte Publikation (Klein u. a. 2023) referenziert.

Das Interesse an SV begründet sich in der Funktion SV als prädiktiver Marker für andere klinische Parameter. Im Vergleich zu den genannten, bisher untersuchten Parametern, die sich mit modernen Messgeräten sehr zuverlässig bestimmen lassen, ist der Befund

SV selbst sehr subjektiv und Untersucher-abhängig. Zudem existiert bisher kein Goldstandard für die SV-Bestimmung. Die meisten vorangehenden Arbeiten zu SV wählten eine binäre Klassifikation (0: keine SV, 1: SV) (Hedström, Albrektsson, und Bergh 2015; Akkaya, Ölmez, und Özkan 2019; Hedström und Bergh 2010; Al-Shayyab und Baqain 2015; Bergh, Albrektson, u. a. 2022; Accardo u. a. 2021; Bergh, Kastberg, u. a. 2022) die jeweils von wenigen medizinischen Experten vorgenommen wurde (Hedström, Albrektsson, und Bergh 2015; Akkaya, Ölmez, und Özkan 2019; Al-Shayyab und Baqain 2015; Bergh, Albrektson, u. a. 2022). In bisherigen Arbeiten wurde die Reliabilität (Messgenauigkeit) der SV-Inspektion nicht systematisch berücksichtigt. Der prädiktive Wert SV hängt dabei von der SV-Reliabilität in der klinischen Inspektion ab und die Inspektionsreliabilität SV beeinflusst direkt die Korrelation SV mit anderen klinischen Parametern.

Fragestellung: Diese Arbeit untersucht erstmals systematisch die Inspektionsreliabilität SV im Rahmen einer diagnostischen Studie. Diese Reliabilität soll in einem geeigneten Messmodell geschätzt und daraus ihr Einfluss auf die Qualität SV als prädiktiver Marker quantifiziert werden. Zur Optimierung der SV-Reliabilität wird der RA (relative area) – Score entwickelt und evaluiert. SV dienen hierbei exemplarisch als klinischer Parameter mit geringer Reliabilität, das methodische Vorgehen soll auch auf andere klinische Marker anwendbar sein.

1.2 Material und Methoden

Die Inspektionsreliabilität SV wurde im Rahmen einer zweiteiligen, diagnostischen Studie untersucht. Die klinische Studie wurde von der Ethikkommission der Medizinischen Fakultät Bonn genehmigt (Kennnummer 296/19) und alle Experimente wurden unter Berücksichtigung der Deklaration von Helsinki durchgeführt.

1.2.1 Klinische Studie

Zunächst wurden zwischen 2019 und 2020 im Rahmen einer diagnostischen Studie an der Klinik und Poliklinik für Zahnärztliche Prothetik, Propädeutik und Werkstoffwissenschaften des Universitätsklinikums Bonn n = 78 Patienten (43 Frauen, 35 Männer) im Alter von 22 bis 82 Jahren rekrutiert. Ausschlusskriterien hinsichtlich Alter oder Vorerkrankungen gab es keine. Die Patienten waren für eine routinemäßige

Untersuchung oder Behandlung in der Zahnklinik. Im Rahmen der Studie wurden vor der zahnärztlichen Behandlung mehrere Fotografien der Unterseite der Zunge erstellt (Kamera: Canon EOS 800D, Canon EF 100 mm Makroobjektiv, Nissin MF18 Makro-Ringblitz).

1.2.2 Experiment zur klinischen Inspektion

Die aufgenommenen Bilder der klinischen Studie wurden in zufälliger Reihenfolge $m = 23$ medizinischen Experten präsentiert und von diesen hinsichtlich des Vorliegens von SV beurteilt. Etwa die Hälfte der Gruppe der medizinischen Experten waren Ärzte oder Studenten in klinischer Ausbildung, die andere Hälfte waren Zahnärzte. Alle rekrutierten medizinischen Experten führten im Rahmen ihrer beruflichen Tätigkeit regelmäßig klinische Untersuchungen der Mundhöhle durch oder waren im Rahmen ihres Studiums damit vertraut. Alle rekrutierten medizinischen Experten wurden im Vorfeld über SV als Befund der klinischen Inspektion anhand von Bildbeispielen instruiert.

1.2.3 Statistische Analyse

Zu jedem der $n = 78$ Patientenbilder wurden die Befunde (1: SV, 0: keine SV) der $m = 23$ Experten in einer Matrix erfasst. Da bisher kein Goldstandard für die SV-Diagnostik existiert, wurden die Mittelwerte der 0/1-Bewertung in Bezug auf ein Bild der Zungenunterseite als Surrogatparameter für SV verwendet und als SV-Prävalenz SV_P bezeichnet.

Quantifizierung der Schwierigkeit der SV-Inspektion

Um die Schwierigkeit der SV-Inspektion zu quantifizieren, wurde eine Divergenz zwischen den Einschätzungen mehrerer medizinischer Experten zu einem Bild als Hinweis auf die Schwierigkeit der Befundung gewertet. Dabei stellt sich ein duales Problem: Einerseits muss bestimmt werden, wie genau die erhobenen Fotografien SV abbilden, eine Reliabilitätsschätzung der Bildserie bezüglich SV. Andererseits muss die Übereinstimmung der medizinischen Experten bezüglich dieser Bilder analysiert werden. Unter der Annahme eines essentiell τ -äquivalenten Messmodells (Moosbrugger und Kelava 2020) wurde die Reliabilität der Bildserie für SV mit Cronbach's α geschätzt; als Maß für die Inter-item-Korrelation gibt α an, wie gut die ausgewählten Bilder SV

repräsentieren. Die Übereinstimmung der medizinischen Experten bezüglich der SV-Diagnostik wurde bei nominal skalierten Daten (0/1) und $m > 2$ medizinischen Experten mit Fleiss' κ analysiert (Moosbrugger und Kelava 2020).

Schätzung der maximalen linearen SV-Korrelation

Um die Prädiktion durch SV zu quantifizieren, muss die Reliabilität der SV-Inspektion geschätzt werden. Dabei ist die Reliabilität R der Anteil der Varianz der wahren Werte τ an der Gesamtvarianz der beobachteten Werte X : $R = \frac{\sigma_{\tau}^2}{\sigma_X^2}$. Unter der Annahme eines parallelen Messmodells (Moosbrugger und Kelava 2020) wurde die SV-Inspektionsreliabilität mit einer speziellen Split-Half-Korrelation geschätzt. Dazu wurden die $m = 23$ medizinischen Experten in zwei etwa gleich große Gruppen aufgeteilt und für jedes Bild die Mittelwerte SV_{P1} , SV_{P2} der SV-Befundung aller Bilder in beiden Gruppen berechnet. Die Test-Retest-Korrelation der SV_P -Werte wurde dann als Reliabilitätsschätzer der SV-Diagnostik verwendet. Mit diesem Vorgehen war es möglich, unter Kenntnis einer geschätzten Reliabilität R einer Messgröße X (hier $X = SV_P$), die maximal mögliche lineare Korrelation r mit einem beliebigen anderen Parameter Y nach oben abzuschätzen: $r(X, Y) \leq \sqrt{R}$ (Danner 2015).

Qualitative Analyse des RA-Score

Die Analyse des RA-Scores $RA := \frac{A}{L^2}$ (A : Fläche der sichtbaren sublingualen Venen (px^2), L : Zungenlänge (px) wurde mit einer Receiver-Operating-Characteristics (ROC)-Kurve gegen SV_P als Surrogatparameter für SV durchgeführt. Die Bestimmung des RA-Scores ist exemplarisch in Abbildung 1 dargestellt. Die gegenständliche Fläche A der sichtbaren sublingualen Venen bezieht sich hierbei ausschließlich auf die Zungenunterseite. Die Zungenlänge L bestimmt sich ebenfalls an der Zungenunterseite durch Abmessung von der Zungenspitze, entlang des *Frenulum linguae* bis zur *Caruncula sublingualis*. Für eine numerische Analyse wurden die Konditionszahlen φ_A , φ_L für die Eingabeparameter A , L berechnet. Für $i = 1, \dots, n$ ist die Konditionszahl φ_i einer Aufgabe $f(x)$ in Bezug auf das i -te Argument definiert als: $\varphi_i := \frac{x_i}{f(x)} \frac{\partial f}{\partial x_i}(x)$. Konditionszahlen sind ein quantitatives Maß für die Abhängigkeit der Zielparameter von Störungen in den Eingangsdaten. Die Konditionszahl φ_i ist der Faktor, um den der Eingangsfehler des Eingangsparameters i

maximal, d.h. im ungünstigsten Fall, durch die Berechnung verstärkt werden kann (Richter und Wick 2017). Die statistische Analyse wurde mit Excel und die ROC-Analyse mit R durchgeführt. Die Einfärbung und Messung der Fläche A und der Länge L, die für den RA-Score verwendet werden, wurden mit GIMP (GNU Image Manipulation Program) durchgeführt. Die Messung der Fläche A und der Länge L wurde von einem Arzt mit Erfahrung in der Oralmedizin und der klinischen SV-Inspektion (einer der Autoren der Publikation) durchgeführt. A und L wurden zweimal mit einem Abstand von mehreren Tagen gemessen.

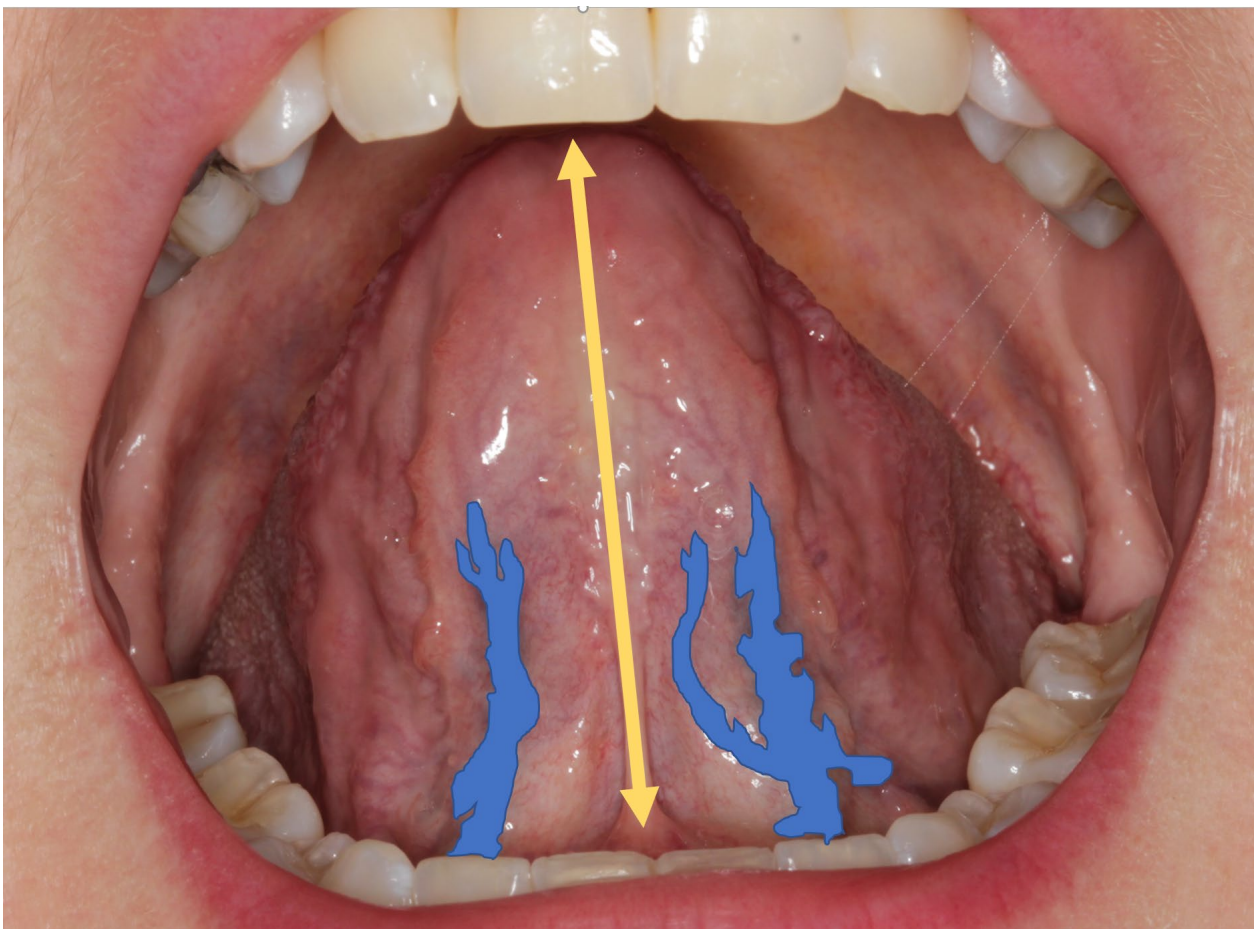


Abbildung 1: Die Fläche A der sichtbaren sublingualen Venen (px^2) ist blau eingefärbt. Die Zungenlänge L (px) ist als gelber Pfeil eingezeichnet. Für die Berechnung des RA-Scores wird die Fläche A durch das Quadrat der Zungenlänge geteilt: $RA := \frac{A}{L^2}$.

1.3 Ergebnisse

1.3.1 Schwierigkeit der SV-Inspektion in der klinischen Untersuchung

Die Schwierigkeit der klinischen SV-Inspektion wurde mit Cronbachs $\alpha = 0,937$ und Fleiss' $\kappa = 0,397$ bestimmt. Die Summe der Varianz der einzelnen Bilder war im Verhältnis zur Gesamtvarianz vergleichsweise klein. Die Varianz der Beobachter in Bezug auf die Bilder war größer als die Summe der Varianzen der Bilder über alle Beobachter. Die geringe nicht-zufällige Übereinstimmung ist somit eher auf die Varianz der Befunde zwischen den medizinischen Experten als auf die Varianz der einzelnen Bildbefunde zurückzuführen. In diesem Sinne ist die Inspektion von SV schwer reproduzierbar und eine schwierige Aufgabe der klinischen Untersuchung. Bei der Interpretation dieser Ergebnisse ist es wesentlich, dass, anders als in der Klassischen Testtheorie, Cronbachs α in dem hier verwendeten Kontext kein direkter Reliabilitätsschätzer der klinischen SV-Inspektion darstellt, sondern ein Maß dafür, wie genau die ausgewählten Bilder SV abbilden. Konsekutiv quantifiziert Fleiss' κ als Interrater-Reliabilitätsmaß die Übereinstimmung der medizinischen Experten bezüglich ihrer SV-Befunde.

1.3.2 Maximale Korrelation von SV mit einem beliebigen anderen Parameter Y

Mit der beschriebenen speziellen Test-Retest-Korrelation (Abbildung 2) als Reliabilitätsschätzer R der SV-Inspektion, fanden wir eine Inspektionsreliabilität für SV von $R = 0,847$. Daraus ergab sich eine maximale lineare Korrelation von $r_{\max}(SV_P, Y) = 0,920$.

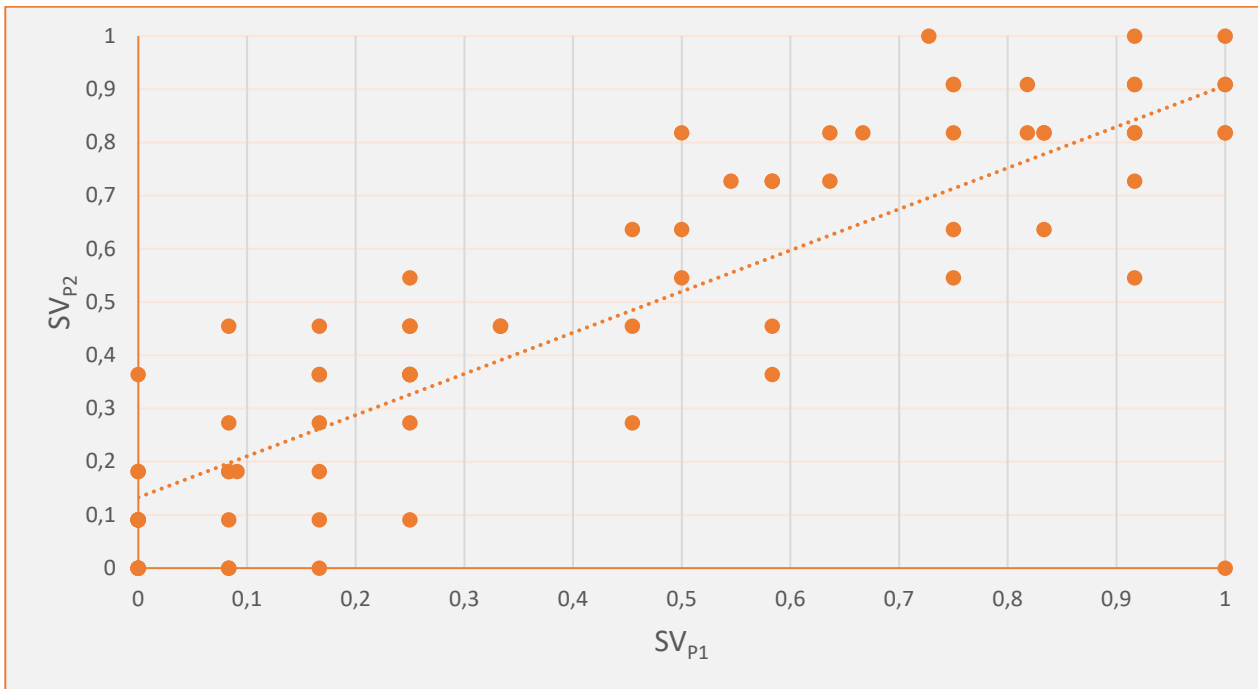


Abbildung 2: Visualisierung der speziellen Test-Retest-Korrelation im parallelen Messmodell. Die dargestellte Korrelation dient als Reliabilitätsschätzer R der SV-Inspektion. Diese beschränkt die Bravais-Pearson-Korrelation von SV_P mit beliebigen anderen Parametern Y .

1.3.3 Der RA-Score als objektiver Marker und kontinuierliches Klassifikationssystem für SV

Um das Problem unzureichender SV-Reliabilität in der klinischen Inspektion zu überwinden, wurde der RA-Score (relative area) als neues objektives Maß und kontinuierliches Klassifizierungssystem für SV eingeführt. Hierbei ist zu betonen, dass ausschließlich Da die absoluten Zungenflächen der sichtbaren Venenwanddilataion der Zungenunterseite aufgrund unterschiedlicher Größen- und Inspektionsverhältnisse nicht gut vergleichbar sind, wird die Zungenlänge als zuverlässige Normierung verwendet: Der RA-Score bestimmt sich als die Fläche der sichtbaren sublingualen Venen A (px^2), normiert auf das Quadrat der Zungenlänge L^2 (px^2): $RA := \frac{A}{L^2}$. Der RA-Score ist ein dimensionsloses Maß mit Wertebereich $0 < RA < 1$. Die Test-Retest-Reliabilität der RA-Score-Bestimmung ist gut, mit einer Varianz < 10 (px^2/px) in A und L . Der RA-Score korreliert besser mit SV_P als die Einzelvariablen A und L . Er stellt somit ein kontinuierliches

und weitgehend vom Untersucher unabhängiges Scoring-System für SV dar. Da bei der Definition des RA-Scores nicht zwischen linker und rechter Seite der Zunge unterschieden wird, ist für eine stabile Berechnung eine Kameraposition nahe der Sagittalachse erforderlich, um die gesamte Zungenunterseite zuverlässig zu erfassen. Mit $\varphi_A = 1$, $\varphi_L = -2$ ist der RA-Score gut konditioniert, sodass bei Störung der Eingangsparameter A, L keine übermäßige Fehlerverstärkung durch die Berechnung selbst zu erwarten ist. Die ROC-Analyse des RA-Score auf Grundlage von SV_P zeigt mit $AUC = 72,5 \%$ eine gute Trennschärfe in Bezug auf SV_P . Da bei der klinischen SV-Inspektion neben der 2D-Projektion der Venenwanddilataion auch andere Parameter berücksichtigt werden, kann keine 100%ige Unterscheidungsgenauigkeit erwartet werden. Da der RA-Score nur die 2D-Projektion der Venenwanddilataion und nicht andere Faktoren wie die Topographie oder die Farbverteilung der sublingualen Venen berücksichtigt, lässt sich daraus schließen, dass die Einschätzung der SV_P durch medizinische Experten ebenfalls stärker von geometrischen Eindrücken beeinflusst wird als von anderen, die der RA-Score nicht messen kann. Ebenso kann der RA-Score auch zur Bewertung der automatischen SV-Segmentierung verwendet werden.

1.4 Diskussion

SV sind ein gut untersuchtes Gebiet der Oralmedizin. Viele frühere Studien diagnostizierten SV auf einer binären Skala (0/1) mit relativ wenigen Untersuchern (Hedström, Albrektsson, und Bergh 2015; Akkaya, Ölmez, und Özkan 2019; Hedström und Bergh 2010; Al-Shayyab und Baqain 2015; Bergh, Albrektson, u. a. 2022; Accardo u. a. 2021; Bergh, Kastberg, u. a. 2022). In dieser Arbeit wurde erstmals systematisch der Einfluss der Reliabilität der SV-Inspektion untersucht. Es wurde gezeigt, dass SV-Inspektion prinzipiell möglich, die Inspektionsreliabilität jedoch relativ gering ist. Die Diagnose SV weist eine deutlich größere klinische Heterogenität auf als bisher angenommen. Die Reliabilität der SV-Inspektion beschränkt die maximal mögliche Korrelation von SV mit beliebigen anderen klinischen Parametern, eine Korrelation mit einem beliebigen weiteren Parameter scheint prinzipiell nie 100 % erreichen zu können. Das limitiert die Qualität SV als prädiktiven Marker. Die Reliabilität ist ein wesentliches Qualitätsmerkmal prädiktiver Marker; bisher gefundene Korrelationen SV mit anderen klinischen Parametern, beispielsweise die populäre Assoziation SV und arterieller

Hypertonie (Hedström, Albrektsson, und Bergh 2015; Bergh, Albrektson, u. a. 2022; Hedström und Bergh 2010), sollten auf dem Hintergrund der SV-Reliabilität kritisch geprüft werden.

Um die Reliabilität und Objektivität der SV-Diagnostik zu erhöhen, wurde der RA-Score eingeführt und evaluiert. Dieser setzt die 2D-Projektion der sublingualen Venen A ins Verhältnis zum Quadrat der Zungenlänge L. Er kann daher als Vergleichs- und Normierungspunkt verschiedener klinischer Inspektionsbefunde im Rahmen der SV-Diagnostik verwendet werden. Der RA-Score ist ein weitgehend untersucherunabhängiges Maß mit hoher Reliabilität in A und L. Er ist algorithmisch gut konditioniert und könnte daher in klinischen SV-Studien und für die automatische SV-Segmentierung (Chiu, Lan, und Chang 2002; Yan, Wang, und Li 2008; Lin u. a. 2014) verwendet werden.

Der metrische Wertebereich von SV_P und RA bildet die die Kontinuität der anatomischen SV-Varianz genauer ab als eine binäre 0/1-Klassifikation. Durch ihr metrisches Skalenniveau wird eine Korrelation mit anderen (metrischen) Parametern methodisch überhaupt erst möglich, was bei der bisher verwendeten binären/diskreten Klassifikation nicht der Fall war.

Die verwendeten statistischen Methoden lassen sich auch auf andere klinische Parameter als SV anwenden. Das demonstrierte Vorgehen ermöglicht allgemein eine Qualitätsanalyse prädiktiver Marker durch Reliabilitätsschätzung. Die Reliabilität des Markers ermöglicht dann a-priori eine Aussage über die maximal mögliche Korrelation des Markers mit einem beliebigen anderen Parameter, da die Korrelation durch die Reliabilität des Markers limitiert wird. Auf diese Weise ist eine Marker-Selektion ohne Kenntnis expliziter Korrelationen zwischen Marker und zu prädizierendem Parameter möglich.

Eine Limitation der Arbeit ist die verhältnismäßig geringe Stichprobengröße von $n = 78$ Patienten in der diagnostischen Studie. Es wäre interessant, die vorgestellten Methoden zur Schätzung der Reliabilität auf eine größere Kohorte anzuwenden. Durch die Verwendung einer frequentistischen SV-Definition (SV_P), wäre es möglich, eine genauere Prävalenzschätzung der SV und eine stabilere Schätzung der maximal möglichen

Korrelation r_{\max} (aus der Reliabilitätsschätzung R) zu erhalten. Zudem wäre eine automatisierte Berechnung des RA-Scores mit Methoden des maschinellen Lernens interessant, wie dies bereits für die binäre (0/1) SV-Klassifikation etabliert wurde (Lu u. a. 2022).

Die Reliabilität der SV-Inspektion ist wichtig für die Qualität der SV als prädiktiver Marker. Dies sollte bei der Interpretation bisheriger SV-Studien berücksichtigt werden und hat Auswirkungen auf zukünftige Studien. Für eine Optimierung der SV-Reliabilität in der klinischen Praxis, kann bei systematischer SV-Diagnostik der RA-Score, statt einer subjektiven, Untersucher-abhängigen Einschätzung, empfohlen werden.

Der RA-Score ermöglicht die SV-Diagnose auch für Ärzte, die nicht routinemäßig orale Untersuchungen durchführen. Damit ermöglicht der RA-Score die Integration der SV-Diagnostik in andere medizinische Fachgebiete außerhalb der Oralmedizin.

1.5 Zusammenfassung

Die vorliegende Arbeit widmet sich der Untersuchung von sublingualen Varizen (SV), erweiterten sublingualen Venen, in der Oralmedizin. Die klinische Bedeutung der SV liegt in ihrer Rolle als prädiktiver Marker für andere klinische Parameter. Bisherige Studien haben einen Zusammenhang zwischen SV und Parametern wie Alter, arterieller Hypertonie, Diabetes mellitus Typ 2, Varikosis der unteren Extremitäten, Rauchen, Diabetes mellitus Typ 2, Dyslipidämie, Leberzirrhose und Allergien hergestellt. Die Diagnose von SV beruht jedoch auf subjektiven, vom Untersucher abhängigen Beurteilungen, und es gibt keinen etablierten Standard für die inspektorische Beurteilung von SV. Ziel dieser Studie ist es, die Zuverlässigkeit der SV-Inspektion systematisch zu untersuchen und ihren Einfluss auf die Verwendung von SV als prädiktiven Marker zu quantifizieren. Die Reliabilität der SV-Inspektion wurde in einer diagnostischen Studie analysiert, in der 78 Patienten von 23 medizinischen Experten hinsichtlich des Vorliegens von SV beurteilt wurden. Die Ergebnisse zeigten, dass die Übereinstimmung zwischen den Experten gering war, was auf die Herausforderung hinweist, SV objektiv zu beurteilen. Die geringe Inspektionsreliabilität und Heterogenität der SV-Diagnose limitiert statistisch die maximal mögliche Korrelation von SV mit beliebigen anderen Parametern und damit die Vorhersagefähigkeit von SV als prädiktiver Marker. Um die Zuverlässigkeit der SV zu

verbessern, wurde der RA-Score (relative area) entwickelt und getestet. Dieser Score basiert auf der Fläche der sichtbaren sublingualen Venen, normiert auf das Quadrat der Zungenlänge. Der RA-Score bietet eine objektivere und kontinuierlichere Methode zur Beurteilung der SV als die herkömmliche subjektive Beurteilung. Insgesamt trägt diese Studie zu einem besseren Verständnis der Schwierigkeiten bei der SV-Diagnose bei und bietet eine mögliche Lösung zur Verbesserung der Genauigkeit und Objektivität dieser Diagnose. Die Ergebnisse zeigten, dass die SV-Inspektion relativ unzuverlässig ist, was die Verwendung von SV als prädiktiver Marker für andere Parameter einschränkt. Die Reliabilität ist ein entscheidender Faktor für die Qualität eines solchen Markers. Die Einführung des RA-Scores stellt jedoch eine vielversprechende Lösung dar, um die Reproduzierbarkeit der SV-Diagnose zu erhöhen und sie für verschiedene medizinische Fachgebiete zugänglicher zu machen.

1.6 Literaturverzeichnis der deutschen Zusammenfassung

Accardo A, Pascazio L, Costantinides F, Gorza F, Silveri G. Influence of hypertension and other risk factors on the onset of sublingual varices. *BMC Oral Health*. 2021; 21(1):235

Ahadian H, Akhavankarbassi MH, Sabaghzadegan Y, Owlia F, Daneshmand AS. Is there any difference based on sublingual varices frequency between hypertensive patients and healthy persons? *BMC Oral Health* 2023; 666

Akkaya N, Ölmez D, Özkan G. Evaluation of the factors associated with sublingual varices: a descriptive clinical study. *Folia Morphol* 2019; 78(2): 325-330

Al-Shayyab MH, Baqain ZH. Sublingual varices in relation to smoking, cardiovascular diseases, denture wearing, and consuming vitamin rich foods. *Saudi Med J*. 2015; 36(3): 310-315

Barzideh N, Alaei A, Azizi A. The Relationship between Smoking and Sublingual Varices in the Elderly. *Oman Med J*. 2021; 36(4): e288

Bergh H, Albrektson M, Kastberg C, Baigi A, Hedström L. The Association Between Sublingual Varices and Cardiovascular Risk Factors. *Vasc Health Risk Manag.* 2022; 18: 319-327

Bergh H, Albrektson M, Kastberg C, Hedström L. Association of Sublingual Varices With Heart- and Cerebrovascular Disease. *Int Dent J.* 2023; 22:S0020-6539(23)00438-0

Bergh H, Kastberg C, Albrektson M, Hedström L. Persistence and stability of sublingual varices over time and their connection to underlying factors: an 8 year follow up study. *BMC Oral Health* 2022; 22(1): 346

Bhatnagar P, Rai S, Bhatnagar G, Kaur M, Goel S, Prabhat M. Prevalence study of oral mucosal lesions, mucosal variants, and treatment required for patients reporting to a dental school in North India: In accordance with WHO guidelines. *J Family Community Med.* 2013; 20(1): 41-48

Chiu CC, Lan CY, Chang YH. Objective assessment of blood stasis using computerized inspection of sublingual veins. *Comput Methods Programs Biomed.* 2002; 69(1): 1-12

Corrêa PH, Nunes LC, Johann AC, Aguiar MC, Gomez RS, Mesquita RA. Prevalence of oral hemangioma, vascular malformation and varix in a Brazilian population. *Braz Oral Res.* 2007; 21(1): 40-45

Danner D. Reliabilität – die Genauigkeit einer Messung. Mannheim, GESIS – Leibniz-Institut für Sozialwissenschaften (SDM Survey Guidelines). 2015

Duarte NT, de Oliveira Godoy A, da Rocha Tenório J, Andrade NS, Franco JB, Pérez-Sayáns M, Ortega KL. Prevalence of sublingual varices in patients with cirrhosis and the correlation with nitrogen compounds. *Oral Surg Oral Med Oral Pathol Oral Radiol.* 2020; 129(1): 39-44

Egito Vasconcelos BC, Novaes M, Sandrini FA, Maranhão Filho AW, Coimbra LS. Prevalence of oral mucosa lesions in diabetic patients: a preliminary study. *Braz J Otorhinolaryngol.* 2008; 74(3):423-428

Eslami H, Halimi Milani F, Salehnia F, Kourehpaz N, Katebi K. Relationship between sublingual varices and hypertension: a systematic review and meta-analysis. *BMC Oral Health.* 2024; (1): 240

Espinoza I, Rojas R, Aranda W, Gamonal J. Prevalence of oral mucosal lesions in elderly people in Santiago, Chile. *J Oral Pathol Med.* 2003; 32(10): 571-575

Ettinger RL, Manderson RD. A clinical study of sublingual varices. *Oral Surg Oral Med Oral Pathol.* 1974; 38(4): 540-545

Gonzalez Ballerga E, Pozo MO, Rubatto Birri PN, Kanoore Edul VS, Sorda JA, Daruich J, Dubin A. Sublingual microcirculatory alterations in cirrhotic patients. *Microcirculation.* 2018; 25(4): e12448

González-Álvarez L, García-Pola MJ. Risk factors associated with tongue lesions: a propensity score-matched case-control study. *Med Oral Patol Oral Cir Bucal.* 2022; 27(1): e25-e34

Hedström L, Albrektsson M, Bergh H. Is there a connection between sublingual varices and hypertension? *BMC Oral Health.* 2015; 15: 78

Hedström L, Bergh H. Sublingual varices in relation to smoking and cardiovascular diseases. *Br J Oral Maxillofac Surg.* 2010; 48(2):136-138

Klein CR, Stoppenbrink D, Geier J, Mayr A, Stark H. The relative area score for sublingual varices reliability measurement: a diagnostic study. *BMC Oral Health.* 2023; 23(1): 366

Kovac-Kovacic M, Skaleric U. The prevalence of oral mucosal lesions in a population in Ljubljana, Slovenia. *J Oral Pathol Med.* 2000; 29(7): 331-335

Lazos J, Rodriguez Marco E, Panico R, Romero JC, Cámara L, Brunotto M, & Piemonte E. Oral varicose veins: Clinical characterization and association with medical conditions. *Oral Surg Oral Med Oral Pathol Oral Radiol.* 2020; 129(1): e150

Lin H-J, Chen Y-J, Damdinsuren N, Tan T-H, Liu T-Y, Chiang J-Y. Automatic Sublingual Vein Feature Extraction System. *International Conference on Medical Biometrics, Shenzhen, China.* 2014: 55-62

Lu PH, Chiang CC, Yu WH, Yu MC, Hwang FN. Machine Learning-Based Technique for the Severity Classification of Sublingual Varices according to Traditional Chinese Medicine. *Comput Math Methods Med.* 2022; 2022: 1-10

Moosbrugger H, Kelava A. *Testtheorie und Fragebogenkonstruktion.* Springer Berlin, Heidelberg 2020: 3. Auflage: 1-770

Mozafari PM, Dalirsani Z, Delavarian Z, Amirchaghmaghi M, Shakeri MT, Esfandyari A, Falaki F. Prevalence of oral mucosal lesions in institutionalized elderly people in Mashhad, Northeast Iran. *Gerodontology.* 2012; 29(2): e930-934

Rabiei M, Kasemnezhad E, Masoudi rad H, Shakiba M, Pourkay H. Prevalence of oral and dental disorders in institutionalised elderly people in Rasht, Iran. *Gerodontology.* 2010; 27(3): 174-177

Richter T, Wick T. *Einführung in die Numerische Mathematik: Begriffe, Konzepte und zahlreiche Anwendungsbeispiele.* Springer Spektrum Berlin, Heidelberg 2017: 1. Auflage: 1-478

Triantos D. Intra-oral findings and general health conditions among institutionalized and non-institutionalized elderly in Greece. *J Oral Pathol Med.* 2005; 34(10): 577-582

Yan Z, Wang K, Li N. Segmentation of Sublingual Veins from near Infrared Sublingual Images. 8th IEEE International Conference on Bioinformatics and BioEngineering, Athens, Greece. 2008: 1–5

2. Veröffentlichung

Klein et al. *BMC Oral Health* (2023) 23:366
<https://doi.org/10.1186/s12903-023-03068-x>

BMC Oral Health

RESEARCH

Open Access



The relative area score for sublingual varices reliability measurement: a diagnostic study

Christian R. Klein^{1*}, David Stoppenbrink¹, Jannik Geier², Andreas Mayr³ and Helmut Stark¹

Abstract

Background Sublingual varices (SV) and their predictive potential for other clinical parameters is a much studied topic in oral medicine. SVs have been well studied as predictive markers for many common diseases such as arterial hypertension, cardiovascular disease, smoking, type 2 diabetes mellitus and age. Despite many prevalence studies, it is still unclear how the reliability of SV inspection affects its predictive power. The aim of this study was to quantify the inspection reliability of SV.

Methods In a diagnostic study, the clinical inspection of 78 patients by 23 clinicians was examined for the diagnosis of SV. Digital images of the underside of the tongue were taken from each patient. The physicians were then asked to rate them for the presence of sublingual varices (0/1) in an online inspection experiment. Statistical analysis for inter-item and inter-rater reliability was performed in a τ -equivalent measurement model with Cronbach's α and Fleiss κ .

Results The interrater reliability for sublingual varices was relatively low with $\kappa=0.397$. The internal consistency of image findings for SV was relatively high with $\alpha\approx 0.937$. This shows that although SV inspection is possible in principle, it has a low reliability R . This means that the inspection finding (0/1) of individual images often cannot be reproduced stably. Therefore, SV inspection is a difficult task of clinical investigation. The reliability R of SV inspection also limits the maximum linear correlation r_{max} of SV with an arbitrary other parameter Y . The reliability of SV inspection $R=0.847$ limits the maximum correlation to $r_{max}(SV, Y)=0,920$ —a 100% correlation was a priori not achievable in our sample. To overcome the problem of low reliability in SV inspection, we propose the RA (relative area) score as a continuous classification system for SV, which normalises the area of visible sublingual veins to the square of the length of the tongue, providing a dimensionless measure of SV.

Conclusions The reliability of the SV inspection is relatively low. This limits the maximum possible correlation of SV with other (clinical) parameters. SV inspection reliability is an important indicator for the quality of SV as a predictive marker. This should be taken into account when interpreting previous studies on SV and has implications for future studies. The RA score could help to objectify the SV examination and thus increase its reliability.

Keywords Sublingual varices, Reliability, Clinical inspection, Aging, Oral science, Maximum correlation, Clinical markers, Preventive health care

*Correspondence:

Christian R. Klein
christian.klein@ukbonn.de

Full list of author information is available at the end of the article



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Background

Sublingual varices (SV) are a widely studied topic in oral medicine. SV are dilated changes in the sublingual veins (*Vv. linguae profundae*, Fig. 1). Previous studies have shown a predictive value of SV for a number of clinical parameters such as arterial hypertension [1–3], smoking [4–7], type 2 diabetes mellitus [6, 8], dyslipidaemia [6, 8], thrombosis [9, 10], cirrhosis [11], allergies [12] and age [13], where SV remain persistent with increasing age [14].

The diagnosis of SV on clinical examination depends on the examiner. Despite many studies on SV, the influence of the examiner on the diagnosis of SV and consequently the accuracy of SV measurement is not clear. Previous studies of SV have been conducted on large patient samples ($n > 400$, dental patients) [1, 2, 5, 6, 8, 14, 15], often with an inclusion criterion of age > 40 years [1, 4, 6, 8, 14]. Many previous studies on SV used a binary 0/1 classification [1, 2, 4–6, 8, 14] by relatively few (1–2) medical experts [1, 2, 5, 6]. There is currently no objective gold standard for the diagnosis of SV on clinical inspection. This study was the first to systematically investigate the influence of the examiner on the diagnosis of SV during oral inspection. This was used to statistically quantify the influence of examiner reliability on SV quality as an indicator of other clinical parameters. In addition, the RA (relative area) score is proposed as a continuous classification system for SV to increase the reliability and objectivity of SV inspection.

Methods

Design and setting

Our study design to investigate the reliability of SV-inspection consists of two parts:

- 1) **Clinical study:** A clinical study was conducted between 2019 and 2020 in the Department of Pros-



Fig. 1 Example of SV: Dilated changes in the sublingual veins (*Vv. linguae profundae*). Written consent obtained from patient

thetic Dentistry, Preclinical Education and Materials Science, Faculty of Dentistry, Rheinische Friedrich-Wilhelms-Universität, University Hospital Bonn, Bonn, Germany. Patients attending the dental clinic for a routine examination or treatment were invited to participate in the study. There were no exclusion criteria regarding age or previous diseases. Prior to dental treatment, several photographs of the underside of the tongue were taken with a special camera for oral cavity photographs (measuring device: Canon EOS 800D, Canon EF 100 mm macro, Nissin MF18 macro ring flash). A total of $n = 78$ patients (43 women and 35 men) aged 22 to 82 years, were included.

- 2) **Experiment on clinical inspection:** The recorded images of the clinical study were presented in an arbitrary order and then rated in by $m = 23$ medical professionals for the detection of SV. Approximately half of the group of medical experts were medical doctors or students in clinical training, and the other half were dentists. All recruited medical experts regularly perform clinical inspections of the oral cavity in their profession or are familiar with it as part of their studies. All recruited medical experts were instructed in advance about SV as a clinical finding on examination by means of image examples.

This experiment was conducted with a freely available computer programme called *yesnomaybe* (figure 2, <https://github.com/richard-vock/yesnomaybe>). The findings of an individual expert were stored in binary form (1 = SV detected, 0 = no SV detected) and the entire inspection was stored in vector form. Subsequently, the data from all medical professionals were collated into in a matrix, a schematic illustration of the procedure is shown in figure 3.

SVP as a surrogate parameter for SV

As there is no practical definition of SV exists yet due to the lack of a gold standard, we used the mean values of the 0/1 ratings with respect to an image of the underside of the tongue as surrogate parameter. We refer to this as SV subject prevalence (SV_p).

Statistical analysis

Statistical analysis was performed with Excel and ROC analysis using R. The colouring and measurement of the area A and length L used in the RA score were performed using GIMP (GNU Image Manipulation Program). The measurement of area A and length L was performed by a physician with experience in oral medicine and clinical SV inspection (one of the authors). A and L were measured twice with an interval of several days.

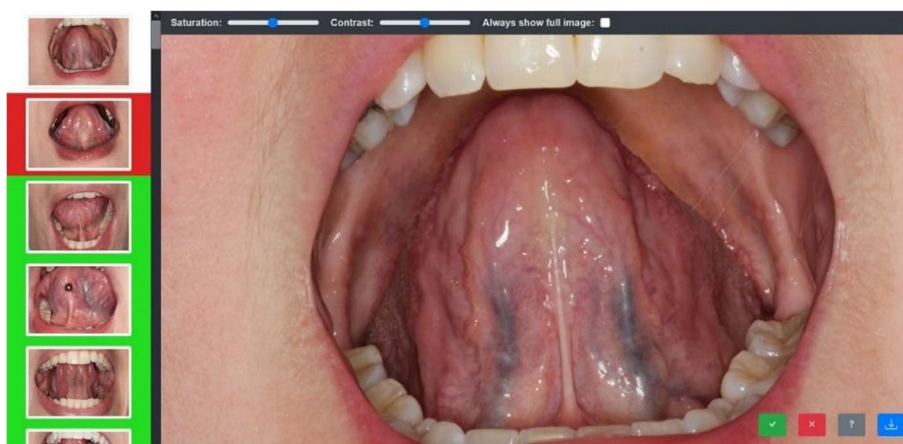


Fig. 2 Exemplary representation of the experiment for the clinical inspection of SV. Written consent obtained from the patients. The yesnomabye programme was used for this purpose. The medical professionals were able to evaluate whether they thought SV were present or not with a simple click of the mouse. The binary findings of the individual images (1 = SV detected, 0 = no SV detected) were summarised as a vector at the end, and the data of all medical professionals were then written into a common matrix

	Expert ₁	Expert _{...}	Expert ₁₁	SV _{P1}	Expert ₁₂	Expert _{...}	Expert ₂₃	SV _{P2}
Patient ₁	0	..	1	$\mu_{1,1}$	0	..	0	$\mu_{2,1}$
Patient ₂	0	..	1	$\mu_{1,2}$	1	..	1	$\mu_{2,2}$
:	:	:
Patient ₇₈	1	..	1	$\mu_{1,78}$	1	..	0	$\mu_{2,78}$

Fig. 3 Schematic representation of the test–retest correlation in the parallel measurement model for estimating the reliability of SV inspection. The group of medical experts (m = 23) was divided into two halves and SV_p values were calculated for each image for both groups. Subsequently, the μ_{ij} were correlated with each other to estimate the reliability of inspection (test–retest correlation). The numbers (0,1) inserted in the matrix are used to illustrate the method and are not real measurements from the study

Quantification of SV inspection difficulty

In the absence of a gold standard for the analysis of SV inspection the reliability of the test scale, together with the interrater reliability, can thus serve as a measure of the difficulty of the task. A disagreement between the scores of several medical experts regarding an item is taken as an indication of the difficulty of the findings.

Since an at least essentially τ -equivalent measurement model [16] is given for the evaluation of the individual pictures (test items), we analysed the reliability of the scale with Cronbach’s $\alpha := \frac{n}{n-1} \left(1 - \frac{\sum_{i=1}^n \sigma_{u_i}^2}{\sigma_V^2} \right)$ [17].

$U_i \in \{0, 1\}$ here is the SV rating of the clinical examination of patient $i \in \{1, \dots, n\}$ with $n = 72$. σ_V^2 is

the variance of the frequency of SV assessment V_j across all medical experts $j \in \{1, \dots, m\}$ with $m = 23$. $V_j := \sum_{i=1}^n \delta_{ij} \in \{1, \dots, n\}$ counts the absolute frequency of SV assessment by medical expert j over all $i = 1, \dots, n$ patients, where the SV assessment δ_{ij} of medical expert j regarding patient i is defined by

$$\delta_{ij} := \begin{cases} 1, & \text{if medical expert } j \text{ detects SV in patient } i : U_i = 1 \\ 0, & \text{else} \end{cases}$$

We quantified the interrater reliability, given multiple medians and nominal data (0/1), with Fleiss $\kappa := \frac{p_\lambda - p_e}{1 - p_e}$ [18]. κ is the proportion of the actually achieved over-random agreement $p_\lambda - p_e$ to the in principle achievable over-random agreement $1 - p_e$ of the medical experts in their SV diagnosis. Let n_{i1} be the number of medical experts who diagnosed SV in patient i and n_{i0} the number of medical experts who did not diagnose SV in patient i . Then the probability of agreement between the experts in their SV diagnosis of patient i is given by $p_i := \frac{1}{m(m-1)}(n_{i0}(n_{i0} - 1) + n_{i1}(n_{i1} - 1))$ and $p_\lambda := \frac{1}{n} \sum_{i=1}^n p_i$ denotes the mean of this agreement in SV diagnosis. The relative frequency of the diagnosis or non-diagnosis of SV in relation to the total of all clinical examinations is given by $p_1 := \frac{1}{n \cdot m} \sum_{i=1}^n n_{i1}(\text{SV})$ or $p_0 := \frac{1}{n \cdot m} \sum_{i=1}^n n_{i0}(\text{no SV})$ and $p_e := p_0^2 + p_1^2$.

By combining an inter-item correlation with an interrater correlation, the clinical inspection of SV can be thoroughly assessed. Cronbach's α as a measure of inter-item correlation quantifies how well the selected images represent SV, Fleiss κ as a measure of inter-rater reliability quantifies how well medical professionals agree in their findings on these data.

Estimation of the maximum linear SV correlation

The reliability R of a measurement is the proportion of the variance of the true values τ in the total variance of the observed values X : $R := \frac{\sigma_\tau^2}{\sigma_X^2}$.

We estimate the SV reliability in the clinical inspection experiment by split-half correlation in the parallel measurement model on the set of $m = 23$ medical experts (Fig. 3).

For this purpose, the $m = 23$ medical experts were divided into two groups of approximately equal size and the mean values SV_{p1} , SV_{p2} of the SV inspection of both groups were calculated for each image. With respect to a fixed image i , we assume that SV_{p1} and SV_{p2} estimate the same true value τ_i with the same errors (parallel measurement model). This is plausible because the image is the same and a priori there is no difference in error variance to be assumed between the groups of the two SV_p values. The test-retest correlation of the two SV_p scores (meaning that first mean value μ_{1i} and then mean value μ_{2i} (see Fig. 3)

estimates the true value τ_i of image i) is then an estimator for the reliability of the SV inspection. Note, that Cronbach's α is not an estimate of reliability because it measures how accurately the selected images represent SV, rather than of the accuracy of the clinical inspection of SV.

Knowing the reliability R of a measurement X limits the maximum possible linear correlation r with an arbitrary other parameter Y by

$$r(X, Y) := \frac{cov(X, Y)}{\sqrt{\sigma_X^2} \sqrt{\sigma_Y^2}} = \frac{cov(X, Y)}{\sqrt{\sigma_\tau^2} \sqrt{\sigma_Y^2}} = \sqrt{R} \frac{cov(X, Y)}{\sqrt{\sigma_\tau^2} \sqrt{\sigma_Y^2}} \leq \sqrt{R} \cdot 1.$$

Qualitative analysis of the RA score

The analysis of the RA score $RA := \frac{A}{L^2}$ (A: Area of visible sublingual veins (px²), L: tongue length (px²)) was performed with a receiver operating characteristics (ROC) curve against SV_p as surrogate parameter for SV. Numerical analysis was performed by calculating the condition numbers φ_A , φ_L for the input parameters A, L. For $i = 1, \dots, n$, the condition number φ_i of a task $f(x)$ with respect to the i -th argument is defined as:

$$\varphi_i := \frac{x_i}{f(x)} \frac{\partial f}{\partial x_i}(x)$$

Condition numbers are a quantitative measure of the dependence of the target parameters on disturbances in the input data. The condition number φ_i is the factor by which the input error of the input parameter i can be amplified maximally, i.e. in the worst case, by the calculation [19].

Results

Inspection of SV is a difficult task of clinical examination

We determined the difficulty of clinical inspection of SV using Cronbach's $\alpha = 0.937$ and Fleiss' $\kappa = 0.397$. The sum of the variance of the individual images $\sum_{i=1}^n \sigma_{U_i}^2$ was comparatively small with respect to the total variance σ_V^2 . The variance of the observers with respect to the images was larger than the sum of the variances of the images across all observers. Cronbach's α thus indicates that the clinical examination of SV was, in principle, possible. The interrater reliability was relatively small with $\kappa = 0.397$. Along these lines, inspection of SV was difficult to reproduce and was a difficult task of clinical examination. With Cronbach's $\alpha \approx 0.937$, the low non-random agreement was due to variance in individual observer ratings rather than variance in individual image ratings (Fig. 4).

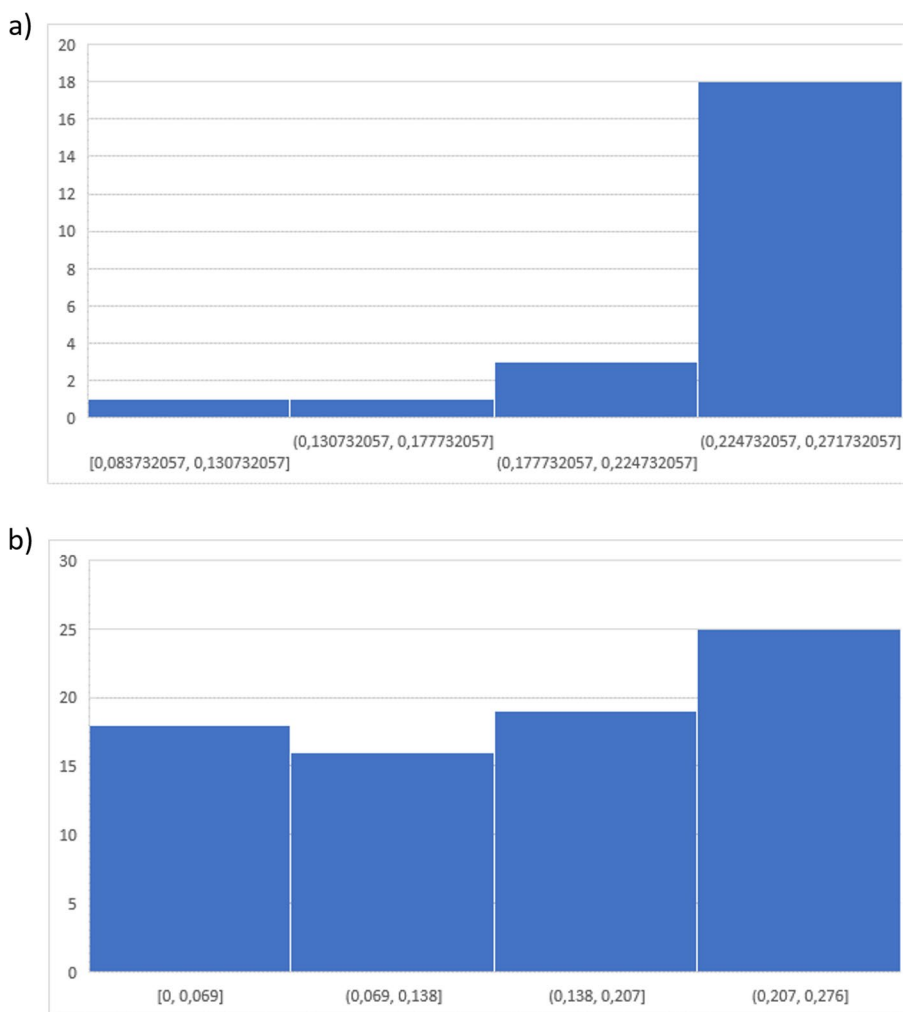


Fig. 4 Clinical SV inspection experiment. Visualisation of Cronbachs α : a) Histogram of observers variances (0/1 classification). b) Histogram of Variances of individual images (0/1 classification)

Maximum correlation of SV with an arbitrary other parameter Y

To estimate the maximum correlation of SV with an arbitrary other parameter Y, we estimated the reliability R of the SV inspection by a special test–retest correlation (Fig. 5). We thus obtained an inspection reliability for SV of $R=0.847$. This resulted in a maximum linear correlation of $r_{max}(SV_p, Y)=0,920$.

RA score as an objective marker and continuous classification system for SV

To overcome the problem of insufficient reliability in clinical inspection, we propose the relative area score (RA) as a new objective measure and continuous classification system for SV. Since absolute tongue areas of visible vein wall dilatation are not well comparable due to different sizes and inspection ratios, tongue

length is used as a reliable normalization: The RA score is determined as the area of visible sublingual veins A (pixels(px)²) normalized to the square of the tongue length L^2 (px²): $RA := \frac{A}{L^2}$ (Fig. 6). This score is a non-dimensional measure in the range of values $0 < RA < 1$. The test–retest reliability of the RA determination is good with a variance of about $< 10 \times$ in A and L . In each case, the RA score correlates better with SV_p than the individual variables A and L . It thus provides a continuous, and largely viewer-independent scoring system for SV. As the definition of the RA score does not distinguish between the left and right sides of the tongue, a camera position close to the saggital axis is required for a stable calculation in order to reliably capture the entire underside of the tongue.

For the condition numbers of the RA value with respect to A, L we obtain

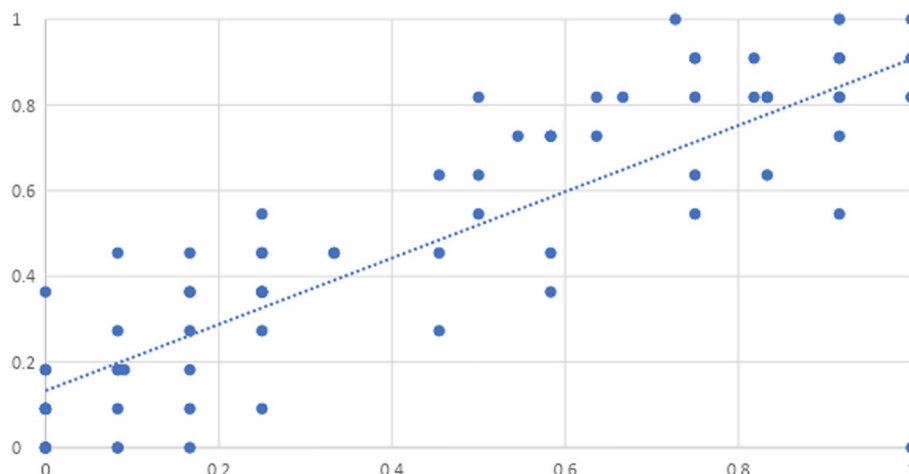


Fig. 5 Special test–retest correlation for estimating reliability in the parallel measurement model. The correlation estimates the reliability of the measurement of SV. This limits the maximum possible linear correlation of SV with an arbitrary other parameter Y

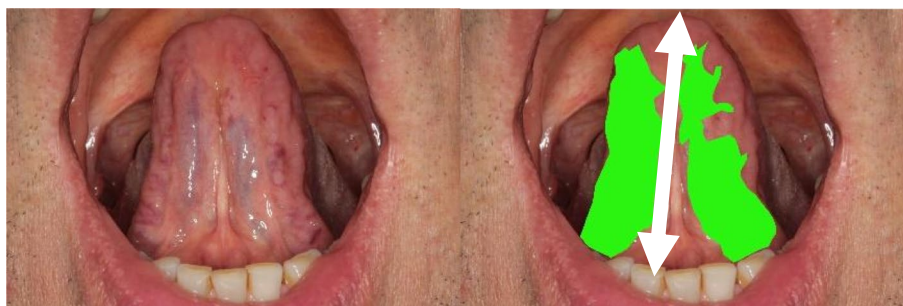


Fig. 6 Exemplary representation of the coloured tongue area for calculating the RA score. Written consent obtained from patient. The area of the sublingual veins (*Vv. linguae profundae*) or their dilatation is measured (e.g. by colouring with GIMP) in px² and then normalised to the square of the tongue length (along the *frenulum linguae*): $RA := \frac{A}{L^2}$

$$\varphi_A := \frac{A}{RA} \cdot \frac{\partial RA}{\partial A}(A, L) = \frac{A \cdot L^2}{A} \cdot \frac{1}{L^2} = 1$$

$$\varphi_L := \frac{L}{RA} \frac{\partial RA}{\partial L}(A, L) = \frac{L^3}{A} (-2) \frac{A}{L^3} = -2$$

Due to small absolute value of the condition numbers, the RA score is well conditioned. This means that if the input parameters A, L are disturbed, no excessive error amplification is to be expected from the calculation itself. In this sense, the RA score is a useful measure from a numerical point of view.

Figure 7 shows the ROC curve for the RA score based on SV_p with AUC=72.5%. Since other parameters besides the 2D projection of vein wall dilatation are taken into account in the clinical SV inspection, no 100% discriminatory accuracy can be expected.

With an AUC of 72.5%, the RA shows a good discriminatory power regarding SV_p. Since the RA score only

takes into account the 2D projection of vein wall dilatation and not other factors such as topography or colour distribution of the sublingual veins, it can be concluded that the estimation of SV_p by medical experts is similarly influenced more by geometric impressions than by others that the RA score cannot measure. Likewise, the RA score can also be used to evaluate automated SV segmentation.

Discussion

SV are a widely studied area of oral medicine. Many previous studies have diagnosed SV on a binary scale (0/1) with relatively few observers [1, 2, 4–6, 8, 14]. We systematically investigated the influence of the SV inspection reliability for the first time and found that SV inspection is possible in principle, but its inspection reliability is relatively small. SV diagnosis in clinical inspection has greater heterogeneity than previously thought. This finding may be important for the design of future SV studies, as the inclusion of a larger number of experts could

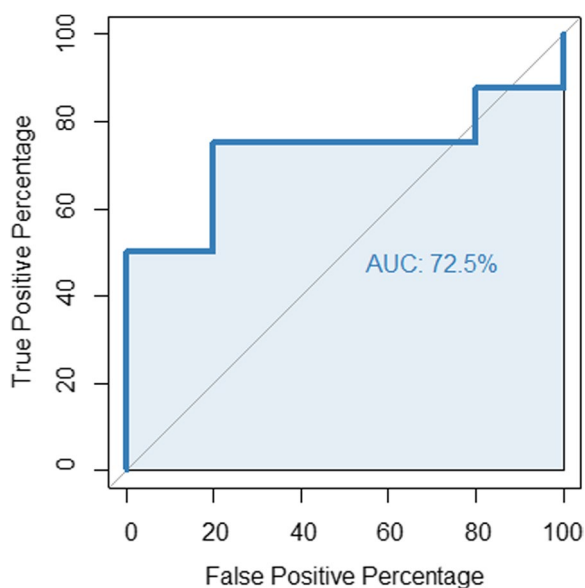


Fig. 7 ROC (receiver operating characteristic) curve for the RA score based on SV_p . The AUC is 72.5%

better represent the SV inspection and thus increase the reproducibility of the SV inspection.

The SV inspection reliability also limits a priori the maximum possible correlation SV with other clinical parameters. It can be seen that the maximum possible correlation with any other parameter Y can in principle never reach 100% due to the measurement uncertainty of SV. Reliability is therefore an essential quality feature for SV as a predictive marker. The design of future studies on SV as a predictive marker can thus benefit from controlling for inspection reliability.

The limited reliability of SV inspection should also be taken into account when interpreting previous studies. Previous work did not calculate a linear correlation when analysing SV as a predictive marker for other metric parameters (such as arterial blood pressure levels), as this only exists for metric variables. Methodologically, we defined a metric variable by introducing the surrogate parameter SV_p , which allows a (linear) correlation of the SV inspection with other metric parameters. This approach may also be helpful for further diagnostic studies on SV as a marker for metric outcome parameters.

To increase reliability and objectivity of SV inspection, we propose the RA score. This relates the 2D projection of the sublingual veins A to the square of the tongue length L. The RA score is a largely examiner-independent measure with a high reliability of A and L. In this way, it can be used as a point of comparison and normalisation of different clinical inspection findings in the context of SV diagnostics. The metric range of values maps

the continuity of the anatomical variance SV more accurately than a binary 0/1 classification and enables many statistical applications such as (linear) correlation. From an algorithmic point of view, the RA score is well conditioned and could therefore be used in automated SV segmentation [20–22] in addition to clinical SV studies. As there is no gold standard for defining and measuring SV, the evaluation of the RA score can only be a relative one. The ROC curve (Fig. 7) was determined on the SV_p as surrogate parameter and shows a relatively good prediction with an AUC=72%. Regardless of this, the RA score must prove itself as a measure in practice.

A probable limitation of this work is the sample size of $n=78$ patients in the clinical study. It would be interesting to apply the presented methods for inspection reliability to a larger cohort. By using a frequentist concept of SV (SV_p), it would be possible to obtain a more precise prevalence estimate of SV and a more stable estimate of the maximum possible correlation r_{max} (from the reliability estimate R). In addition, RA score calculation could be automated using machine learning methods, as has already been established for binary (0/1) SV classification [23].

Conclusions

For the first time, the inspection reliability SV has been systematically investigated. It turned out that SV inspection is possible in principle, but the inspection results were not always stably reproducible across different medical experts. This limits the maximum possible correlation of SV with other parameters, a correlation of 100% does not seem to be achievable a priori. The reliability of SV inspection is important for the quality of SV as a predictive marker. This should be taken into account when interpreting previous studies on SV and has implications for future studies. The RA score could help to objectify the SV examination and thus increase its reliability. To optimise the accuracy of SV diagnosis in clinical practice, we recommend calculating the RA score in systematic examinations of SV rather than using a subjective examiner-dependent assessment as the basis for diagnosis. In this way, the quality of examiners’ findings can be compared. In addition, the use of the RA score in clinical practice allows clinicians who do not routinely perform oral examinations to include the diagnosis of SV in their assessments and diagnoses. In this way, the RA score opens up the possibility of including the diagnosis of SV in the diagnosis of other medical specialties outside of oral medicine.

Abbreviations

- Px Pixels
- R Reliability

ROC	Receiver operating characteristics
SV	Sublingual varices
SV _p	SV subject prevalence

Acknowledgements

C.K., D.S. and H.S. thank Al Dente Uni-Bonn e.V. for their financial support of the study. All authors thank all medical professionals who participated in the clinical inspection experiment. The authors thank Richard Vock for making the yesnomabye software freely available to conduct the inspection experiment.

Authors' contributions

Idea, concept, study design: C.K., D.S. funding acquisition: D.S., H.S., C.K. data collection: D.S., J.G., C.K. data curation, statistical analysis, visualisation: C.K. clinical supervision: H.S. statistical supervision: A.M. writing- original draft: C.K. writing: review & editing: C.K., D.S., H.S., A.M., J.G. The author(s) read and approved the final manuscript.

Funding

Open Access funding enabled and organized by Projekt DEAL. This study was supported by Al Dente Uni-Bonn e.V.

Availability of data and materials

The data generated in this study are not publicly available due to patient data privacy but are available upon reasonable request from the corresponding author. The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Declarations

Ethics approval and consent to participate

Written informed consent was obtained from all participants and the study was approved by the Ethics Committee of the Medical Faculty of the University of Bonn (identification number: 296/19). All data were anonymised in accordance with EU Regulation 2016/679 (General Data Protection Regulation). All experiments were conducted in accordance with the Declaration of Helsinki.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interest.

Author details

¹Department of Prosthetic Dentistry, Preclinical Education and Materials Science, Dental School, Rheinische Friedrich-Wilhelms University, University of Bonn, Welschnonnenstr. 17, 53111 Bonn, Germany. ²Department of Oral, Maxillofacial and Plastic Surgery, Center of Dento-Maxillo-Facial Medicine, University Hospital Bonn, Welschnonnenstr. 17, 53111 Bonn, Germany. ³Department of Medical Biometry, Informatics and Epidemiology, University Hospital Bonn, Venusberg-Campus 1, 53127 Bonn, Germany.

Received: 9 March 2023 Accepted: 21 May 2023

Published online: 06 June 2023

References

- Hedström L, Albrektsson M, Bergh H. Is there a connection between sublingual varices and hypertension? *BMC Oral Health*. 2015;15:78. <https://doi.org/10.1186/s12903-015-0054-2>.
- Akkaya N, Ölmez D, Özkan G. Evaluation of the factors associated with sublingual varices: a descriptive clinical study. *Folia Morphol*. 2019;78:6.
- Lazos J, Rodríguez Marco E, Brunotto M, Panico R, Piemonte E. Oral varicose veins: Clinical features and its association with medical conditions. *J Oral Maxillofacial Surg Med Pathol* 2019:S2212555819302030. <https://doi.org/10.1016/j.ajoms.2019.11.007>
- Hedström L, Bergh H. Sublingual varices in relation to smoking and cardiovascular diseases. *Br J Oral Maxillofac Surg*. 2010;48:136–8. <https://doi.org/10.1016/j.bjoms.2009.05.005>.

- Al-Shayyab M, Baqain Z. Sublingual varices in relation to smoking, cardiovascular diseases, denture wearing, and consuming vitamin rich foods. *Saudi Med J*. 2015;36:310–5. <https://doi.org/10.15537/smj.2015.3.10429>.
- Bergh H, Albrektsson M, Kastberg C, Baigi A, Hedström L. The association between sublingual varices and cardiovascular risk factors. *VHRM*. 2022;18:319–27. <https://doi.org/10.2147/VHRM.S354021>.
- Barzideh N, Alaaee A, Azizi A. The relationship between smoking and sublingual varices in the elderly. *Oman Med J*. 2021;36:e288–e288. <https://doi.org/10.5001/omj.2021.94>.
- Accardo A, Pascazio L, Costantinides F, Gorza F, Silveri G. Influence of hypertension and other risk factors on the onset of sublingual varices. *BMC Oral Health*. 2021;21:235. <https://doi.org/10.1186/s12903-021-01604-1>.
- Eguchi M, Shigematsu H, Oku Y, Kikuchi K, Okada M, Sakashita H. A case of a large thrombosed lingual varix. *J Oral Maxillofacial Surg Med Pathol*. 2019;31:180–4. <https://doi.org/10.1016/j.ajoms.2018.11.003>.
- Tomita R, Kitamura N, Yoshizawa Y, Sasabe E, Kudo Y, Yamamoto T. A case of large varix including partially organizing thrombosis on the oral floor. *J Oral Maxillofacial Surg Med Pathol*. 2020;32:313–5. <https://doi.org/10.1016/j.ajoms.2020.03.003>.
- Duarte NT, de Oliveira GA, da Rocha TJ, Andrade NS, Franco JB, Pérez-Sayáns M, et al. Prevalence of sublingual varices in patients with cirrhosis and the correlation with nitrogen compounds. *Oral Surg Oral Med Oral Pathol Oral Radiol*. 2020;129:39–44. <https://doi.org/10.1016/j.oooo.2019.09.009>.
- González-Álvarez L, García-Pola Mj. Risk factors associated with tongue lesions: a propensity score-matched case-control study. *Med Oral*. 2022;27:e25–34. <https://doi.org/10.4317/medoral.24836>.
- Lazos JP, Piemonte ED, Panico RL. Oral varix: a review. *Gerodontology*. 2015;32:82–9. <https://doi.org/10.1111/ger.12074>.
- Bergh H, Kastberg C, Albrektsson M, Hedström L. Persistence and stability of sublingual varices over time and their connection to underlying factors: an 8 year follow up study. *BMC Oral Health*. 2022;22:346. <https://doi.org/10.1186/s12903-022-02379-9>.
- Jafari A, Alaaee A, Rezaei M, Masoudi M. Evaluation of sublingual varices prevalence and its respective factors in two Iranian nursing homes in 2019. *Iran J Otorhinolaryngol*. 2022;34:165. <https://doi.org/10.22038/ijorl.2022.63364.3170>.
- Reynolds CR, Altmann RA, Allen DN. *Mastering Modern Psychological Testing: Theory and Methods*. Cham: Springer International Publishing; 2021. <https://doi.org/10.1007/978-3-030-59455-8>.
- Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika*. 1951;16:297–334. <https://doi.org/10.1007/BF02310555>.
- Fleiss JL. Measuring nominal scale agreement among many raters. *Psychol Bull*. 1971;76:378–82. <https://doi.org/10.1037/h0031619>.
- Quarteroni, A., Sacco, R., Saleri, F. (2006). Principles of Numerical Mathematics. In: Numerical Mathematics. Texts in Applied Mathematics, vol 37. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-49809-4_2 n.d.
- Chiu C-C, Lan C-Y, Chang Y-H. Objective assessment of blood stasis using computerized inspection of sublingual veins. *Comput Methods Progr Biomed*. 2002;69:1–12. [https://doi.org/10.1016/S0169-2607\(01\)00181-X](https://doi.org/10.1016/S0169-2607(01)00181-X).
- Yan Z, Wang K, Li N. Segmentation of sublingual veins from near infrared sublingual images. 2008 8th IEEE International Conference on Bioinformatics and BioEngineering, Athens, Greece: IEEE; 2008, p. 1–5. <https://doi.org/10.1109/BIBE.2008.4696791>.
- Lin H, Chen Y-J, Damdinsuren N, Tan T-H, Liu T-Y, Chiang JY. Automatic Sublingual Vein Feature Extraction System. 2014 International Conference on Medical Biometrics, China: IEEE; 2014, p. 55–62. <https://doi.org/10.1109/ICMB.2014.17>.
- Lu P-H, Chiang C-C, Yu W-H, Yu M-C, Hwang F-N. Machine learning-based technique for the severity classification of sublingual varices according to traditional Chinese medicine. *Comput Math Methods Med*. 2022;2022:1–10. <https://doi.org/10.1155/2022/3545712>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

3. Danksagung

Mein erster Dank gilt meinem Doktorvater Prof. Dr. Helmut Stark für die Betreuung meiner Promotion. Die stetige Unterstützung meiner Ideen, die Implementierung der diagnostischen Studie an seiner Klinik und die mentorielle Supervision in allen Fragen, weiß ich sehr zu schätzen.

Al Dente Uni-Bonn e.V. danke ich sehr für die Sachmittelförderung der klinischen Studie.

Für die statistische Betreuung und Supervision meiner Dissertation danke ich Herrn Prof. Dr. Andreas Mayr.

Meinem guten Freund Dr. David Stoppenbrink danke ich für unsere gemeinsame Begeisterung an Human- und Zahnmedizin, die Motivation für ein gemeinsames wissenschaftliches Projekt und unsere Freundschaft.

Meinem Freund, ehemaligem Kommilitonen und Kollegen Dr. Jannik Geier danke ich für seine Unterstützung bei der Datenerhebung im Rahmen der klinischen Studie.

Mein größter Dank gilt meiner Familie. Diese Arbeit ist ein Ergebnis ihrer größtmöglichen Förderung. Ich danke ihr von Herzen für ihre bedingungslose Unterstützung von Anfang an.