



Concerning Assumptions in Laryngology

Jan Wouter Brunings

Concerning Assumptions in Laryngology

Copyright © Jan Wouter Brunings 2020

All right reserved. No part of this thesis may be reproduced or distributed in any form or by any means, without the prior written permission of the author or the publisher.

Cover: Jan Wouter Brunings, own vocal folds

Layout: Tiny Wouters

Printing: www.proefschriftmaken.nl

ISBN/EAN: 978-94-6380-989-4

Printing of this thesis was financially supported by: ALK-Abelló BV, Allergy Therapeutics Netherlands BV, DosMedical BV, Meditop Medical Products BV, Pentax Nederland BV, Soluvos Medical BV, Xion GmbH and Stöpler Medical BV.

Concerning Assumptions in Laryngology

PROEFSCHRIFT

ter verkrijging van
de graad van doctor aan de Universiteit Maastricht,
op gezag van de Rector Magnificus, Prof. dr. Rianne M. Letschert,
volgens het besluit van het College van Decanen,
in het openbaar te verdedigen
op donderdag 10 december 2020 om 16.00 uur

door

Jan Wouter Brunings

Geboren op 19 november 1973 te Gorssel

Promotor

Prof Dr B. Kremer

Copromotoren

Dr L.W.J. Baijens

Dr A.E.W. Hamaekers

Beoordelingscommissie

Prof Dr W.F.F.A. Buhre (voorzitter)

Prof Dr F.G. Dikkers (Amsterdam Universitaire Centra)

Dr E.V. Sjögren (Leiden Universitair Medisch Centrum)

Prof Dr E.J.M. Speel

Table of contents

	Prologue	7
Chapter 1	General Introduction and Outline of the Dissertation	11
PART I	Assessment of Voice and Vocal Function	21
Chapter 2	Maximum Phonation Time: Variability and Reliability <i>Journal of Voice. 2010 Oct;24(3):281-4</i>	23
Chapter 3	Observer Agreement for Measurements in Videolaryngostroboscopy <i>Journal of Voice. 2018 Nov;32(6):756-762</i>	35
Chapter 4	The Expression of Oestrogen and Progesterone Receptors in the Human Larynx <i>Journal of Voice. 2013 May;27(3):376-80</i>	51
PART II	Perioperative Voice and Vocal Fold Situation	63
Chapter 5	Adverse Laryngeal Effects Following Short-term General Anaesthesia: a Systematic Review <i>Arch Otolaryngol Head Neck Surg. 2012;138(3):257-264</i>	65
Chapter 6	Voice and Vocal Fold Condition Following Short-term General Anaesthesia: a Prospective Study <i>Journal of Voice. 2020 Jan; Epub ahead of print</i>	85
PART III	General Discussion, Valorisation Addendum and Summary	105
Chapter 7	General Discussion	107
	Valorisation Addendum	115
	Summary	121
	Dutch Summary (Samenvatting)	127
	Epilogue (Een terugblik)	133
	List of Publications	141
	Curriculum Vitae	147

Prologue

Assumption is the mother of all... mistakes.

Every day, every hour, and everywhere our lives are filled with assumptions:

"I thought you would do that for me." "I take it for granted that the appointment was made."

"They will probably wait for me." "I assume the doctor will call me with the test results."

A contradiction in the assumptions above will lead to frustration, misunderstanding, quarrel, and disappointment. If we do not have complete, correct, and timely information at the moment we need it, assumptions fill the gaps.

Not only in normal daily life, but also in healthcare the same applies, as well as in the art of medicine. We take as "unconditional truth" many things that our supervisors have told and taught us...and even this last sentence is an assumption.

"More babies are born during a full moon." "Red-haired people have an increased tendency to bleed." "Scar tissue tickles when the weather changes."

These quotes sound more like idle gossip, or an old farmer's wisdom, but they are still repeated in daily healthcare practice.

To embrace or reject these assumptions, evidence is needed. The Evidence-Based Medicine (EBM) phenomenon has been introduced since the early 90s of the last century. This was not a novelty.

Lord Byron [William Thomson, 1st Baron Lord Kelvin of Largs 1824–1907] said:

[Quote] *To measure is to know: When you can measure what you are speaking about, and express it in numbers, you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind; it may be the beginning of knowledge, but you have scarcely, in your thoughts, advanced to the stage of science, whatever the matter may be.*¹

[Unquote]

Centuries earlier, Roman and Greek physicians, of whom Hippocrates and Galenus are well known, have already conducted scientific research to improve health care.

Research is needed to find answers and to tackle assumptions. Whether the study is observational, randomised controlled, etc. is another dispute.²



The only limits that prevent us from investigating an assumption or hypothesis are of pure physical, ethical or perhaps religious nature.

So, the birth-rate is not related to the lunar cycle,³ and there's still no evidence of an increased tendency to bleed in red-haired individuals.⁴ There are no hits in PubMed for 'scar' and 'weather', so that assumption has yet to be resolved.

Laryngology also has its assumptions and beliefs. This dissertation deals with the following assumptions:

Intubation tubes damage the vocal folds; The vocal folds and the voice are influenced by sex hormones; All laryngologists have the same level of observer agreement when assessing videolaryngostroboscopy; the Maximum Phonation Time is a reliable measurement tool.



References

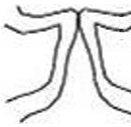
- 1 Kelvin WT. Lecture on "Electrical Units of Measurement" Popular Lectures and Addresses volume 1. London: Macmillan & Co, 1889
- 2 Sehon SR, Stanley DE. A philosophical analysis of the evidence-based medicine debate. BMC Health Serv Res 2003; 3: 14
- 3 Staboulidou I, Soergel P, Vaske B, Hillemanns P. The influence of lunar cycle on frequency of birth, birth complications, neonatal outcome and the gender: a retrospective analysis. Acta Obstet Gynecol Scand 2008; 87: 875-9
- 4 Leebeek FW, Dehghan A, Kruip MJ, Hofman A, Uitterlinden AG, EM DEW, Witteman JC, MP DEM. The presumed increased bleeding tendency in red-haired individuals is not associated with von Willebrand factor antigen levels in older individuals. J Thromb Haemost 2011; 9: 2509-11





Chapter 1

General Introduction and Outline of the Dissertation





General Introduction

The voice

Since the moment a human being is born, the vocal folds are used to express oneself, from the first cry until the last spoken words of comfort on someone's deathbed.

Talking, singing, crying, yelling, laughing, screaming, and moaning are all produced by the vocal folds. Other sounds generated in the same body part are unvoiced consonants, clicks, whistling, and whispering.

The mechanism to vocalise can be divided into three parts;

Firstly, the lungs, like a bellows, produce airflow and pressure underneath the vocal folds to let them vibrate. Secondly, the vocal folds are the true instrument (Figure 1.1); by adducting and vibrating on the passing airflow, they create audible pulses thus constituting the laryngeal sound source.

Lastly, the articulators in the mouth, oropharynx, nasopharynx, and nasal cavities play their part as a sound box or resonance box, amplifying or weakening the sound source.

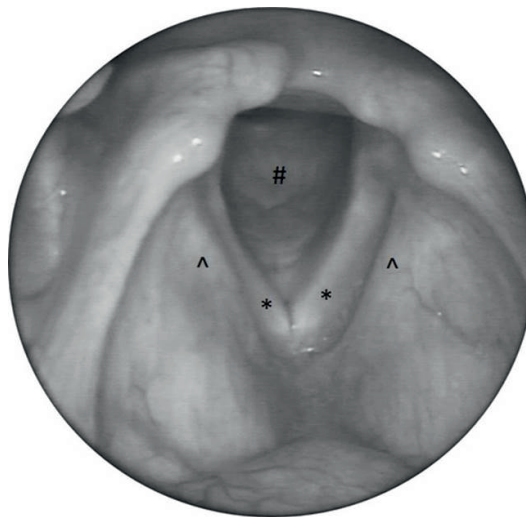


Figure 1.1 Endoscopic view of the larynx.

* - True vocal folds, plicae vocales; ^ - False vocal folds, plicae ventriculares; # - Trachea

The complex system of mucosa, muscles, and soft tissues adjusts the vocal folds in length and tension to provide tune, pitch and tone.

The difference between a normal adult male and female voice is usually clearly audible. This can be explained by the sexual dimorphism in the postpubescent larynx by a generalised increase in all dimensions of the male larynx.¹

The alterations of the male and female voice during life are thought to be influenced by the levels of sex-hormones.¹⁻⁴ The change of the children's voice towards an adolescent voice is called the mutation. In boys, this vocal change is mainly due to growth of the laryngeal cartilage, or framework, during the sex maturation influenced by androgen, the male sex hormone. Women are particularly affected in their vocal performance by the hormonal changes, such as menstruation, pregnancy, and menopause. Over three-quarters of all female singers report complaints before or during menstruation, usually described as abnormal sensations, mucilage, weak breathing, and/or poor supportive function.^{5,6}

It is understandable that professional voice users such as singers, theatre players, but also teachers and other professionals using their speaking voice for hours daily, care about this delicate vocal instrument, especially when in need for an operation requiring airway management.

In the Netherlands more than 800 operations per 10.000 inhabitants are performed yearly, so daily, thousands of people worldwide are intubated to undergo elective or emergency surgery.⁷

Intubation devices and techniques

The first endotracheal intubation was probably performed by the Belgian anatomist Andreas Vesalius around 1542 via a thoracotomy; it was not until two centuries later that in 1754 an 'air pipe' was designed by Benjamin Pugh, a Chelmsford surgeon, that was introduced transorally into the trachea, which was done without any view on the larynx.⁸ It was not until 1854 when Manuel Garcia, a Spanish vocal pedagogue, was the first to have a view on the larynx of a living person.⁹ From that time, scientists and physicians started to develop techniques and devices to secure the airway. Periods of war (both WWI and WWII) produced the most inventions. Sir Robert Reynolds Macintosh (1897-1989) introduced the curved laryngoscope blade in 1943, which is still the most widely used laryngoscope blade.¹⁰ (Figure 1.2). Since then orotracheal tubes were developed, nowadays known as endotracheal tubes (ETT). (Figure 1.3)

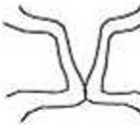




Figure 1.2 Macintosh laryngoscope.



Figure 1.3 Endotracheal Tube (ETT).



Nowadays, endotracheal intubation is one of the most widely accepted techniques in anaesthetic practice, though it is not without complications, mostly due to the need to visualise and penetrate the laryngeal opening, and thereby potentially damaging the vocal folds, both during placement, but also during prolonged stay.¹¹⁻¹³

Besides the endotracheal intubation, face mask ventilation can be performed too, when the vision on the larynx is difficult and intubation is at risk, but also when a procedure takes only a few moments. In those cases the hands of the anaesthesiologist are bound to the patient as the ventilation mask is held into place manually. This issue was the primary reason for Archie Brain to invent and develop the first supraglottic airway (SGA) in 1981, at that time called a Laryngeal Mask Airway (LMA).^{14,15} (Figure 1.4). The LMA, or SGA, is a device covering the laryngeal aperture. The laryngeal mask offers some advantages over endotracheal intubation since the vocal cords need be neither visualised nor forced apart.

Further engineering underlined the value of the SGA compared to the ETT in terms of gas-flow; the thought of two tubes with different diameter, the trachea and the plastic tube, being connected with a gas-tight sealing is rather unsatisfactory. This thought, and the risk of laryngeal or tracheal complications during placement, led to the further development of an SGA device in which both “tubes”, with the same diameter, connect end-to-end.¹⁴



Figure 1.4 Supraglottic Airway Device (SGA), Laryngeal Mask Airway (LMA).



Assessment of voice

For the assessment of voice or pathology of the voice, the vocal folds and the supporting systems such as the lungs and sound box can be evaluated.

According to the European Laryngological Society (ELS), the following measurements for voice evaluation need to be considered: auditory-perceptual rating, videolaryngostroboscopy, acoustic analysis, aerodynamic analysis, and subjective rating by the patient.¹⁶

Auditory-perceptual rating

For auditory-perceptual rating the GRB-scale is used. This scale, derived from the GRBAS scale as proposed by Hirano¹⁷, scores for grade (G), the quality of the voice on the whole, roughness (R), and breathiness (B) and is performed by a trained health care professional, usually a speech therapist.

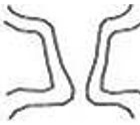
Videolaryngostroboscopy

Videolaryngostroboscopy is until now the gold standard assessment tool for clinical evaluation of the vocal folds. A rigid transoral or flexible transnasal endoscope, attached to a stroboscopic device, is used to make video recordings of vocal fold

vibration during repeated stable phonation of a sustained vowel /i:/ at comfortable pitch and loudness. During the examination, subjects are seated upright. The field of the videolaryngostroboscopic image includes the laryngeal vestibule, vocal folds, anterior commissure, and the arytenoids. According to the guidelines of the ELS, visuoperceptual variables can be scored such as glottal closure, regularity, mucosal wave pattern, and symmetry.^{16,18} The videolaryngostroboscopy is mainly performed and rated by a physician, usually a laryngologist, phoniatician, or general otorhinolaryngologist.

Acoustics

Acoustical analysis of the voice concerns the following variables; the percentage of jitter gives an indication of the variability of the pitch period within the analysed voice sample. It represents the relative period-to-period variability, so the variability in frequency. The percentage of shimmer gives an indication of the period-to-period variability of the peak-to-peak amplitude, so the variability in loudness. The evaluation of the noise present in the signal is expressed as the noise-to-harmonic ratio (NHR): the average ratio of the energy of the harmonic components in the 1500-4500 Hz range to the harmonic components energy in the 70-4500 Hz range. It is performed by attaching a condenser microphone to a recorder. It is placed in a harmonica holder at a mouth-to-mouth distance of approximately five centimetres and an angle of approximately 45 degrees. The subject is asked to produce three sustained vowels /a:/ of several seconds duration on a comfortable pitch and loudness level. The speech samples are recorded and analysed digitally using a computerised Speech Lab Model.¹⁹ The three acoustic parameters are computed for each sample.



Aerodynamics

The most used and simplest aerodynamic parameter of voicing is the maximum phonation time (MPT).¹⁷ Dysphonic people have shorter MPT compared with healthy controls.

Subjective self-report rating

The Voice Handicap Index (VHI) is a quality of life questionnaire. The VHI-30 (Dutch version) is a validated self-perception 30-item questionnaire about the emotional, physical and psychological effects of dysphonia scored on a five grade scale (0-4) with a maximum score of 120. The higher the score, the higher the degree of patient-perceived vocal handicap.²⁰

Voice disorders and vocal fold pathology

The pathophysiology of dysphonia is complex. The Phonosurgery Committee of the ELS proposed a division in two major groups: *vocal fold lesions impairing the vibratory movement* and *vocal fold movement disorders*.²¹ The latter is divided into vocal fold immobility, neurologic movement disorders, and dysfunctional disorders without somatic cause.

Vocal fold lesions impairing the vibratory movement can be divided into lesions at the different layers and parts of the vocal fold: epithelium, lamina propria, arytenoid, and anterior commissure.

Figure 1.5 illustrates the five different layers of the vocal fold. According to Hirano, the vocal fold structures act more like three layers: a cover (epithelium and Reinke's space), a transition (intermediate and deep layer), and the body (the vocalis muscle).¹⁷

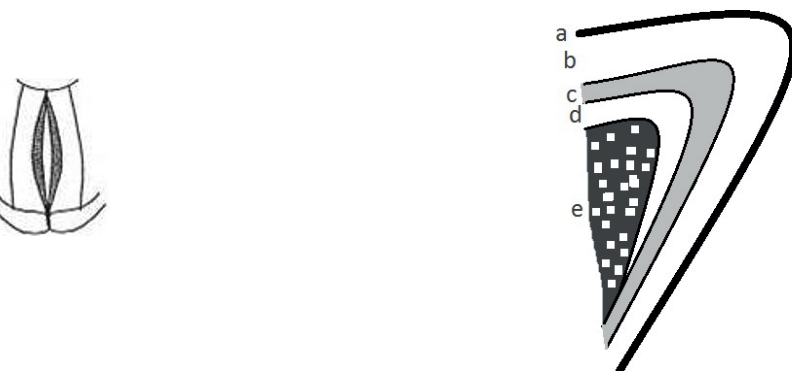


Figure 1.5 Schematic frontal section of the larynx according to Hirano.¹⁷

a – epithelium; b – superficial layer of lamina propria, Reinke's space; c – intermediate layer of lamina propria; d – deep layer of lamina propria, vocal cord/vocal ligament; e – vocalis muscle.

Epithelial diseases impairing the vocal fold movement can be e.g. papillomatosis, laryngitis, or carcinoma. Diseases affecting the lamina propria are e.g. nodules, polyps, cysts, or mucosal atrophy.

Also, iatrogenic trauma can occur during e.g. emergency intubation, leading to lacerations, cricoarytenoid subluxation, and oedema.

Outline of the Dissertation

Assumptions and suppositions in laryngology triggered the research presented in this dissertation on the following subjects:

- The maximum phonation time is a reliable instrument
- The vocal folds, and the voice, are influenced by sex hormones
- All laryngologist have the same observation judging videolaryngostroboscopy
- Intubation tubes damage the vocal folds

Maximum phonation time is a frequently used measure in voice assessment, mainly because it's a quick, cheap and non-invasive assessment method. **Chapter 2** shows the results of a reliability study in dysphonic subjects compared with control subjects.

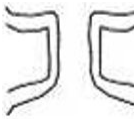
Clinicians around the world apply videolaryngostroboscopy in daily clinical practice, using many variables to describe vocal fold functioning and the vibratory pattern, although data on the reliability and validity of these variables remain scarce. The study in **chapter 3** evaluates the levels of intraobserver and interobserver agreement for measurements of visuoperceptual variables in videolaryngostroboscopic examinations and compared the observers' behaviour during independent versus consensus panel rating.

The human female voice changes in quality during the menstrual cycle, pregnancy, and menopause. The underlying pathophysiological mechanisms are not clearly known yet. The aim of the study in **chapter 4** therefore, was to evaluate the existence of oestrogen receptors (ERs) and progesterone receptors (PRs) in the female human vocal fold.

The widespread belief that endotracheal intubation leads to postoperative hoarseness was addressed in **chapter 5**. A systematic review was conducted to determine the occurrence and type of vocal cord injury, as well as the occurrence of hoarseness in adults using an endotracheal tube or laryngeal mask (supraglottic airway) during routine short term anaesthetic care.

Following the findings in chapter 5, a prospective study was set up to obtain the prevalence of vocal fold injury following short-term routine anaesthesia using an endotracheal tube (ETT) or supraglottic airway (SGA) in adults, which is described in **chapter 6**.

The main findings and conclusions of this dissertation are presented in **chapter 7**, which also outlines some perspectives for the future.



References

- 1 Kahane JC. Growth of the human prepubertal and pubertal larynx. *J Speech Hear Res* 1982; 25: 446-55
- 2 Abitbol J, Abitbol P, Abitbol B. Sex hormones and the female voice. *J Voice* 1999; 13: 424-46
- 3 Beckford NS, Rood SR, Schaid D, Schanbacher B. Androgen stimulation and laryngeal development. *Ann Otol Rhinol Laryngol* 1985; 94: 634-40
- 4 Kahane JC. Connective tissue changes in the larynx and their effects on voice. *J Voice* 1987; 1: 27-30
- 5 Flach M, Schwickardi H, Simon R. [What effect does menstruation and pregnancy have on the trained singing voice?]. *Folia Phoniatri (Basel)* 1969; 21: 199-210
- 6 Lacina O. [The influence of menstruation on the voice of female singers. Premenstrual laryngopathy]. *Folia Phoniatri (Basel)* 1968; 20: 13-24
- 7 Statistiek CBvd. Operaties in het ziekenhuis; soort opname, leeftijd en geslacht, 1995-2010 CBS, 2014
- 8 White GM. Evolution of endotracheal and endobronchial intubation. *Br J Anaesth* 1960; 32: 235-46
- 9 Garcia M. Physiological observations on the human voice *Proceedings of the Royal Society London* 1855; 7: 339
- 10 Macintosh RR. A NEW LARYNGOSCOPE. *Lancet* 1943; 241: 205
- 11 Kambic V, Radsel Z. Intubation lesions of the larynx. *Br J Anaesth* 1978; 50: 587-90
- 12 Ellis SF, Pollak AC, Hanson DG, Jiang JJ. Videolaryngoscopic evaluation of laryngeal intubation injury: incidence and predictive factors. *Otolaryngol Head Neck Surg* 1996; 114: 729-31
- 13 Echternach M, Mencke T, Richter B, Reber A. [Laryngeal alterations following endotracheal intubation and use of larynx masks]. *HNO* 2011; 59: 485-98
- 14 Brain AI. The laryngeal mask—a new concept in airway management. *Br J Anaesth* 1983; 55: 801-5
- 15 van Zundert TC, Brimacombe JR, Ferson DZ, Bacon DR, Wilkinson DJ. Archie Brain: celebrating 30 years of development in laryngeal mask airways. *Anaesthesia* 2012; 67: 1375-85
- 16 Dejonckere PH, 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 Phoniatics of the European Laryngological Society (ELS). *Eur Arch Otorhinolaryngol* 2001; 258: 77-82
- 17 Hirano M. Clinical examination of voice.: Springer, New York, 1981
- 18 Hirano M, Bless DM. Videostroboscopic examination of the larynx San Diego, California Singular Publishing Group Inc., 1993
- 19 Titze I. Workshop on acoustic voice analysis: summary statement.: National Center for Voice and Speech, The University of Iowa, Iowa City, 1995
- 20 Hakkesteegt MM, Wieringa MH, Gerritsma EJ, Feenstra L. Reproducibility of the Dutch version of the Voice Handicap Index. *Folia Phoniatri Logop* 2006; 58: 132-8
- 21 Remacle M, Friedrich G, Dikkers FG, de Jong F. Phonosurgery of the vocal folds: a classification proposal. *Eur Arch Otorhinolaryngol* 2003; 260: 1-6



Chapter 2

Maximum Phonation Time: Variability and Reliability



Renée Speyer, Hans CA Bogaardt, Valéria Lima Passos, Nel PHD Roodenburg,
Anne Zumach, Mariëlle AM Heijnen, Laura WJ Baijens, Stijn JHM Fleskens,
Jan Wouter Brunings

Journal of Voice. 2010 Oct;24(3):281-4

Abstract

Objectives

The objective of the study was to determine maximum phonation time reliability as a function of the number of trials, days, and raters in dysphonic and control subjects.

Material and Methods

Two groups of adult subjects participated in this reliability study: A group of outpatients with functional or organic dysphonia versus a group of healthy control subjects matched by age and gender. Over a period of maximally 6 weeks, three video recordings were made of five subjects' maximum phonation time trials. A panel of five experts was responsible for all measurements, including a repeated measurement of the subjects' first recordings.

Results

Patients showed significantly shorter maximum phonation times compared with healthy controls (on average, 6.6 seconds shorter). The averaged interclass correlation coefficient (ICC) over all raters per trial for the first day was 0.998. The averaged reliability coefficient per rater and trial for repeated measurements of the first day's data was 0.997, indicating high intrarater reliability. The mean reliability coefficient per day for one trial was 0.939. When using five trials, the reliability increased to 0.987. The reliability over five trials for a single day was 0.836; for 2 days, 0.911; and for 3 days, 0.935.

Conclusion

The maximum phonation time has proven to be a highly reliable measure in voice assessment. A single rater is sufficient to provide highly reliable measurements.



Introduction

Voice is a multidimensional phenomenon.¹ The multidimensional voice assessment proposed by the European Laryngological Society includes aerodynamic measurements.² Maximum phonation time is usually used for practical reasons: determination of maximum phonation time is a non-invasive, fast, and low-budget measurement. The ability to maximally sustain a vowel after having taken a maximal inspiration is considered an objective measure of the efficiency of the respiratory mechanism during phonation.³ In the literature, maximum phonation time has been used to objectify the degree of severity of dysphonia and to determine the effects of voice therapy.⁴ However, some authors question the usefulness of this measure as an evaluation tool in therapy, as no significant change could be demonstrated after therapy in a group of patients with vocal nodules (eg, Treole and Trudeau⁵).


Many studies have provided normative data on maximum phonation time in diverse subject populations (see for an overview, Baken and Orlikoff⁶). The variability is high, depending on subject characteristics, such as gender or age⁷ and testing conditions.^{3,8} In 1968, Hirano et al stated that although tests, such as maximum phonation time, indicate the degree of vocal function, no diagnosis of laryngeal disease could be established,⁹ an assertion that has since been confirmed by other authors.^{10,11} In a study on procedural aspects of eliciting maximum phonation time, Neiman and Edeson³ concluded that subjects should receive a complete verbal and visual model of the experimental task before its elicitation, instructions should be standardised, and at least three trials of maximum phonation time were required before performance could be expected to approximate the criterion. However, in a study by Lewis et al,⁷ the authors found that using three trials to determine maximum phonation time in children yielded inadequate data, as most subjects had not yet reached their maximum achievement after three trials. Using more trials, higher achievement levels were found. On the other hand, Finnegan¹² demonstrated the presence of fatigue effects after practice effects. Furthermore, Shanks and Mast¹³ considered the differential operation of fatigue to be partially supported by the progressive increase in standard deviations found when raising the number of maximum phonation trials. Thus, the outcomes of studies on fatigue versus practice effects while performing maximum phonation tasks are not quite consistent. Although quite a few studies have described maximum phonation time in diverse subject populations and under various testing conditions, limited information is available on the reliability of the data over time.⁷



To our knowledge, no study has thus far determined how many days the subjects should be repeatedly measured, nor have any authors indicated how many trials and raters would be necessary to obtain reliable maximum phonation time measurements using dysphonic subjects and control subjects matched by gender and age. The purpose of the present study is to determine the reliability of maximum phonation time as a function of the number of trials, days, and raters in dysphonic and control subjects.

Methods

Subjects



This study was conducted on two groups of adult subjects: patients with functional or organic dysphonia versus healthy subjects who did not suffer from any voice problems. The patients were diagnosed by a laryngologist at the Otorhinolaryngology Department of the University Hospital Maastricht. The same exclusion criteria were applied to both groups: pneumopathy, including chronic obstructive pulmonary disease (COPD) or atopic syndrome; upper respiratory infection; inhalation of corticosteroids; extensive laryngeal surgery; substitution voice after laryngectomy; neuromuscular diseases; or pregnancy. All subjects had to be older than 18 years to rule out bias by voice maturation or mutation, but had to be younger than 65 years. The patients had received no voice therapy for at least 3 months before inclusion. In total, 27 adult outpatients were included. Table 2.1 presents the frequency of the etiologic categories as diagnosed by a laryngologist using laryngostroboscopy. The group of patients consisted of 13 men and 14 women ranging in age from 18 to 63 years. The average age for the female participants was 32 years and for the male subjects, 52 years. A control group was matched by age and gender.

Procedure

Over a period of maximally 6 weeks, three digitised video recordings were made of individual subjects' performances (Mini-DV Camera-Recorder AG-DVC30; Panasonic, Matsushita, Electric Industrial Co., Osaka, Japan). During each recording, the subjects were asked to produce a sustained vowel /a:/ for as long as possible. The subjects were allowed five trials in a row with a 15-second break between each one. Before each recording, the subjects received verbal instructions according to a strict protocol in addition to a visible and audible trial performance by the same researcher.

Panel of expert listeners

The panel of expert listeners consisted of one laryngologist and four speech therapists. Each panel member received a complete set of all digitised video recordings in randomised order. The maximum phonation time for all patients per trial was determined using a stopwatch. By means of computers, the experts were allowed to listen to the stimuli as often as convenient in individual listening sessions. Three to 6 weeks after doing the initial rating, all experts received the randomised first recordings once more for a repeated rating.

Statistical analysis

Variance components analysis was performed for each group separately to determine the proportions of total maximum phonation time variability that can be attributed to its different sources, that is, trial, days, subjects, and raters. To estimate the variance components, a linear mixed model (proc mixed; SAS Institute, Inc., version 9.1) was fit, assuming a hierarchical structure of the data, whereby trials were nested within days and days were nested within patients (a three-level model, under the premise of no interaction among the explanatory variables). Interclass correlation coefficients (ICC) and Cronbach's alpha were used as measures of agreement (absolute) and internal consistency (SPSS Inc., Chicago, USA, version 11.5). Group differences between patients versus healthy controls were determined using a paired-samples t-test. The Spearman-Brown formula was applied to compute the minimum necessary number of trials, days, and raters to obtain reliable measurements.¹⁴

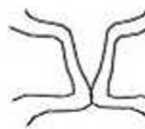


Table 2.1 Distribution of patients by diagnostic categories.

Phoniatric Diagnosis	n=27
Muscle tension dysphonia	9
Submucosal swelling	4
Vocal fold nodules	2
Vocal fold polyps	2
Unilateral vocal fold paralysis	1
Other: slight vocal fold abnormalities	7
Other: severe vocal fold abnormalities	2

Table 2.2 Variance components table for patients and controls.

Source	Estimated σ^2	SE	P	% Total
Patients				
Rater	0.07492	0.05416	0.0833	0.1333
Subject	45.4401	13.2675	0.0003	80.877
Day	6.3272	1.3677	<0.0001	11.2615
Position	3.6766	0.3022	<0.0001	6.5438
Residual	0.6651	0.02362	<0.0001	1.837
Total	56.1839			
Controls				
Rater	0.06671	0.04750	0.0801	0.0849
Subject	61.2618	18.1398	0.0004	78.0182
Day	11.0112	2.3589	<0.0001	14.0230
Position	6.0011	0.4818	<0.0001	7.6425
Residual	0.1816	0.006483	<0.0001	2.1308
Total	78.5224			

SE, standard error



Results

Variance components and reliability coefficients

The estimated variances of raters, subjects, days, and positions (trials) are displayed together with their *P* values in Table 2.2 for patients and controls, respectively. (The null hypothesis assumes the variance components to be zero.) Note that the variance attributed to “rater” did not significantly contribute to the total variances for maximum phonation time. In both groups, the subjects themselves are the predominant source of variation, accounting for approximately 80% of the total variance in maximum phonation time. The second significant source is attributed to “day” (11% and 14%), followed by “trials” (6.5% and 7.6%). Very little (about 2%) of the total variance remains unaccounted for. The control group showed more variance than the patient group (78.5 as against 56.1). The results for the variance components are largely confirmed by the reliability coefficients (ICCs under total agreement and Cronbach’s alpha) for raters, days, and trials, as shown in Table 2.3. The ICCs given for raters, for instance, are averages taken from 15 individual ICCs (for each of the five trials and each of the 3 days). Similarly, there were 25 ICCs for days (five trials and five raters) and 15 ICCs for trials (five raters and 3 days). Using a paired-samples *t*-test, a significant group effect was found for the matched subjects’ overall mean maximum phonation

time ($P=0.003$). The patients showed significantly shorter maximum phonation times than the healthy controls (on average, 6.6 seconds shorter).

Interrater reliability

The average ICC over all raters per trial for the first day was 0.998 (all ICCs \geq 0.997). ICCs were computed for repeated ratings of the initial data. The resulting average ICC per rater and per trial was 0.997 (all ICCs \geq 0.982), indicating highly reliable scoring by all raters (intrarater reliability).

Between-days reliability

The average ICC over 3 days per trial averaged over all raters was 0.829. Although still high, in perspective, the internal consistency of the repeated measurements over the days is lower than that of the raters. Cronbach's alpha was also determined for reliability analysis of one single day and used in the Spearman-Brown formula.¹⁴ Figure 2.1 presents the reliability change as a function of the number of days. The reliability over five trials for 1 day was 0.836; for 2 days, 0.911; and for 3 days, 0.935.

Trials reliability

Cronbach's alpha was determined for the reliability analysis of a single trial. Inserting this value in the Spearman-Brown formula,¹⁴ we calculated the reliability change as a function of an increased number of trials (Figure 2.2). The reliability per day for one trial varied between 0.930 and 0.943 (mean reliability per day 0.939). When using five trials, the reliability varied between 0.985 and 0.988 (mean reliability per day: 0.987).

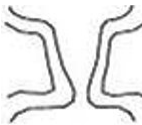


Table 2.3 Mean reliability coefficients for raters, days, and trials.

Factor	Mean ICC	95% Confidence Interval
Raters	0.994 [*]	0.990–0.997
Trials	0.936 ^y	0.932–0.941
Days	0.827 ^z	0.865–0.912

^{*} Mean over 15 individual ICCs (for each of five trials and each of 3 days); ^y Mean over 15 individual ICCs (for each of five raters and each of 3 days); ^z Mean over 25 individual ICCs (for each of five raters and each of five trials).

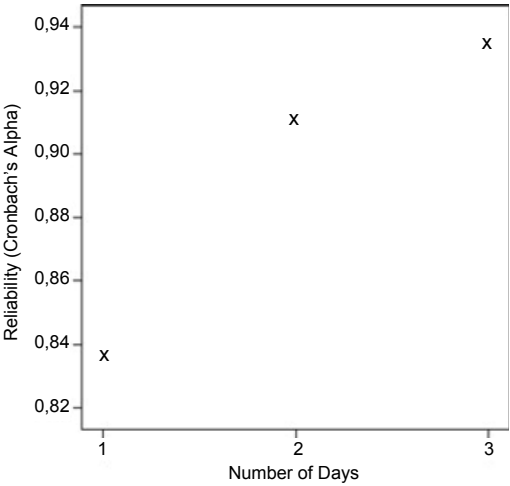


Figure 2.1 Reliability versus number of days. Reliability is presented (over five trials) as a function of the number of days (1–3 days).

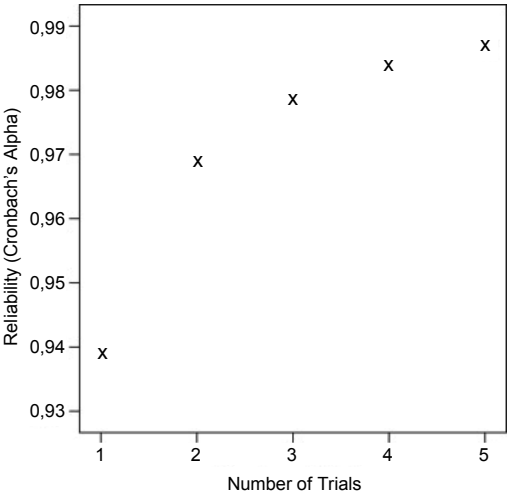


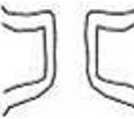
Figure 2.2 Reliability versus number of trials. Reliability is presented (mean reliability per day) as a function of the number of trials (one to five trials).

Discussion

Maximum phonation time is a frequently used measure in voice assessment. It owes its popularity mainly to the fact that determining the maximum phonation time is a cheap, quick, and non-invasive assessment method. Nonetheless, little has been done thus far to establish its reliability. As a result, practitioners generally lack guidelines on how the measurement should be carried out.

The objective of the present study was to determine the reliability of maximum phonation time in dysphonic and control subjects. To restrict the influence of subject characteristics, all patients and control subjects were matched by gender and age. Any difference between the dysphonic versus the control group was regarded as the result of dysphonia and not of any other subject characteristic. Testing conditions were standardised.³ All recordings were made by one single researcher according to a strict protocol. Subjects received verbal instructions and a visible and audible trial performance by the researcher. Therefore, all variabilities in the subjects' performances were considered to be the result of subject variability and not the result of any "error," such as changed procedural aspects.

As expected and in corroboration of other studies (eg, Raes and Clement¹⁰), the maximum phonation time of the dysphonic group was significantly shorter than that of the control group. The results on interrater reliability are unambiguous. Repeated measurements showed an average ICC per rater and per trial as high as 0.997 (all ICCs ≥ 0.982). It is concluded that a single rater can provide highly reliable data, thus making a panel of listeners redundant. The results of the trials presented here, that is, the relatively small variance component and the high agreement, suggest that neither a learning nor a fatigue effect occurred in either group. This finding diverges from those reported elsewhere; for instance, Shanks and Mast¹³ demonstrated the progressive effect of fatigue in repeated maximum phonation trials, whereas authors, such as Neiman and Edeson³ and Lewis et al,⁷ found higher phonation times when using repeated trials. Neiman and Edeson recommended conducting at least three trials of maximum phonation time; only then could the performance be expected to approximate the criterion. According to Lewis et al, even three trials would be inadequate to determine maximum phonation time, because they do not represent a "true" measure of maximum duration. However, our present data suggest that, when using maximum phonation time as a measure in voice assessment in a healthy or dysphonic population, a single trial does provide an approximation of a subject's maximum performance, as indicated by a reliability coefficient of 0.939. If a higher level of reliability is required, it is recommended to take the average of more trials, thereby



yielding an increased reliability, as shown in Figure 2.1. Surely, a reliability coefficient of about 0.939 is high enough. However, because the non-observation of either a fatigue or a learning effect does not necessarily amount to its nonexistence, and the fact that previous studies did seem to find such effects, using several trials remains a recommendable alternative. One should bear in mind that replicating a trial on the same day requires minimal effort from both the subjects and researchers. The same does not apply to days' replications. Measuring maximum phonation time on two different days instead of 1 day resulted in an increased reliability, respectively, of 0.911 compared with 0.836. However, because days' replication may be impractical in a clinical setting, a single day showed to be sufficiently reliable.

Maximum phonation time has proven to be a highly reliable measure in voice assessment. However, other less frequently used aerodynamic measurements, such as airflow rate, may be more useful. Furthermore, although commonly used as an evaluation tool to objectify voice therapy effects, maximum phonation time is still under discussion.⁵ Voice therapy may not necessarily result in a prolonged maximum phonation time. Thus, although maximum phonation time is a reliable measurement in voice assessment, it has not yet been proven to be a useful discriminative tool resulting in changed maximum phonation times after therapy.

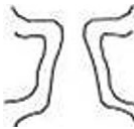


Conclusion

The purpose of the present study was to determine maximum phonation time reliability as a function of the number of trials, days, and raters in dysphonic subjects compared with control subjects matched by gender and age. When using a single trial, a reliability coefficient as high as 0.939 was found. If a higher level of reliability is essential, averaging over several trials may be considered. Nonetheless, as increasing the number of trials requires minimal extra effort, researchers may want to use more trials anyway. Measuring maximum phonation time on two different days instead of on a single day resulted in an increased reliability, respectively, of 0.911 compared with 0.836. Although repeated measurement may be rather impractical in a clinical setting, the results indicated a considerable positive change in reliability scores. As far as interrater reliability is concerned, it can be concluded that a panel of listeners is redundant. A single rater is sufficient to provide highly reliable measurements (reliability coefficient: 0.997).

References

- 1 Speyer R, Wieneke GH, Dejonckere PH. Documentation of progress in voice therapy: perceptual, acoustic, and laryngostroboscopic findings pretherapy and posttherapy. *J Voice* 2004; 18: 325-40
- 2 Dejonckere PH, Crevier-Buchman L, Marie JP, Moerman M, Remacle M, Woisard V. Implementation of the European Laryngological Society (ELS) basic protocol for assessing voice treatment effect. *Rev Laryngol Otol Rhinol (Bord)* 2003; 124: 279-83
- 3 Neiman GS, Edeson B. Procedural aspects of eliciting maximum phonation time. *Folia Phoniatri (Basel)* 1981; 33: 285-93
- 4 Speyer R. Effects of voice therapy: a systematic review. *J Voice* 2008; 22: 565-80
- 5 Treole K, Trudeau MD. Changes in sustained production tasks among women with bilateral vocal nodules before and after voice therapy. *J Voice* 1997; 11: 462-9
- 6 Baken RJO, R.F. *Clinical Measurement of Speech and Voice: 2nd edition Edition*. San Diego: Singular Publishing Group, 2000
- 7 Lewis K, Casteel R, McMahon J. Duration of sustained /a/ related to the number of trials. *Folia Phoniatri (Basel)* 1982; 34: 41-8
- 8 Lee L, Stemple JC, Kizer M. Consistency of acoustic and aerodynamic measures of voice production over 28 days under various testing conditions. *J Voice* 1999; 13: 477-83
- 9 Hirano M, Koike Y, Von Leden H. Maximum phonation time and air usage during phonation. Clinical study. *Folia Phoniatri (Basel)* 1968; 20: 185-201
- 10 Raes JP, Clement PA. Aerodynamic measurements of voice production. *Acta Otorhinolaryngol Belg* 1996; 50: 293-8
- 11 Heyning van de PHR, M.; Cauwenberge van P, et al. Part II: research work of the Belgian study group on voice disorders. *Acta Otorhinolaryngol Belg* 1996; 50: 321-386
- 12 Finnegan DE. Maximum phonation time for children with normal voices. *Folia Phoniatri (Basel)* 1985; 37: 209-15
- 13 Shanks SJ, Mast D. Maximum duration of phonation: objective tool for assessment of voice. *Percept Mot Skills* 1977; 45: 1315-22
- 14 Berger MPFI, T.; Janssen MPE. *Methodologie en statistiek: Deel 2.: Universiteit Maastricht*, 2000





Chapter 3

Observer Agreement for Measurements in Videolaryngostroboscopy



Jan Wouter Brunings, Sophie Vanbelle, Annemarie Akkermans, Nienke MM Heemskerk,
Bernd Kremer, Robert J Stokroos, Laura WJ Baijens

Journal of Voice. 2018;32(6):756-62

Abstract

Objective

This retrospective study evaluated the levels of intraobserver and interobserver agreement for measurements of visuoperceptual variables in videolaryngostroboscopic examinations and compared the observers' behaviour during independent versus consensus panel rating.

Material and Methods

Sixty-four patients with dysphonia of heterogeneous aetiology were included. All subjects underwent a standardised videolaryngostroboscopic examination. Two experienced and trained observers scored the same examinations, first independently and then on a consensus panel. Specific visuoperceptual variables and the clinical diagnosis (as recommended by the Committee on Phoniatrics and the Phonosurgery Committee of the European Laryngological Society and advised by the American Speech-Language-Hearing Association) were scored. Descriptive and Kappa statistics were used.

Results

In general, the intraobserver agreement was better than the agreement between observers for measurements of several variables. The intrapanel observer agreement levels were slightly higher than the intraobserver agreement levels on the independent rating task. When rating on the consensus panel, the observers deviated considerably from the scores they had previously given on the independent rating task.

Conclusion

Observer agreement in videolaryngostroboscopic assessment has important implications not only for the diagnosis and treatment of dysphonic patients but also for the interpretation of the results of scientific studies using videolaryngostroboscopic outcome parameters. The identification of factors that can influence the levels of observer agreement can provide a better understanding of the rating process and its limitations. The results of this study suggest that future research could achieve better agreement levels by rating the visuoperceptual variables in a panel setting.



Introduction

Laryngoscopy was first used to examine the larynx and the vocal folds at the end of the 19th century when Manuel Garcia visualised the vocal folds.¹ Since then, the technique has been refined.² Now, videolaryngostroboscopy is an important tool for the clinical assessment of vocal pathologies and function, and it can be used to evaluate the effectiveness of treatments.³ Clinicians around the world apply videolaryngostroboscopy in daily clinical practice, using many variables to describe vocal fold functioning and the vibratory pattern, although data on the reliability and validity of these variables remain scarce. Several studies report the levels of intraobserver and interobserver agreement for measurements in videolaryngostroboscopy, but few mention which visuoperceptual variables are used in daily clinical practice. Thus, it is unknown whether those studies used the visuoperceptual variables recommended by the Committee on Phoniatrics of the European Laryngological Society (ELS)³ and advised by the American Speech-Language-Hearing Association (ASHA),⁴ most of which are based on the parameters defined by Hirano and Bless.⁵

External validation of clinical decisions based on the interpretation of videolaryngostroboscopic exams requires accurate and reliable measurements.⁶ The description of aspects that can influence observer behaviour and agreement on measured videolaryngostroboscopic variables can elucidate the rating process and thereby facilitate developing procedures to increase the observer agreement levels.

This study has two aims: (1) to evaluate the intraobserver, interobserver, and intrapanel agreement for measurements in videolaryngostroboscopic examinations and (2) to compare the observers' behaviour during independent versus consensus panel rating. It was hypothesised that intrapanel agreement would sustain higher levels of observer reproducibility.



Materials and Methods

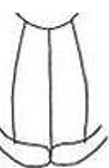
Selection of participants

Videolaryngostroboscopic recordings of 64 patients with dysphonia of heterogeneous aetiology were selected by an independent co-worker laryngologist from the clinical research archives of the Maastricht University Medical Center. All patients were referred to the multidisciplinary outpatient clinic for their dysphonia and underwent a standardised videolaryngostroboscopic examination as part of the regular health-care program. During the selection process, videolaryngostroboscopic recordings were

excluded if the patient presented severe discomfort or gagging. Reasons for exclusion were disturbing body movements and a diagnosis of head and neck cancer. Only recordings made at least 1 year previously to the study in 2014 were included to avoid recognition of patients by the laryngologists and to ensure blinding to patient identity.

Videolaryngostroboscopic examination

All videolaryngostroboscopic examinations and measurements were performed in the same hospital by the same two laryngologists (each with over 10 years of experience in performing and rating videolaryngostroboscopy). All subjects underwent a standardised videolaryngostroboscopic examination. The videos were recorded using a 70- or 90-degree rigid Hopkins endoscope (model 8706 CA; Karl Storz GmbH & Co KG, Tuttlingen, Germany), which was attached to an Alphontron Stroboscopy ACLS camera, Alphontron Lightsource, IVACX computerised video archiving system (Alphontron Medical Systems Rotterdam, the Netherlands), and were recorded on a DVD. A rigid endoscopy was carried out to obtain an image of a quality superior to that attainable with flexible equipment. If necessary, a topical anaesthetic (Xylocaine 10% (AstraZeneca AB, Södertälje, Sweden)) was applied. The examination had to be stable enough to allow continuous stroboscopic “tracking” of vocal fold vibration. Care was taken to ensure that each selected video contained a phonation time long enough to allow for the registration of a sustained phonation of the vowel /i:/, at comfortable pitch and loudness, and at least one complete cycle of vibration. The audio track was removed from the recordings. During the examination, subjects were seated upright wearing their dental prosthesis if present. The field of the image included the laryngeal vestibule, vocal folds, anterior commissure, and arytenoids. Informed consent was obtained from all subjects.



Study design and measurements

Prior to data collection, the two laryngologists received training in visuoperceptual measurement for the nominal and ordinal variables given in Table 3.1. The categorical scales of the variables recommended by the ELS and ASHA were modified as described in Table 3.1. The videos used for training were not included in the experimental set of videos. Training in the exact interpretation of each scale category took place in sessions with both observers and was intended to generate substantial to almost perfect levels of intraobserver and interobserver agreement (kappa 0.61–0.99). The duration of the program was predetermined. It consisted of two training sessions, approximately 1 hour each, interspersed over the course of a month, with practice periods when the observers had to do test runs separately. A written manual with well-defined

descriptions of the scales' levels was available during the training and the rating process for the observers to consult. Visuo-perceptual nominal and ordinal variables and the clinical diagnosis (nominal variable) were scored for each videolaryngostroboscopic video at varying speeds if desired by the rater (slow motion, normal, up to frame-by-frame) as often as necessary using the software program *Windows Movie Maker* version 5.1 (Microsoft Corporation, Redmond, WA). Clinical diagnosis of the vocal pathology was derived from reports of the Phonosurgery Committee of the ELS (Table 3.2).⁷

Table 3.1 Visuo-perceptual nominal and ordinal variables used to rate videolaryngostroboscopic recordings.

Variable name	Definition	Scale
Diagnosis of vocal pathology [*]	Diagnosis derived from reports of Phonosurgery Committee of the European Laryngological Society (ELS) ⁷	Not applicable (see Table 3.2)
Amplitude, left and right vocal fold [†]	Extent of vocal fold displacement near glottic opening	0 = normal 1 = impaired 2 = absent
Periodicity, left and right vocal fold [†]	Temporal regularity of vibratory cycles	0 = normal 1 = impaired
Symmetry [†]	Symmetry of mucosal displacement	0 = normal 1 = impaired
Closure [†]	Degree of glottic closure during the closed phase of vibration	0 = normal 1 = impaired
Defect [†]	Type of glottic closure: predominant mucosal closure patterns	0 = normal 1 = oval 2 = hourglass 3 = anterior 4 = posterior (<50%) 5 = complete (>50%)

^{*} Visuo-perceptual nominal variable; [†] Visuo-perceptual ordinal variable.

The rating process comprised two separate tasks: independent rating and consensus panel rating of the same randomised videolaryngostroboscopic videos. While the two observers were blinded to each other's ratings during the independent task, the decision on the score was reached in consensus during the panel task, which took place 1 month later. Differences between the raters during the panel task were solved by discussion using the manual with well-defined descriptions of the scales' levels and controlled mutual feedback.

The observers were blinded to the vocal sound recording and the medical history and identity of the patients.



Table 3.2 Clinical diagnosis of vocal pathology, as assessed by the panel, derived from reports of the Phonosurgery Committee of the European Laryngological Society, in absolute numbers (*n*) and percentages (%).

Diagnosis	<i>n</i> (%)
Anterior webbing	2 (3.2)
Atrophy	3 (4.7)
Granuloma vocal process	5 (7.8)
Intracordal hematoma	1 (1.6)
Intracordal cyst	8 (12.5)
Laryngitis	10 (15.6)
Laryngitis sicca	1 (1.6)
Laryngocele	1 (1.6)
Nodules	1 (1.6)
Normal larynx	9 (14.1)
Polyp	0 (0.0)
Papillomatosis	0 (0.0)
Paralysis	11 (17.2)
Reinke space oedema III/IV	3 (4.7)
Reinke space oedema I/II	7 (10.9)
Scarification	2 (3.1)



To determine how well the intraobserver scores arrived at jointly agreed with those obtained independently, the laryngologists (again blinded) repeated the assessments within a period of 4 months for all variables on 32 videos randomly selected from the 64 initial videos. They were advised to limit the duration of the measurement sessions (maximum 1 hour) to avoid fatigue and prevent instability of observers' input.

Statistical analysis

Results were expressed as means (standard deviation) for quantitative variables, while frequencies and proportions (%) were used for categorical variables. For the nominal variables, i.e., periodicity, symmetry, closure, diagnosis, and defect, disagreement on a video was defined as a difference in the classification made by the two observers. For ordinal variables, i.e., amplitude, disagreement was defined as the number of categories separating the classification made by the two observers. In this latter case, a disagreement on two adjacent categories of the same measurement scale is therefore considered less important than a disagreement on more distant categories. The percentage of videos on which two observers agree was reported. However, agreement between two observers can also occur even if they randomly allocate the videos. Consequently, the level of agreement was adjusted for the amount of agreement expected by chance, leading to Cohen's kappa coefficient for nominal variables and the linear weighted kappa coefficient for ordinal variables. The intraobserver and

interobserver agreement level, quantified through kappa coefficients, was reported with a 95% confidence interval (CI).

The standard error of the kappa coefficient used to construct the 95% CI was adjusted for the presence of repeated measurements when necessary, i.e., when a variable was assessed on both sides of the larynx of the same patient.⁸ Not accounting for the repeated measurements will artificially increase the sample size and lead to CIs that are too narrow.

A comparison of the intraobserver agreement levels obtained by the two observers individually and jointly was made using Hotelling's T squared statistic, which is a generalisation of Student' t-test to more than two variables. The variance-covariance matrix used in the calculations was estimated using a bootstrap method.^{9,10} Missing data were not replaced, i.e., no imputation. Data analysis was performed using R (version 3.0.1 for Windows [Microsoft Corporation])

Results

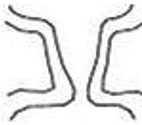
Characteristics of the participants

The demographics of the dysphonic patients and their clinical characteristics, as assessed by the panel, are displayed in Table 3.3. The frequency distribution of patients per category of the videolaryngostroboscopic variables is shown in Table 3.3, indicating the average voice function of the study population.

Patients (30 men and 30 women, four missing values) were, on average, 50.3 years old, the men being older than the women (mean: 56.6 versus 44.3 years, $P=0.007$). The study population was a fair reflection of the dysphonic patients who consult our outpatient clinic. The most frequently observed conditions, as assessed by the panel, were paralysis ($n=11$; 17.2%) laryngitis ($n=10$; 15.6%), normal larynx ($n=9$; 14.1%), intracordal cyst ($n=8$; 12.5%), and Reinke oedema grades I and II ($n=7$; 10.9%) (Table 3.2).

Agreement study

Interobserver agreement levels are given in Table 3.4. Levels of intraobserver agreement for the two individual observers and the panel are shown in Table 3.5. Table 3.6 gives the number of changes in the video scores made by the observers during the panel meeting compared with their individual assessment of the same videos.



Amplitude and periodicity were measured separately for the left and right sides. For these variables, agreement is determined for each side separately and at the patient level.

Table 3.3 Demographic and clinical characteristics of the 64 dysphonic patients and the frequency distribution of patients per category of the videolaryngostroboscopic variables in absolute numbers (*n*) and percentages (%).

Variable		<i>n</i> (%)	Mean (SD)
Gender	Man	30 (50.0)	
	Woman	30 (50.0)	
Age	Total		50.3 (17.8)
	Man		56.6 (17.3)
	Woman		44.3 (16.4)
Amplitude- left side	Normal	18 (34.6)	
	Impaired	28 (53.8)	
	Absent	6 (11.5)	
Amplitude- right side	Normal	26 (44.2)	
	Impaired	26 (50.0)	
	Absent	3 (5.8)	
Periodicity- left side	Normal	33 (71.7)	
	Impaired	13 (28.3)	
Periodicity- right side	Normal	36 (78.2)	
	Impaired	10 (21.7)	
Symmetry	Normal	8 (14.0)	
	Impaired	49 (86.0)	
Closure	Normal	8 (13.1)	
	Impaired	53 (86.9)	
Defect	Normal	8 (13.1)	
	Oval	12 (19.7)	
	Hourglass	9 (14.8)	
	Anterior	3 (4.9)	
	Posterior	16 (26.2)	
	Complete	13 (21.3)	



Table 3.4 Interobserver agreement levels for the two independent observers on the 64 dysphonic patients.

	<i>n</i>	% agreement	Kappa*	SE	95% CI
Diagnosis	64	67.1	0.63	0.064	0.50-0.75
Amplitude	101	57.4	0.37	0.090	0.19-0.55
Left	50	62.0	0.46	0.10	0.14-0.77
Right	51	52.9	0.28	0.11	-0.15-0.57
Periodicity	94	71.0	0.30	0.11	0.084-0.52
Left	46	72.0	0.36	0.12	0.12-0.61
Right	48	71.0	0.22	0.11	0.0021-0.43
Symmetry	57	86.0	0	NA	NA
Closure	63	98.0	0.85	0.15	0.56-1.00
Defect	54	70.4	0.62	0.077	0.47-0.77

*Kappa coefficient; <0, less than chance agreement. 0.01–0.20, slight agreement. 0.21– 0.40, fair agreement. 0.41–0.60, moderate agreement. 0.61–0.80, substantial agreement. 0.81–0.99, almost perfect agreement. NA, not available; SE, standard error.

Table 3.5 Intraobserver agreement levels for each observer and the consensus panel on 32 dysphonic patients.

	Observer 1			Observer 2			Panel		P-value
	n	%	Kappa* (SE)	n	%	Kappa (SE)	n	%	
	agreement			agreement			agreement		
Diagnosis	32	71.9	0.67 (0.086)	32	78.1	0.75 (0.081)	32	81.3	0.57
Amplitude	48	58.3	0.42 (0.13)	47	55.3	0.37 (0.17)	52	78.8	0.016
Left	24	58.3	0.45 (0.24)	23	56.5	0.38 (0.25)	32	76.9	0.053
Right	24	54.2	0.36 (0.22)	24	54.2	0.37 (0.24)	26	80.8	0.008
Periodicity	46	76.1	0.51 (0.18)	46	100	1 (NA)	43	72.1	0.37 (0.20)
Left	23	78.3	0.56 (0.29)	23	100	1 (NA)	22	68.2	0.32 [†]
Right	23	73.9	0.46 (0.27)	23	100	1 (NA)	21	76.2	0.76 [†]
Symmetry	29	93.1	-0.036 (0.025)	26	53.8	0.20 (0.10)	26	92.3	0.0010
Closure	30	96.7	0.78 (0.21)	31	96.7	0.78 (0.21)	30	83.3	0.012
Defect	24	75.0	0.59 (0.25)	31	54.8	0.46 (0.24)	30	60.0	0.24

[†] Comparing only the intraobserver agreement for observer 1 and the panel; *Kappa coefficient. <0, less than chance agreement. 0.01–0.20, slight agreement. 0.21–0.40, fair agreement. 0.41–0.60, moderate agreement. 0.61–0.80, substantial agreement. 0.81–0.99, almost perfect agreement. NA, not available; SE, standard error.

Table 3.6 Scores given independently compared to scores given on the consensus panel for the same measurement. Number (%) of changes in the scores arising when comparing the individual to consensus measurements.*

	Observer 1 independently; first measurement	Observer 1 independently; repeated measurement	Observer 2 independently; first measurement	Observer 2 independently; repeated measurement	Total
	Number (%) of changes	Number (%) of changes	Number (%) of changes	Number (%) of changes	Number (%) of changes
Diagnosis	11 (34.4)	2 (6.3)	6 (18.8)	8 (25.0)	27 (25.8) [†]
Amplitude	17 (34.0)	10 (20.0)	18 (35.0)	16 (32.0)	61 (30.3)
Left	7 (28.0)	5 (20.0)	8 (32.0)	10 (40.0)	30 (30.0)
Right	10 (40.0)	5 (20.0)	10 (38.5)	6 (24.0)	31 (30.7)
Periodicity	16 (39.0)	9 (18.7)	6 (14.6)	18 (36.0)	49 (27.2)
Left	8 (38.0)	5 (20.8)	3 (14.3)	10 (40.0)	26 (28.6)
Right	8 (40.0)	4 (16.7)	3 (15.0)	8 (32.0)	23 (25.8)
Symmetry	5 (17.9)	3 (12.0)	2 (7.1)	12 (48.0)	22 (20.8)
Closure	4 (12.9)	1 (3.4)	4 (12.9)	2 (6.7)	11 (9.1)
Defect	13 (52.0)	6 (21.4)	19 (61.3)	13 (43.3)	51 (44.7)

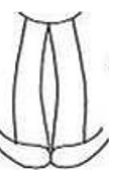
*To facilitate the interpretation of the table, a more detailed description is given here. For instance, in the row "Diagnosis," in 11 videos observer 1 changed his initial score (given during the independent rating task) when he rated these same video recordings in the panel setting. Observer 1 rated 35 video recordings independently a second time to obtain his intraobserver agreement level. On these previously rated video recordings, he changed his score twice when he rated the same videos in the panel setting. [†]This column represents the summed score of all the changes made in the initial scores after discussion by the two observers in the panel setting.

Diagnosis

Interobserver agreement for clinical diagnosis was 0.63 (95% CI: 0.50–0.75) (Table 3.4). Intraobserver agreement was 0.67 (0.50–0.84) for observer 1, 0.75 (0.59–0.91) for observer 2, and 0.78 (0.62–0.93) for both during the panel rating task (Table 3.5). No particular disagreement pattern was observed. Nor was any significant difference in intraobserver agreement found between the panel and the individual assessments ($P=0.57$). However, the clinical diagnosis based on the same videos during the panel rating task and the individual assessments changed for 27 (25.8%) patients (Table 3.6).

Amplitude

Interobserver agreement for the variable amplitude was 0.37 (0.19–0.55) (Table 3.4). Disagreement arose on the 61 videos classified as “impaired” by the first observer. Among these videos, 14 (23%) were classified as “normal” and 12 (19.7%) as “absent amplitude” by the second observer. Intraobserver agreement was 0.42 (0.18–0.67) for observer 1, 0.37 (0.15–0.59) for observer 2, and 0.76 (0.53–0.92) for the observers together during the panel rating task (Table 3.5). Here too, more than 50% of the videos initially classified as “impaired” were subsequently classified as “normal” or “absent amplitude” by the team. Intraobserver agreement was higher during the panel meeting than during individual assessments ($P=0.016$) (Table 3.5). After discussion, the observers modified the amplitude classification of 61 (30.3%) videos, most of which had been classified as “impaired” or “absent” during their individual assessments (Table 3.6).



Periodicity

Interobserver agreement for the variable periodicity was 0.30 (0.084–0.52) (Table 3.4). While the first observer classified 36 (38.3%) videos as impaired, only 11 (11.7%) were classified thus by the second observer. Intraobserver agreement was 0.51 (0.17–0.86) for observer 1, 1 (not applicable) for observer 2, and 0.37 (–0.023–0.76) for the observers together (Table 3.5). No particular disagreement pattern was observed. Forty-nine (27.2%) of the videos scored individually by the observers were rescored at the panel meeting (Table 3.6).

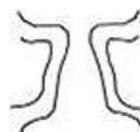
Symmetry

Interobserver agreement for the variable symmetry was 0, although 86% of the videos were classified in the same category by the two observers (Table 3.4). All videos were classified as “impaired” by the first observer, explaining the difference between the

high percentage of agreement and the low value of Cohen's kappa coefficient. Among these, 49 (86%) were classified as impaired by the second observer. Intraobserver agreement was -0.036 (-0.085 ; 0.014) for observer 1, 0.20 (0.0040 – 0.40) for observer 2, and 0.70 (0.32 – 1.00) for the panel (Table 3.5). This result is in contrast with the percentage of agreement between the two measurement occasions (93.1% for observer 1, 53.8% for observer 2, and 92.3% for the panel) (Table 3.5). Here too, the difference between the percentage of agreement and the kappa coefficient is explained by the low number of patients classified with normal symmetry. Normal symmetry was observed only in 1 (3.4%) patient for observer 1, 10 (38%) patients for observer 2, and 4 (15.4%) patients for the panel rating task. Intraobserver agreement was higher for the panel task than for the independent assessments ($P=0.0010$) (Table 3.5). The individually assigned classifications of 22 (20.8%) videos were changed during the panel meeting (Table 3.6).

Closure

Interobserver agreement for closure was 0.85 (0.56 – 1.00) (Table 3.4). Intraobserver agreement was 0.78 (0.02 – 1.00) for observer 1, 0.78 (0.37 – 1.00) for observer 2, and 0.19 (-0.28 – 0.67) for the observers jointly (Table 3.5). No particular pattern of disagreement was seen on this variable. Normal closure was rarely noted (maximum two [6.7%] videos by observer 1, three [10%] for observer 2, and four [13.3%] for the panel). Intraobserver agreement was low for the panel, even though 25 (83.3%) videos were classified in the same category by the panel on both measurement occasions. Intraobserver agreement was higher for individual assessments than for the panel ($P=0.012$) (Table 3.5). During the panel meeting, 11 (9.1%) videos were reclassified (Table 3.6).



Defect

Interobserver agreement for the variable defect was 0.62 (0.47 – 0.77) (Table 3.4). Intraobserver agreement was 0.59 (0.47 – 0.90) for observer 1, 0.46 (0.24 – 0.67) for observer 2, and 0.52 (0.28 – 0.72) for the panel (Table 3.5). No particular disagreement pattern was seen. There was no significant difference between the individual and consensus panels for intraobserver agreement ($P=0.24$) (Table 3.5). The classification of 51 (44.7%) videos was changed during the panel meeting (Table 3.6).

Discussion

Visuoperceptual analysis of videolaryngostroboscopic images has some limitations and is subject to error. However, no other alternative is available to analyse these images in daily clinical practice. While recourse to videolaryngostroboscopy is increasing, research on the standardisation and validation of measurement criteria in these exams lags behind. Crucially, the interpretation of videolaryngostroboscopic images rests on visual judgment and is thus subjective.⁶ It might be influenced by factors such as experience, observer fatigue, the severity of vocal fold disease, etc.¹¹ Investigating observer agreement is the first step to demonstrating the validity of the procedure because measurement criteria can only be valid if they are reproducible in terms of observer agreement.¹² Some studies have addressed observer agreement on some well-known visuoperceptual variables.¹¹ Nonetheless, the variability in the scoring of videolaryngostroboscopic exams remains underexplored. Given its role in clinical decision making, such as on (surgical) treatment, these exams must be both accurate and reliable.



The present study considers intraobserver and interobserver agreement for visuoperceptual measurements in videolaryngostroboscopic exams and explores any discrepancies that arise between independent and consensus ratings to better understand the causes of disagreement among observers. In this study, observer agreement was determined for specific visuoperceptual variables and clinical diagnoses that are used in daily clinical practice derived from the recommendations of the Committee on Phoniatrics and the Phonosurgery Committee of the ELS and advised by the ASHA. For the independent rating task, intraobserver agreement on several variables (diagnosis, amplitude, and closure) was similar for both observers. This suggests that the two observers had a similar interpretation of the nominal and ordinal scoring systems and were consistent when repeating the measurements. In general, the intraobserver agreement was better than the agreement between different observers for measurements of almost all variables except for “closure” and “defect.” The finding that the intraobserver agreement was better than the agreement between observers for the majority of variables might be explained as follows. First, even though the observers understood the nominal and ordinal scoring systems well, they did not reach consensus on the cut-off points. The description of the rating scale does not give the precise range of each nominal or ordinal category, which leaves it up to the observers to set their own boundaries. Accordingly, the interpretation of videolaryngostroboscopic variables is often subject to a wide degree of variability because of the subjective discrimination of the levels of these scales.

In an overview of the literature, studies investigating the intraobserver and interobserver agreement of measurements in videolaryngostroboscopic exams were found to have methodological shortcomings. Notably, interobserver agreement varied considerably with the method of measurement, the level of experience of the observers, the number and pathology of the subjects, and the presence or absence of the audio track (auditive information).¹³⁻²⁰ Often, it was not reported whether the observers had received pre-experimental training.^{19,21,22} In general, the conclusions could not be compared across the studies because of diverging study designs, small populations, and different ways of evaluating the videolaryngostroboscopic exams. Given their heterogeneous methodologies, the studies could not be compared with one another. Some had issues concerning the generalisation of results. For instance, performing several measurements on a large number of participants or healthy subjects in a limited time period makes the exercise susceptible to bias and does not reflect clinical practice.^{14,17,19} Therefore, observer agreement varies across raters and populations.^{11,15,16,19,20,23,24}

In the present study, intrapanel observer agreement was slightly higher than the intraobserver levels on the independent rating task (except for the variables periodicity and closure). That difference suggests that consensus rating might be a useful alternative to the independent rating of videolaryngostroboscopic exams, as the discussion of cases in a panel may improve concordance. However, the level of agreement between two separate consensus panels with different members still needs to be explored, particularly in comparison with individual interobserver agreement levels. Also, the validity of the measurement criteria of the different variables should be investigated.



Observers were consistent for several measured variables (diagnosis, periodicity, closure, defect) when rescoring videos independently or on the consensus panel. However, when repeating the task on the panel, they frequently adjusted the scores they had given previously when rating the same measurements independently. That tendency to change in a panel setting reflects the observers' interpretation of the videolaryngostroboscopic scoring system.

Limitations of the study

The results of the study might be affected by the limited number of videos analysed. However, considering the amount of time, observers spent during the pre-experimental training period and the rating process, a larger number of videos would have made the

process last so long that it would be difficult to engage experienced observers. This study presents the videolaryngostroboscopic scores of just two observers. Comparing scores of a larger number of observers could yield different results. However, the situation in which the present study was conducted resembles most clinical settings where one or two professionals are usually responsible for the interpretation of videolaryngostroboscopic exams.

Conclusion

Observer agreement for videolaryngostroboscopic measurements has important implications not only for the diagnosis and treatment of dysphonic patients but also for the interpretation of the results of studies using videolaryngostroboscopic outcome parameters. The identification of factors that can influence the observer agreement levels such as the setting of the rating task (independent versus panel) can provide a better understanding of the rating process and its limitations. For research purposes, this study suggests that the visuoperceptual variables should be rated in a panel to achieve better agreement levels. The same procedure is suggested for the clinical setting: to use an experienced panel composed, for example, of the laryngologist and speech and language pathologist who work together at the outpatient voice clinic.



References

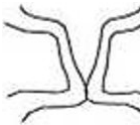
1. Garcia M. Physiological observations on the human voice Proceedings of the Royal Society London 1855;7:339.
2. Heman-Ackah YD. Diagnostic tools in laryngology. *Curr Opin Otolaryngol Head Neck Surg.* 2004;12: 549-52.
3. Dejonckere PH, Bradley P, Clemente P, et al. A basic protocol for functional assessment 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.
4. Association AS-L-H. Vocal tract visualization and imaging: technical report 2004.
5. Hirano M, Bless DM. Videostroboscopic examination of the larynx San Diego, California Singular Publishing Group Inc.; 1993.
6. Roy N, Barkmeier-Kraemer J, Eadie T, et al. Evidence-based clinical voice assessment: a systematic review. *Am J Speech Lang Pathol.* 2013;22:212-26.
7. Friedrich G, Remacle M, Birchall M, Marie JP, Arens C. Defining phonosurgery: a proposal for classification and nomenclature by the Phonosurgery Committee of the European Laryngological Society (ELS). *Eur Arch Otorhinolaryngol.* 2007;264:1191-200.
8. Yang Z, Zhou M. Kappa statistic for clustered matched-pair data. *Stat Med.* 2014;33: 2612-33.
9. Vanbelle S, Albert A. A bootstrap method for comparing correlated kappa coefficients. *Journal of Statistical Computation and Simulation.* 2008;78:1009-15.
10. Vanbelle S. Comparing dependent kappa coefficients obtained on multilevel data. *Biom J.* 2017.
11. Rosen CA. Stroboscopy as a research instrument: development of a perceptual evaluation tool. *Laryngoscope.* 2005;115:423-8.
12. de Vet HC, Terwee CB, Knol DL, Bouter LM. When to use agreement versus reliability measures. *J Clin Epidemiol.* 2006;59:1033-9.
13. Gelfer MP BD. Evaluation of Vocal Fold Vibratory Patterns in Normal Voices. *J Voice.* 1990;4:335-45.
14. Kendall KA. High-speed laryngeal imaging compared with videostroboscopy in healthy subjects. *Arch Otolaryngol Head Neck Surg.* 2009;135:274-81.
15. Nawka T, Konerding U. The Interrater Reliability of Stroboscopy Evaluations. *J Voice.* 2012.
16. Poburka BJ. A new stroboscopy rating form. *J Voice.* 1999;13:403-13.
17. Pemberton C, Russell A, Priestley J, Havas T, Hooper J, Clark P. Characteristics of normal larynges under flexible fiberoptic and stroboscopic examination: an Australian perspective. *J Voice.* 1993;7:382-9.
18. Dejonckere PH, Crevier L, Elbaz E, et al. Quantitative rating of video-laryngostroboscopy: a reliability study. *Revue de laryngologie - otologie - rhinologie.* 1998;119:259-60.
19. Lundy DS, Casiano RR, Sullivan PA, Roy S, Xue JW, Evans J. Incidence of abnormal laryngeal findings in asymptomatic singing students. *Otolaryngol Head Neck Surg.* 1999;121:69-77.
20. Nospes S, Kuhr K, Napiontek U, Keilmann A. [Stroboscopy findings: a comparison of flexible CCD-videostroboscopy and rigid stroboscopy]. *Laryngorhinootologie.* 2011;90:218-23.
21. Teitler N. Examiner bias: influence of patient history on perceptual ratings of videostroboscopy. *J Voice.* 1995;9:95-105.
22. Olthoff A, Woywod C, Kruse E. Stroboscopy versus high-speed glottography: a comparative study. *Laryngoscope.* 2007;117:1123-6.
23. Chau HN, Desai K, Georgalas C, Harries M. Variability in nomenclature of benign laryngeal pathology based on video laryngoscopy with and without stroboscopy. *Clin Otolaryngol.* 2005;30:424-7.
24. Cipriani NA, Martin DE, Corey JP, et al. The clinicopathologic spectrum of benign mass lesions of the vocal fold due to vocal abuse. *Int J Surg Pathol.* 2011;19:583-7.





Chapter 4

The Expression of Oestrogen and Progesterone Receptors in the Human Larynx



Jan Wouter Brunings, Janneke J BFG Schepens, Carine J Peutz-Kootstra,
Kenneth W Kross

Journal of Voice. 2013 May;27(3):376-80

Abstract

Objectives

The human female voice changes in quality during the menstrual cycle and during pregnancy and menopause. The underlying pathophysiological mechanisms are as yet not known. The aim of this study, therefore, was to evaluate the existence of oestrogen receptors (ERs) and progesterone receptors (PRs) in the human vocal fold.

Material and Methods

Biopsies of benign vocal fold lesions from 37 female patients were obtained during surgery. Immunohistochemistry for expression of ERs and PRs was performed and evaluated on a semi-quantitative scale by two independent pathologists.

Results

In series 1, immunohistochemical staining showed six sections positive for ER and three sections for PR. One section had positive staining for both receptors. In series 2, immunohistochemical staining showed 10 of the 15 oedema biopsies were positive for ER and six for PR. Six biopsies expressed both receptors. Four of the 10 laryngocele biopsies were positive for ER and two for PR. One was positive for both receptors.

Conclusion

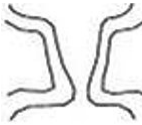
Our study demonstrates that ERs and PRs are expressed in the larynx of the female human vocal fold in conjunction with oedema. The function of these receptors has to be elucidated in future studies.



Introduction

Vocal disorders occur in endocrine diseases with dysregulation of the hypopituitary gland, thyroid gland, adrenal glands, testicles, or ovaries. During a severe hypothyroid state, the voice might be hoarse because of a marked oedema of the vocal mucosa or a congested vocal muscle. This hoarseness of the voice can also occur during menstruation, pregnancy, and menopause.¹ The vocal harmonics are hormonally dependent. This is illustrated by changes in voice occurring during male and female puberty. The larynx develops in response to the production of sex steroid hormones with significant differences between male and female.² As such, it is an androgen-dependent organ.^{3,4} Illustrating is that the absence of testosterone before puberty in a castrato will evolve in a man with feminine voice.⁵

The premenstrual and menopausal vocal syndrome is characterised by lowered vocal intensity, vocal fatigue, a decreased range with loss of the high tones, and a loss of vocal quality.⁵⁻⁹ Hormonal changes associated with the menopause exert an effect on the female vocal folds as well as on other target organs. A decrease in serum levels of oestrogen and progesterone and increase of androgen dominance during the menopausal transition has been shown to affect voice quality.⁶ Because voice quality is influenced by hormonal changes, we hypothesise that oestrogen receptors (ER) and progesterone receptors (PR) are expressed in the human female larynx.



The ER is a nuclear protein exerting important functions in hormone-dependent tissues.⁴ Oestrogens have a hypertrophic and proliferative effect on mucosa. One of the functions is the reduction of desquamation of the superficial layers of the squamous epithelium and differentiation and complete maturation of the fat cells. Earlier findings showed ER in the mesenchyme of the vocal folds and lateral to the laryngeal aditus of mice.¹⁰ In baboons, the vocalis muscle and other mesenchymal tissues contain the largest number of receptors.¹¹ Presence of ER in the human vocal fold has not been shown convincingly.^{4,12-15}

Several studies have been performed in cadaver specimen, but ER/PR expression has not been conclusively shown in biopsy material. Schneider et al.¹³ used specimens of human vocal folds, which were obtained 4–8 hours after death. They could not find any nucleus staining of ER and PR. Newman et al.¹² observed the ER and PR expression in the vocal folds by using immunohistochemical staining within 18 hours after the death of 42 donors. Specimens showed positive staining in epithelial cytoplasm and nucleus, glandular cytoplasm and nucleus, and fibroblasts within the lamina propria. Ferguson

et al.² took four laryngeal tissue specimens from autopsy cases within 4–16 hours of death and maintained at -70°C until immunohistochemical analysis could be performed.²

One of the reasons that they found no nuclear staining without cytoplasmic staining might be that nuclear stability was lost because of post-mortem autolysis and therefore receptor degradation or migration from the nucleus to the cytoplasm can occur. Fodor et al.¹⁶ clarified the effect of post-mortem delay of sex steroid receptors. They used immunohistochemistry to study the stability of ER, PR, and androgen receptor (AR) in the rat hypothalamus and adjacent structures obtained within 0–24 hours after death after fixating for 20 days. Both AR and ER but not PR immunoreactivity were decreased after immersion fixation compared with the perfused sections at time point zero. In brains fixed by immersion, all three receptors decreased gradually with increasing post-mortem time, and ER became hardly detectable after 24 hours post-mortem. Newman et al.^{1,2} found that time interval for receptor degradation was up to 18 hours after death. Ferguson et al.² found that an interval of 16 hours before degradation of the receptors in four laryngeal tissue specimens.



In this study, we investigated the presence of ER and PR expression in benign laryngeal tissue obtained from female patients operated for benign vocal fold lesions. The specimens were fixated immediately after the removal.

Materials and Methods

Patient material

Histological material was obtained from two groups of patients. Group 1 consisted of archived data (period 2006) from 25 female patients treated for benign vocal fold lesions. This group provided 34 vocal fold samples. The samples contained polyps, cysts, laryngeal (Reinke space) oedema, granulomas, nodules, and laryngoceles. Age range was 18–74 years. Group 2 consisted of archived data (period 2003–2006) from 12 female patients, providing vocal fold tissue reflecting oedema and laryngocele. This group provided 25 vocal fold samples. Age range was 35–78 years.

Immunohistochemistry

Specimens were, after the operation, immediately fixated in formaldehyde and embedded in paraffin after processing. For this study, 3- μ thick sections were cut and

placed overnight in a 37°C stove followed by deparaffinisation in xylene and rehydration in graded ethanol (100% and 70%). In between the ethanol steps, the sections were placed for 15 minutes in 5% H₂O₂ in methanol for endogenous blockage. The sections were then separated for antigen retrieval. The ER sections and controls were placed in boiling Tris-ethylenediaminetetraacetic acid solution (Klinipath, Duiven, The Netherlands) and pre-treated for 10 minutes at 750 W in a microwave. The PR sections and controls were placed in citrate buffer (Dakocytomation; DAKO, Heverlee, Belgium) in the same procedure. Benign breast tissue served as a positive control. After cooling down for 30 minutes, sections were washed with Tris buffer 0.005 mol/L and 120 µL antibodies (ER 1:100, PR 1:50) were applied for 45 minutes. After washing with Tris buffer, 120 µL Powervision (Poly-HRP-GAM/R/R IgG; Immunologic, Duiven, The Netherlands) was applied for 30 minutes. After washing with Tris buffer, 200 µL diaminobenzidine (DAB) (liquid DAB + substrate chromogen system [Dakocytomation]) was applied for 20 minutes. After washing with Tris buffer, staining with hematoxylin was performed followed by a bath in cold tap water. Dehydration in graded ethanol (70–100%) and xylene. All procedures were performed at room temperature.

Histological evaluation

Slides were evaluated independently by two senior pathologists. The pathologists were blinded for patient characteristics and the slides were scored in randomised order. Histological sections were scored for ER and/or PR positivity in different tissue compartments. Staining of surface epithelium, fibrous stroma, and gland epithelium was scored on a semi-quantitative scale; 0 was scored as no staining and 3 as intense staining in the whole tissue compartment. Staining with positivity of more than 2 was judged as positive staining and thus as ER/PR positivity in the patient. Before interpreting the study outcome, the interrater variability was determined. To obtain the interrater variability, the measurements of each rater were compared using descriptive statistics and the percentage of the total amount. The reliability between raters was expressed as per cent agreement. A difference in scoring by the pathologists of more than 1 was judged as not consistent. These sections were re-evaluated by both pathologists together in a separate session, and staining intensity was scored together. This way of semi-quantitative hormone receptor scoring is a widely used method.^{17,18}

Statistics

Descriptive statistics were used. The sample sizes were too small to perform statistical analysis as Kappa tests. Numerical descriptions are given in Tables 4.1 and 4.2. In Table 4.2, the sample size, the subgroups, and demography distribution are showed.

Table 4.1 Disagreement percentage between the raters.

	Rater 1 (n)	Rater 2 (n)	Disagreement (n (% of total))	Result after conference
Series 1, n=34				
ER positivity	7	7	5 (14.7)	6
PR positivity	1	10	9 (26.5)	4
Series 2, n=25				
ER positivity	19	16	10 (40)	14
PR positivity	6	11	11 (44)	8

Table 4.2 Diagnosis, number of patients, number of section, and positivity for ER and PR.

Lesions	No. of Patients	No. of Sections	ER-Positive Section	PR-Positive Section	ER- and PR-Positive Section	ER- and PR-Positive Patients	Age Range of Positive Patients (mean)
Group 1							
Cyst	11	14	1	0	0	0	54 (54)
Granuloma	2	2	0	0	0	0	
Noduli	3	4	0	0	0	0	
Polyp	4	8	1	1	0	0	67 (67)
Oedema	4	5	4	2	2	2	45–63 (56)
Laryngocele	1	1	0	1	0	0	8 (78)
Total 1	25	34	6	4	2	2	45–78 (61)
Group 2							
Oedema	7	15	10	6	6	3	46–78 (60)
Laryngocele	5	10	4	2	1	1	43–78 (65)
Total 2	12	25	14	8	7	4	43–78 (62)
Total	37	59	20	12	9	6	43–78 (62)



Results

Benign breast tissue served as a positive control. All laryngeal slides showed less intense staining than these positive controls. The ER and PR staining was mainly in fibroblasts aligning small vessels and more prominent in the specimen from laryngeal oedema (Figure 4.1). In Table 4.1, the disagreement percentage between the two raters is given.

In the first series ($n=34$), “Rater 1” found seven sections positive for ER and one for PR and “Rater 2” found seven for ER and 10 for PR. Of both the seven positive ER findings, there was no agreement for five sections. After conference, a total of six ER-positive remained. Of the PR-positive findings, the disagreement concerned nine sections. Four sections remained positive after conference.

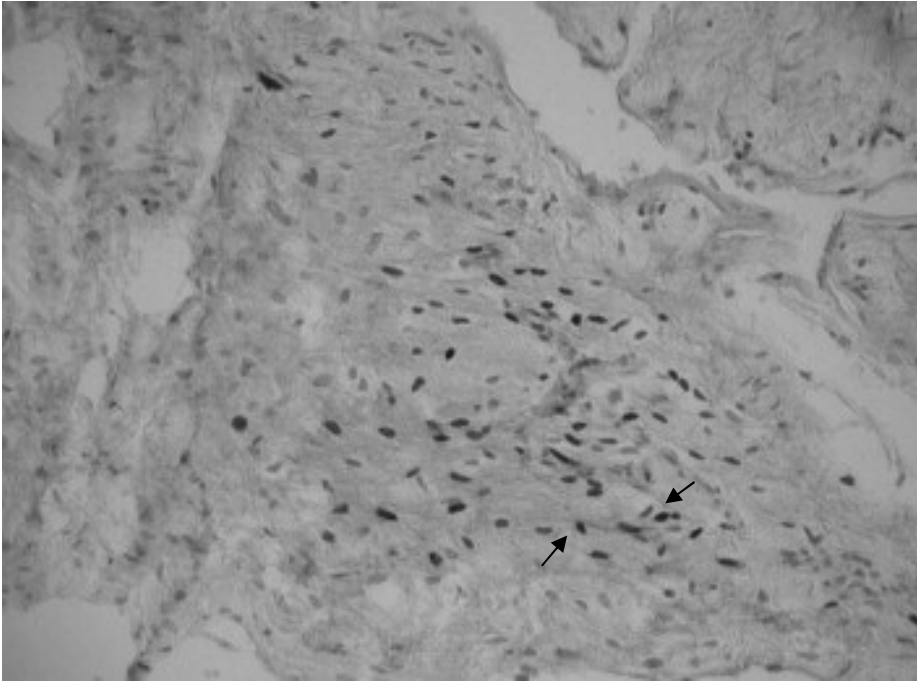


Figure 4.1 In ER staining sections ($\times 200$ magnification), fibroblasts aligning small vessels, the nuclei of these cells are stained dark in greyscale visualisation (arrows).

In the second series ($n=25$), “Rater 1” found 19 sections positive for ER and six for PR and “Rater 2” found 16 for ER and 11 for PR. Of the ER-positive findings, there was no agreement for 10 sections. After conference, a total of 14 ER-positive sections remained. Of the PR-positive findings, the disagreement concerned 11 sections. Eight sections remained positive after conference.

In the first group, 34 sections of 25 female patients were examined. Immunohistochemical staining showed six sections positive for ER and four for PR. Two sections had positive staining for both receptors.

Of the six positive ER sections, four were from oedema specimens, one from a cyst, and one from a polyp. PR was positive in one polyp, two oedema, and one laryngocele section. The laryngocele sections were deemed negative (Table 4.2).

In the second group, 25 sections of 12 female patients were examined. The series contained 15 sections of oedema and 10 sections of laryngocele. Ten of the 15 oedema sections were positive for ER and six for PR. Six sections expressed ER and PR. Four of 10 laryngocele sections were positive for ER and two for PR. Only one was positive for both (Table 4.2).

In Table 4.2, the age range and mean age of the ER- and/or PR-positive patients are given. The results samples are too small to conclude anything at all out of it.

Discussion

Because voice quality is influenced by hormonal changes,¹ we wanted to determine if ER and PR could be identified in the female vocal folds. These receptors might contribute to different pathologic conditions in the vocal folds. However, ER and PR in the female vocal folds have not been shown convincingly.^{4,12-15} We studied ER and PR expression by immunohistochemistry in benign vocal fold lesions of the females and found that both ER and PR were present in the vocal folds. Tissue was immediately fixated postoperatively, which might be the primary cause of these findings in contrast to inconclusive reports based on the biopsies from cadaver specimens.



Although the number of patients was small, nearly all subjects with positive staining were already in the menopausal age, the mean age was 62 years. Ferguson et al.² states that age is associated with ER staining, with more ER staining in older patients compared with younger patients. These data have to be interpreted with caution due to the small number of subjects expressing positive ER staining. Newman et al.¹² suggested that density of hormone receptors decreases with age in vocal fold tissue possibly because of lower serum hormone levels but that it also might be a product of ageing. Age and gender were found to be correlated with PR. But in this study group, men were also included. Also in our study, the number of female subjects was too small to indicate a correlation of receptor expression to age.

We anticipated the expression of ER and PR in all lesions as our hypothesis suggested that ER and PR would be present in laryngeal mucosa. Our results though only showed ER and PR expression in oedema and laryngocele tissue. Because the goal in these types of surgery is minimal damage surgery, it was not possible to harvest great amounts of tissue for histopathological studies. When resection of more tissue around polyps, cysts, and nodules would have been possible, we might have found more ER and PR

expression. This might be studied further in patients in need of bigger laryngeal resections such as total laryngectomies where more tissue is available for histopathological investigation. Ferguson et al.² analysed eight laryngeal or hypopharyngeal squamous cell carcinomas and 12 specimens of laryngeal tissue in males, which were not involved with tumour. In the first group, all tissue was harvested in the operating room and immediately placed at 0°C until frozen at -70°C. The ER and PR localisation was in the nucleus and occurred primarily in the vocalis muscle with some evidence of receptor in the lamina propria and mesenchymal tissues. Detectable quantities of ER or PR were not present in the epithelial tissues of the larynx.² They confirmed the findings in baboons¹¹ and mice.¹⁰ Results of this study cannot be extrapolated to our study as our study population consists of females.

Although past studies have shown clear structural and functional voice changes associated with age and gender, the underlying factors that cause these changes have not been fully explained. In the search for hormone receptors in the vocal folds, this study proved only the existence of ER and PR in the vocal folds in laryngeal oedema and laryngoceles. This indicates that there might be an influence of oestrogen and progesterone on laryngeal tissues. As mentioned earlier, we hoped to find more expression in the other, smaller specimens, but ethics prevented us from extensive removal of healthy laryngeal tissue. A recent study of Voelter et al.¹⁹ confirmed in a study group including males the existence of ER and PR in the vocal fold.



We found ER and PR in laryngeal tissue with clinical and histopathological changes. It cannot be excluded that ER and PR density is a cause or a result of this pathology. A recent study showed higher serum levels of testosterone but also progesterone in males with Reinke oedema.²⁰ An interesting study was recently published where voice quality was studied in different phases of the menstrual cycle compared and in postmenopausal women.²¹ The authors concluded that women in the menstrual phase with the highest oestrogen level had the best voice quality as compared with the phase with the lowest levels. Voice quality was also significantly worse in postmenopausal women. To study the function of ER and PR in laryngeal tissue, one could prospectively study the density of ER and PR and correlate that to serum levels of oestrogen or progesterone. More interestingly, to study the interactions between the receptors on a cellular level, one could use the so-called three-dimensional cell cultures, for example, spheroids.²² In this way, one could co-culture the spheroids derived from laryngeal tissue with, for example, different concentrations of oestrogen and progesterone thus mimicking the in vivo situation. This could gain valuable insights into the functions of

oestrogen and progesterone in laryngeal tissues. Future steps might also involve the influence of oestrogens in laryngeal oedema and carcinogenesis.

Our study cannot explain the vocal changes in menopause or during sexual maturation in male and female puberty. Although our results might be considered as a first step toward obtaining more knowledge of the function of oestrogen and progesterone in laryngeal oedema.



References

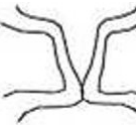
- 1 Abitbol J, Brux, J., Millot, G., Masson, M.F., Mimoun, O.L., Pau, H., & Abitbol, B. . Does a hormonal vocal cord cycle exist in women? Study of vocal premenstrual syndrome in voice performers by videoscropy-glottography and citology on 38 women. *J Voice* 1989; 3: 157-162
- 2 Ferguson BJ, Hudson WR, McCarty KS, Jr. Sex steroid receptor distribution in the human larynx and laryngeal carcinoma. *Arch Otolaryngol Head Neck Surg* 1987; 113: 1311-5
- 3 Jenkins JS. The voice of the castrato. *Lancet* 1998; 351: 1877-80
- 4 Marsigliante S, Muscella A, Resta L, Storelli C. Human larynx expresses isoforms of the oestrogen receptor. *Cancer Lett* 1996; 99: 191-6
- 5 Abitbol J, Abitbol P, Abitbol B. Sex hormones and the female voice. *J Voice* 1999; 13: 424-46
- 6 Amir O, Biron-Shental T, Muchnik C, Kishon-Rabin L. Do oral contraceptives improve vocal quality? Limited trial on low-dose formulations. *Obstet Gynecol* 2003; 101: 773-7
- 7 Hirano M, Kurita S, Sakaguchi S. Ageing of the vibratory tissue of human vocal folds. *Acta Otolaryngol* 1989; 107: 428-33
- 8 Wendler J, Siegert C, Schelhorn P, Klinger G, Gurr S, Kaufmann J, Aydinlik S, Braunschweig T. The influence of Microgynon and Diane-35, two sub-fifty ovulation inhibitors, on voice function in women. *Contraception* 1995; 52: 343-8
- 9 Lindholm P, Vilkman E, Raudaskoski T, Suvanto-Luukkonen E, Kauppila A. The effect of postmenopause and postmenopausal HRT on measured voice values and vocal symptoms. *Maturitas* 1997; 28: 47-53
- 10 Narbaitz R, Stumpf WE, Sar M. Estrogen target cells in the larynx: autoradiographic studies with 3H-diethylstilbestrol in fetal mice. *Horm Res* 1980; 12: 113-7
- 11 Aufdemorte TB, Sheridan PJ, Holt GR. Autoradiographic evidence of sex steroid receptors in laryngeal tissues of the baboon (*Papio cynocephalus*). *Laryngoscope* 1983; 93: 1607-11
- 12 Newman SR, Butler J, Hammond EH, Gray SD. Preliminary report on hormone receptors in the human vocal fold. *J Voice* 2000; 14: 72-81
- 13 Schneider B, Cohen E, Stani J, Kolbus A, Rudas M, Horvat R, van Trotsenburg M. Towards the expression of sex hormone receptors in the human vocal fold. *J Voice* 2007; 21: 502-7
- 14 Cohen E, Kolbus A, van Trotsenburg M, Rudas M, Horvat R, Schneider B. Immunohistochemical examinations of sex hormone receptors in benign vocal fold lesions. *Folia Phoniatr Logop* 2009; 61: 259-62
- 15 Nacci A, Fattori B, Basolo F, Filice ME, De Jeso K, Giovannini L, Muscatello L, Matteucci F, Ursino F. Sex hormone receptors in vocal fold tissue: a theory about the influence of sex hormones in the larynx. *Folia Phoniatr Logop* 2011; 63: 77-82
- 16 Fodor M, van Leeuwen FW, Swaab DF. Differences in postmortem stability of sex steroid receptor immunoreactivity in rat brain. *J Histochem Cytochem* 2002; 50: 641-50
- 17 Cohen DA, Dabbs DJ, Cooper KL, Amin M, Jones TE, Jones MW, Chivukula M, Trucco GA, Bhargava R. Interobserver agreement among pathologists for semiquantitative hormone receptor scoring in breast carcinoma. *Am J Clin Pathol* 2012; 138: 796-802
- 18 Wasielewski R, Hasselmann S, Ruschoff J, Fisseler-Eckhoff A, Kreipe H. Proficiency testing of immunohistochemical biomarker assays in breast cancer. *Virchows Arch* 2008; 453: 537-43
- 19 Voelter C, Kleinsasser N, Joa P, Nowack I, Martinez R, Hagen R, Voelker HU. Detection of hormone receptors in the human vocal fold. *Eur Arch Otorhinolaryngol* 2008; 265: 1239-44
- 20 Kravos A, Hocevar-Boltezar I, Gersak K. Serum levels of sex hormones in males with Reinke's edema. *Eur Arch Otorhinolaryngol* 2012
- 21 Raj A, Gupta B, Chowdhury A, Chadha S. A study of voice changes in various phases of menstrual cycle and in postmenopausal women. *J Voice* 2010; 24: 363-8
- 22 Kross KW, Heimdal JH, Olsnes C, Olofsson J, Aarstad HJ. Co-culture of head and neck squamous cell carcinoma spheroids with autologous monocytes predicts prognosis. *Scand J Immunol* 2008; 67: 392-9





PART II

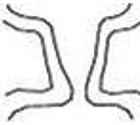
Perioperative Voice and Vocal Fold Situation





Chapter 5

Adverse Laryngeal Effects Following Short-term General Anaesthesia: a Systematic Review



Elodie J Mendels, Jan Wouter Brunings, Ankie EW Hamaekers, Robert J Stokroos,
Bernd Kremer, Laura WJ Baijens

Arch Otolaryngol Head Neck Surg. 2012;138(3):257-64

Abstract

Objective

To conduct a systematic review to determine the occurrence and type of vocal cord injury, as well as the occurrence of hoarseness, in adults using an endotracheal tube or laryngeal mask during routine anaesthetic care.

Data Sources

Two reviewers independently performed a literature search using PubMed, EMBASE, and Cochrane Central Register of Controlled Trials. The search was limited to articles published in English, German, French, or Dutch. Subsequently, reference lists of the included articles were searched manually.

Data Extraction

Studies describing vocal cord injury and/or hoarseness following short-term general anaesthesia (<5 hours) using an endotracheal tube or any type of laryngeal mask were included. To obtain a reliable outcome regarding the occurrence of anaesthesia-related laryngeal morbidity, only studies reporting both preoperative and postoperative measurements of vocal cord function were included.

Data Synthesis

A total of 4119 articles were identified; of these, 13 studies met the inclusion criteria. The studies were found to be heterogeneous and hardly comparable. Hoarseness and vocal cord injuries were common findings in most investigations.

Conclusions

Hoarseness and vocal cord injuries are clinically relevant complications related to short-term general anaesthesia using an endotracheal tube or laryngeal mask. However, more well-designed prospective studies are necessary to generate reliable data as well as to investigate techniques to reduce adverse laryngeal effects. For future research, a proposal to categorise the vocal cord lesions due to general anaesthesia is presented. Furthermore, the use of a preoperative and postoperative standardised measurement protocol using acoustic analysis and the Voice Handicap Index is advised.



Introduction

Each year, millions of patients undergo instrumentation/manipulation of the airway as part of routine anaesthetic care. Although both endotracheal intubation and the use of a laryngeal mask (LM) are associated with postoperative laryngeal morbidity, the incidence of hoarseness and vocal cord injuries is not clear.¹

Vocal cord injuries due to prolonged or difficult intubation are well known.²⁻⁴ However, thorough knowledge of the occurrence, as well as the type of injury related to short-term general anaesthesia using an endotracheal tube (ET) or LM, is needed to develop techniques for reducing the risk of adverse laryngeal effects and provide improved patient care. According to anaesthesia-related claims in a closed claims database,⁵ 7% of all claims are related to airway injury. The most frequent site of injury is the larynx, with the most common lesions being vocal cord paralysis, hematomas and granulomas (severe vocal cord lesions usually seen after prolonged intubation) of the vocal cords, and arytenoid luxation.⁶ These injuries may result in severe, prolonged laryngeal dysfunction.⁷⁻¹³

Postoperative hoarseness, an important clinical sign of laryngeal injury or dysfunction, can be distressing to a patient.^{4,8,12} It may have a negative effect on the patients' degree of satisfaction, as well as on their level of activity after hospital discharge.¹⁴ However, patients usually consult an otorhinolaryngologist only if postoperative hoarseness persists for a longer time (e.g., 6 weeks) since it is considered to be a typical effect of general anaesthesia and is expected to recover spontaneously.¹⁵

The primary aim of this article is to systematically review all published studies on the occurrence of vocal cord injuries and hoarseness in adults following short-term anaesthetic care. The types of injury are also described.

Methods

Search strategy

We performed a literature search using PubMed, the Cochrane Central Register of Controlled Trials, and EMBASE. The search was restricted to reports that were published in English, German, French, or Dutch until November 30, 2010. In PubMed, the Medical Subject Heading (MeSH) term *intubation*, *intratracheal* was combined with



the MeSH terms *laryngeal diseases* or *voice* or *speech* or *phonation* or *glottis*. More detailed information on the MeSH terms is reported in Table 5.1.

Table 5.1 Medical Subject Heading (MeSH) Terms.

Intubation, intratracheal	Laryngeal masks
Laryngeal diseases	Granuloma, laryngeal
	Laryngeal oedema
	Laryngeal neoplasms
	Laryngismus
	Laryngitis (Croup)
	Laryngomalacia
	Laryngostenosis
	Tuberculosis, laryngeal
	Vocal cord paralysis
	Voice disorders (aphonia, dysphonia, hoarseness)
Voice	Voice quality
Speech	Speech acoustics
	Speech intelligibility
Phonation	
Glottis	Vocal cord



The same search was used in the Cochrane Central Register of Controlled Trials. In EMBASE, the thesaurus terms *endotracheal intubation* or *endotracheal tube* or *laryngeal mask* were combined with *voice* or *dysphonia* or *hoarseness* or *larynx disorder* or *phonation* or *glottis* or *vocal cord* or *voice change*. To identify the most recent publications, the search was expanded by using free-text words (truncation or wildcard) *voice ** or *vocal ** combined with *intub ** in PubMed from November 17, 2008, until November 30, 2010. The reference lists of all included articles were searched for additional literature. Inclusion and exclusion criteria were defined before the systematic search (Table 5.2).

Table 5.2 Inclusion and exclusion criteria.

Type of Criteria	Items
Inclusion criteria	Studies describing vocal cord injuries and/or hoarseness following short-term general anaesthesia using an endotracheal tube or any type of laryngeal mask Short term defined as <5 h Studies reporting both pre- and postoperative measurements of vocal cord function.
Exclusion criteria	Data from children aged ≤18 y Studies describing emergency intubation and rapid sequence induction Studies including critically ill patients, patients with morbid obesity, or a history of disorders or surgery in the neck, larynx or airway

Data collection and analysis

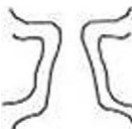
Two reviewers (E.J.M. and J.W.B.) independently based their first selection on titles and abstracts. Differences in their findings were settled by discussion. Duplicate articles from the databases were excluded manually. The original articles were used to make the definitive decision on inclusion and were independently assessed by the 2 reviewers. Only vocal cord injuries and postoperative hoarseness as outcomes were extracted for analysis.

To determine the level of evidence of the included articles, the Dutch EBRO (Evidence-Based *Richtlijn Ontwikkeling* [Guideline Development]) guideline was used (Table 5.3).¹⁶ Level D articles presenting an expert opinion were excluded from this review.

Table 5.3 Classification of Level of Evidence.*

A1	Meta-analysis containing at least some trials of level A2 and of which the results of individual trials are consistent
A2	Randomised comparative clinical trials of good quality (randomised double-blind controlled trials) of sufficient size and consistency
B	Randomised clinical trials of moderate (weak) quality or insufficient size or other comparative trials (non-randomised, cohort studies, patient-control studies)
C	Non-comparative trials
D	Expert opinion

*Translated from the Ditch EBRO guideline (Evidence-Based *Richtlijn Ontwikkeling* [Guideline Development]).



The intention was to assess the quality of the randomised controlled trials based on the methods, applied statistics, and conclusions of the authors, using the method described by Schulz et al.¹⁷ in 1995. To assess the quality of the observational studies, the Dutch EBRO guideline was used.¹⁶

Results

A total of 5079 articles were found: 2051 in PubMed, 2751 in EMBASE, 154 in the Cochrane Central Register of Controlled Trials, and 123 in PubMed using free-text words. Because of overlap, 960 articles were excluded, leaving 4119. Of these, 4037 articles were excluded when a review of the title and abstract showed that they did not meet the inclusion criteria. Thirteen of 82 full-text articles met the inclusion criteria (Figure 5.1); review of the reference lists identified no additional eligible articles. The details of the included studies are described in Table 5.4.^{15,18-29}

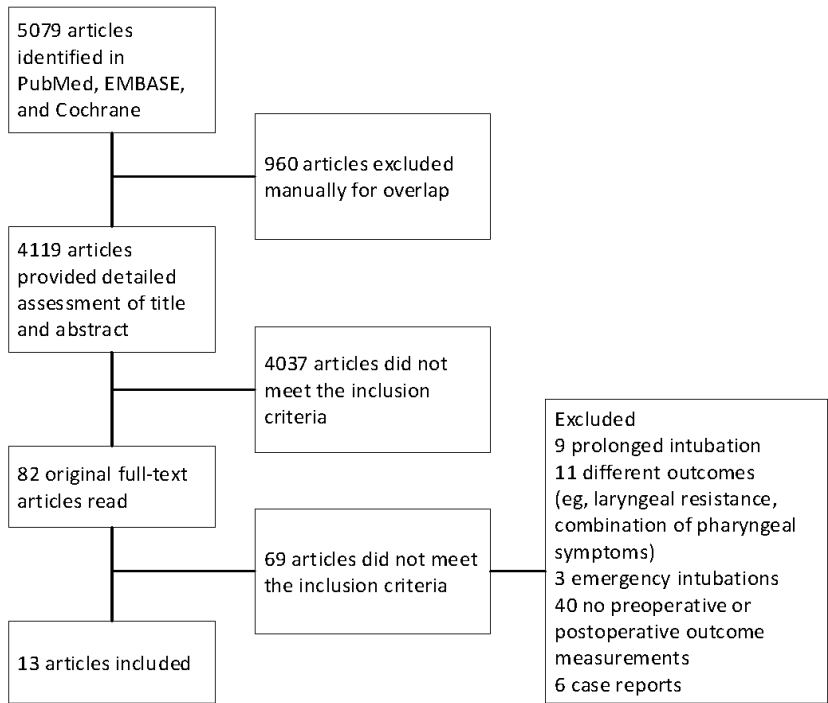


Figure 5.1 Flow diagram of the literature search. Cochrane indicates Cochrane Central Register of Controlled Trials.

General results

Of the 13 included articles, 3 were considered to meet the criteria of level A2,¹⁸⁻²⁰ 4 of level B,²¹⁻²⁴ and 6 of level C^{15,25-29} (Table 5.4). Two studies^{22,24} examined vocal cord injuries as well as hoarseness. Six studies^{18,19,23,25,26,29} examined vocal cord injuries, of which two^{26, 29} investigated only vocal cord paralysis. Five studies^{15,20,21,27,28} examined postoperative hoarseness. All studies included ET intubation; two^{21,22} also included LM insertion compared with ET intubation. The mean duration of intubation ranged from 65 to 195.5 minutes. The number of patients in the studies ranged from 6 to 235. All investigations (from 1976 to 2010) used descriptive statistics and several used statistical analysis to evaluate their results.

Measurement tools

All studies used various measurement tools and outcome assessments (Table 5.4). For the measurement of vocal cord injury, laryngoscopy or videolaryngostroboscopy was used. For the measurement of hoarseness, various instruments were applied, including several subjective interviews, acoustic analysis, and audioperceptual analysis in voice recordings. The timing of postoperative follow-up measurements varied from once immediately after the operation to 6 months postoperatively.

Occurrence of postoperative vocal cord injuries and hoarseness

Postoperative vocal cord injuries and hoarseness were common in most studies using the ET. Vocal cord injuries ranged from none of 6 participants to 69 of 100 participants. Hoarseness ranged from 4 of 10 participants immediately after the operation to 54 of 167 patients within the first postoperative week.

In one study²² using the LM, 1 participant of 21 was diagnosed as having vocal cord injury and 1 participant of 28 reported hoarseness. In the study of Hamdan et al.,²¹ hoarseness after use of the LM was also reported.

In all investigations, the number of participants having hoarseness or vocal cord injuries decreased after the first postoperative measurement. Not all studies described a follow-up until total resolution of hoarseness or vocal cord injuries had occurred. In the studies of Friedrich et al.²⁶ and Mencke et al.,¹⁸ both using an ET, persistent vocal cord injuries were described as varying from 1 week to 6 months. Jones et al.²⁸ reported on 5 participants with persistent hoarseness up to 99 days following the use of an ET. With use of laryngoscopy, vocal cord injuries were diagnosed in these patients. In Table 5.5 and Table 5.6, the included studies are summarised regarding available data on the occurrence and time of evaluation of vocal cord injuries and hoarseness.



Type of injury

Seven studies^{18,19,22,24-26,29} described the type of vocal cord injuries. The most frequent types of injuries were hematoma and thickening of the mucosa with oedema, equally located bilaterally and unilaterally (left more than right). In 5 patients, granulomas were seen (Table 5.5).

Table 5.4 Summary of the included studies.

Source	Level of evidence	Measurement tools	Outcome variables	Statistics
Mencke et al. ¹⁸	A	VCI: videolaryngostroboscopy	VCI: location and type of injury	Descriptive and statistical analysis
Mencke et al. ¹⁹	A	VCI: videolaryngostroboscopy	VCI: location and type of injury	Descriptive and statistical analysis
Jaensson et al. ²⁰	A	H: interview hoarseness	H: binary scale: yes/no Discomfort graded on a 4-point scale (0, none; 1, mild; 2, moderate; 3, severe)	Descriptive and statistical analysis
Hamdan et al. ²¹	B	H: patient questionnaire laryngopharyngeal discomfort	H: binary scale: yes/no.	Descriptive and statistical analysis
Zimmert et al. ²²	B	VCI: videoendoscopy, videolaryngostroboscopy H: preoperatively, audioperceptual analysis; postoperatively, interview hoarseness	VCI: binary scale: yes/no, changes in vocal cord function H: preoperatively, irregular voice variables; postoperatively, "How do you do?" and specific questions concerning hoarseness (not furthermentioned)	Descriptive
Lesser and Lesser ²³	B	VCI: laryngoscopy	NR	Descriptive
Beckford et al. ²⁴	B	VCI: laryngoscopy, videostroboscopy H: audioperceptual analysis	VCI: binary scale: yes/no. H: GRBAS scale	Descriptive
Chilla et al. ²⁵	C	VCI: laryngoscopy	VCI: 5 subgroups (no, little, mild, moderate, severe), as a result of a sum of points given to different types of vocal cord injuries	Descriptive
Friedrich et al. ²⁶	C	VCI: laryngoscopy	VCI: presence of movement and morphologic characteristics	Descriptive
Hamdan et al. ²⁷	C	H: patient questionnaire of laryngopharyngeal discomfort	H: binary scale: yes/no.	Descriptive and statistical analysis
Jones et al. ²⁸	C	H: preoperatively and postoperatively, audioperceptual analysis; postoperatively, interview hoarseness	H: relating a Visual Analogue Scale for hoarseness of voice recording (assessed by speech analyst) to subjective assessment of hoarseness (assessed by patient)	Descriptive
Walts et al. ²⁹	C	VCI: laryngoscopy	VCI: presence of paresis or paralysis	Descriptive
Lesser and Williams ¹⁵	C	H: acoustic analysis and interview hoarseness	H: relating fundamental frequency to a subjective assessment (asking patients how their voice had felt during reading, and marking responses this on a 10 cm linear analogue scale)	-

GRBAS: Grade, Roughness, Breathiness, Asthenia, Strain; H: Hoarseness; NR: not reported; VCI: vocal cord injury.

Comment

The present systematic review aimed to evaluate the occurrence and type of vocal cord injury as well as the occurrence of hoarseness after short-term general anaesthesia using an ET or an LM in adults. The systematic search generated a large number of articles on the occurrence of vocal cord injuries and/or hoarseness; however, few measured both hoarseness and vocal cord injuries preoperatively and postoperatively. Preoperative and postoperative assessments are necessary to measure the effects of ET or LM on vocal cord function and morphology.

It may be concluded that vocal cord injuries, as well as hoarseness, were common after short-term general anaesthesia using an ET in most studies (Tables 5.5 and 5.6). Even vocal cord injuries following LM insertion were reported²² although the LM does not traverse the vocal cords. Furthermore, 5 patients from different studies were reported to have granulomas. In addition, persistent hoarseness was reported in several studies up to 6 months after routine short-term anaesthetic care. However, the results of the included studies vary widely. Because of heterogeneous designs, such as different measurement tools, outcome variables, and postoperative measurement times, the included studies can hardly be compared with each other. Other methodological shortcomings include, for example, the absence of a control group to identify confounders,^{15,21-23,25-27,29} selection bias,^{15,21-23,27} and loss to follow-up^{21-24,27}; these limitations result in weak overall conclusions.



Times of measurement

To evaluate and compare the results of studies examining vocal cord injuries and/or hoarseness, it is important to know the time of the postoperative measurement. If more time elapses between the end of anaesthesia and the first measurement, more factors will interfere and affect the results. Six studies^{18-21,27} used standardised times of postoperative measurement. Eight studies^{18-22,24,27,29} performed the first postoperative measurement within 1 day. One author²⁴ did not report any information or included incomplete information on the time of postoperative measurement. A wide range of times of postoperative measurement, for example, within 1 to 5 days,²⁵ 4 to 9 days,²⁶ and within 1 week²⁸ was observed in several studies, resulting in less-standardised study conditions.



Table 5.5 Characteristics, occurrence, and investigators' conclusions per study regarding VCIs.

Source	Patients examined postoperatively, <i>n</i>	Time of postoperative measurement	Mean duration (range) of intubation, min	VCI preoperatively, <i>n</i>	VCI postoperatively, <i>n</i> (%)	Type of VCI, <i>n</i>	Investigators' conclusions
Beckford et al. ²⁴	9 1 drop-out	Immediately after surgery	66	0	<i>n</i> =1	Hematoma, 1	Laryngoscopy did not demonstrate consistent changes in glottic mucosal function
Chilla et al. ²⁵	100	1-5 d	100	NR	69	Not clear. Intubation complications divided into 5 groups from none to severe VCI: 26 little, 24 mild, 19 moderate	Postoperative complications were dependent on intubation time, age, sex and tube size. Asymmetry of the larynx favoured post-intubation complications
Friedrich et al. ²⁶	206 4 drop-outs	1. 4-9 d 2. 2 wk 3. 6 mo	145 (50-420)	4 excluded with recurrent paresis	13 (2 of 3 total resolution in 6 mo)	Hematoma, 3; oedema, 7; granuloma, 3; hyperemia, 7; paralysis, 3 NR	Experience of the anaesthesiologist and difficulty and duration of intubation were not related directly to recurrent nerve palsy Temporary harshness in speech following intubation was directly related to length of intubation and not to act of intubation
Lesser and Lesser ²³	6	NR	78.3 (25-195)	0	0		Quality of tracheal intubation may have affected the incidence of laryngeal morbidity Adding atracurium improved the quality of intubation
Mencke et al. ¹⁸	G1: 37 ^a G2: 36 ^b 7 drop-outs	1. 24 h 2. 72 h 3. >72h	Duration of operation G1: 70.5 (20-195) G2: 65 (10-300)	VCI excluded	G1: 1. 3 (8.1) 2. 1 (2.7) 3. 0 G2: 1. 15 (41.7) 2. 8 (22.2) 3. 2 (5.6) 2 persistent injuries for 1 and 2 wk	Hematomas, 11; thickening, 9; granulomas, 2	

Table 5.5 (continued)

Source	Patients examined postoperatively, <i>n</i>	Time of postoperative measurement	Mean duration (range) of intubation, min	VCI preoperatively, <i>n</i>	VCI postoperatively, %	Type of VCI, <i>n</i>	Investigators' conclusions
Mendcke et al. ¹⁹	G1: 27 ^c G2: 25 ^d 8 drop-outs	1. 24 h	G1: VCI excluded	G1:	Hematomas, 4; thickening, 13		Preoperative monitoring improved intubating conditions
		2. 72 h	133 (35)	1.5 (18.5)			No decrease of VCI regarding injection of atracurium at maximum neuromuscular block
		3. >72h		2.1 (3.7)			
			G2: 146 (45)	3.0			
Walts et al. ²⁹	100	< 2h		G2: 1.9 (36.0)			
			156 (25-430)	2.0	Paralysis, 0		Although soft-tissue trauma detected, no patient had evidence of vocal cord paralysis
Zimmert et al. ²²	G1: 22 ^e G2: 21 ^f 13 drop-outs	NR		G1: 6 (27.3)			Lower incidence of laryngeal sequelae and fewer reports of laryngeal disorder following LM vs ET insertion
			G1: VCI excluded	G2: 1 (4.8)			
			87 (35-140)		G1: hematoma, 2; hyperemia, 1; supraglottic hematoma, 2; subglottic hematoma		
			G2: 89 (25-215)		G2: hyperemia, 1		

ET, endotracheal tube; G, group; LM, laryngeal mask; NR, not reported; VCI, vocal cord injury. ^a Neuromuscular agent group; ^b Saline group; ^c 2-minute group; ^d Monitor group; ^e Endotracheal intubation group; ^f Laryngeal mask insertion group.





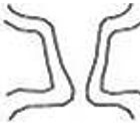
Table 5.6 Characteristics, occurrence, and investigators' conclusions per study regarding hoarseness.

Source	Examined postoperatively, n	Time of postoperative measurement	Mean duration (range) of intubation, min	Hoarseness, % preoperatively	Hoarseness, % postoperatively	Investigators' conclusions
Beckford et al. ²⁴	10	Immediately after surgery	66	NR	n= ±4	Subjective speech analysis: decreased intensity, increased roughness, lowered affect Objective not to find: multifactorial dysphonia
Hamdan et al. ²⁷	35	1. 2 h 2. 24 h	20-290	17.1	1. 47.1 2. 11.5	Vocal changes shortly after ET insertion were significant but subside within 24h. ET insertion variables: cuff mean pressure and volume are most important variables associated with increase in vocal symptoms Both LM and ET insertion can be regarded as non-traumatic, because of lack of significant or permanent vocal symptoms and acoustic changes detected 24h postoperatively
Hamdan et al. ²¹	G1:17 ^a G2:10 ^b	1. 2 h 2. 24 h	94.79	G1: 5.9 G2: 10	G1: 1. 47.1 2. 5.9 G2: 1. 44.4 2. 10	
Jaensson et al. ²⁰	G1:48 ^c G2: 49 ^d 3 drop-outs	1. PACU 2. 24 h 3. 72 h 4. 96h	G1: 195.5 G2: 190.4	NR	G1: 1. 38.3 2. 29.2 3. 12.5 4. 6.1 G2: 1. 46.9 2. 40.8 3. 12.2 4. 8.3	Use of ET size 6 instead of size 7 in women reduced incidence of sore throat as well as discomfort from a sore throat in early postoperative recovery In some patients, sore throat and hoarseness lasted up to 96h postoperatively
Jones et al. ²⁸	G1: 167 ^e G2: 68 ^f	< 1 week	NR	NR	G1: n=54 (32.3) n=5 (9, 10, 12, 54, 99 d; n=2 of 5, granulomas) G2: NR	Supports low incidence of prolonged hoarseness following short-term ET insertion

Table 5.6 (continued)

Source	Examined postoperatively, <i>n</i>	Time of postoperative measurement	Mean duration (range) of intubation, min	Hoarseness, % preoperatively	Hoarseness, % postoperatively	Investigators' conclusions
Lesser and Williams ¹⁵	G1: 22 ^e G2: 3 ^g	NR	Approximately 33.5 (20-210)	Not clear	Not clear	Subjective change in voice correlated well with objective measure of vocal cord trauma following intubation Transient postoperative hoarseness is due to laryngeal damage
Zimmert et al. ²²	G1: 28 ^a G2: 28 ^b	1. 4 h 2. 24 h	G1: 87 (35-140) G2: 89 (25-215)	Hoarseness was excluded	G1: 1. <i>n</i> =8 2. <i>n</i> =3 G2: 1. <i>n</i> =1 2. <i>n</i> =0	Lower incidence of laryngeal sequelae and fewer reports of laryngeal disorder following LM vs ET insertion

ET, endotracheal tube; G, group; LM, laryngeal mask; NR, not reported; PACU, post anaesthetic care unit. ^{a7} Endotracheal intubation group; ^{b8} Laryngeal mask group; ^{c9} Endotracheal intubation using tube size 6; ^{d10} Endotracheal intubation using tube size 7; ^{e11} General anaesthesia group; ^{f12} No anaesthesia; ^{g13} Local anaesthesia.



Measurement tools and outcome variables

Laryngoscopy is a routine procedure to examine the vocal cords.³⁰ Furthermore, videolaryngostroboscopy is a well-established technique that is used extensively for the analysis of vocal cords and voice disorders using visuoperceptual outcome variables. It can reveal a number of abnormalities.³¹⁻³⁴ All studies examining the vocal cords used these universally accepted instruments. However, the definition of vocal cord injury was not uniform²²⁻²⁴ and the classification of the type of injury varied.^{18,19,25,26,29} For example, one study²³ did not describe a classification and another²⁵ divided the vocal cord injuries into 5 subgroups (no, little, mild, moderate, and severe) as a result of a sum of points given to different types of vocal cord injuries. Furthermore, several studies reported whether lesions were present^{22,24,29} or used classifications varying in detail concerning the type and location of the lesions.^{18,19,26} Therefore, uniform nomenclature for vocal cord injury is necessary to compare the outcomes of future studies.

A classification can be made according to the impairment of the vibratory movement of the vocal cord and movement disorders of the vocal cords as suggested by the Phonosurgery Committee of the European Laryngological Society.³⁵ A revision in the classification of lesions due to ET or LM insertion in general anaesthesia is proposed (Table 5.7).



Table 5.7 Lesions due to endotracheal intubation or laryngeal mask insertion.

Type		Description
I	A	Vocal cord lesions impairing the vibratory movement: Epithelial
		1. inflammation 2. pigmentation
	B	Lamina propria
		1. Reinke's space oedema 2. hematoma 3. acquired scar/laceration
	C	Arytenoid granuloma
II	Movement disorders of the vocal cords	
	A	Neurologic: paralysis/paresis of laryngeal nerve(s)
	B	Cricoaarytenoid joint disorders (arytenoid luxation)

To evaluate hoarseness, several measurements are available.³⁶ Even though in many studies audioperceptual evaluation is considered the criterion standard for voice evaluation, it is a less reliable assessment technique because of its subjective nature. Audioperceptual evaluation involves problems such as the unstable internal standards for comparing speech stimuli and the lack of universally accepted definitions for

perceptual concepts.³⁷ Because of this, some scope for subjectivity is left to the rater in the discrimination of the ordinal audioperceptual variables. Furthermore, even though instrumental analysis is considered to be more objective, it also has limitations, including imperfection in the acoustic analysis (e.g., errors in pitch tracking, the inability to perform acoustic analysis in very aperiodic vocal vibrations, and the analysis of non-physiologic speech samples, such as sustained vowels).^{38,39} For that reason, no measurement alone can diagnose or characterise dysphonia, and multidimensional voice assessment is recommended.^{39,40} Concerning the 7 included studies on postoperative hoarseness, 4 studies used audioperceptual measurements,^{15,22,24,28} including the Grade, Roughness, Breathiness, Asthenia, Strain scale used by Beckford et al.²⁴ All studies used instrumental acoustic analysis measuring different variables.^{15,20-22,24 27,28} However, most investigations used different subjective measurements, including patient interviews and questionnaires varying in validity and reliability.^{20-22,27} For future research, combining the use of acoustic measurements and patient questionnaires—for example, the validated self-perception questionnaire Voice Handicap Index—is proposed.⁴¹ The Voice Handicap Index is a validated test used to quantify the patient's perception of his or her vocal function. This self-administered questionnaire consists of 30 questions; the patient responds according to the appropriateness of each item (0, never; to 4, always). The Voice Handicap Index is scored from 0 to 120, with the latter representing the maximum perceived disability due to voice difficulties based on the patient's response. Acoustic analysis is performed by selecting a particular segment from a voice sample and analysing it using defined acoustic algorithms: the percentage of jitter, shimmer, and noise to harmonic ratio.⁴² The percentage of jitter gives an indication of the variability of the pitch period within the analysed voice sample. It represents the relative period-to-period variability. The percentage of shimmer gives an indication of the period-to-period variability of the peak-to-peak amplitude. The evaluation of the noise present in the signal is expressed as noise to harmonic ratio: the average ratio of energy of the inharmonic components in the 1500- to 4500-Hz range to the harmonic components' energy in the 70- to 4500-Hz range.³⁹

An important condition to obtain sufficiently reliable study results is blinding of the raters combined with consensus training and independent scoring of the visuoperceptual or audioperceptual outcome variables for vocal cord injuries and hoarseness. Only 3 studies noted the presence of an expert rater.^{18,19,26} The remaining studies did not present clear information about the raters, resulting in less robust conclusions. Blinding of the raters and reliability analysis of their results are necessary to generate reliable conclusions.

Anaesthesia-related factors

Laryngeal injury caused by an ET may occur during induction of anaesthesia as direct intubation trauma, an operation, or tracheal extubation.⁴³ In addition, several risk factors leading to laryngeal injury have been described. These include ET size, cuff design, cuff pressure, type of tube, use of an introducer, use of a gastric tube, muscle relaxation, use of propofol, duration of the operation, intubation conditions, and movement of the tube, as well as demographic factors such as sex, weight, history of smoking and gastroesophageal reflux, or even the type of operation.^{1,4,7-11,18} Furthermore, removal of the LM with an inflated cuff, forced traction, or twisting of the LM may cause rotation of the larynx and possible dislocation of the arytenoids.²¹ Severe cranial nerve injuries arising from pressure trauma have been associated with the LM. More than 20 case reports of recurrent laryngeal nerve injuries have been published.⁴⁴ Two studies applied a standardised anaesthesia protocol.^{18,19} The remaining studies used a variety of anaesthesia protocols.^{15,20-22,24-29} One investigation did not mention details concerning the anaesthesia technique used.²³ Overall, this could have affected the outcomes. For future research, it is important to use a standardised anaesthesia protocol and report potential contributing factors to vocal cord injuries.

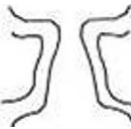


Conclusions

The clinical objective examined by the present systematic review was to evaluate the occurrence and the type of vocal cord injury as well as the occurrence of hoarseness following short-term general anaesthesia using an ET or LM in adults. All 13 of the included studies applied preoperative and postoperative measurements. Hoarseness and vocal cord injuries were common findings following the use of an ET applied for short-term anaesthesia in most studies. In addition, several investigations reported persistent hoarseness and injury for up to 6 months. However, because of heterogeneous study designs, such as different assessment tools, different outcome variables, and a wide range of postoperative measurement moments, the included studies could hardly be compared. Nevertheless, hoarseness and vocal cord injuries seem to be clinically relevant complications related to use of an ET or LM during short-term general anaesthesia.

For future research, more well-designed prospective studies with both preoperative and postoperative examinations, validated measurement tools, and reliable outcome assessments are necessary to generate reliable data as well as develop techniques to reduce adverse laryngeal effects. Furthermore, a preoperative and postoperative standardised measurement protocol using acoustic analysis and the Voice Handicap

Index is advised. We recommend that vocal cord lesions due to general anaesthesia be categorised.



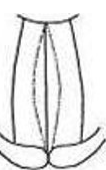
References

- 1 Maktabi MA, Smith RB, Todd MM. Is routine endotracheal intubation as safe as we think or wish? *Anesthesiology* 2003; 99: 247-8.
- 2 Ellis SF, Pollak AC, Hanson DG, Jiang JJ. Videolaryngoscopic evaluation of laryngeal intubation injury: incidence and predictive factors. *Otolaryngol Head Neck Surg* 1996; 114: 729-31.
- 3 Hamdan AL, Sabra O, Rameh C, El-Khatib M. Persistent dysphonia following endotracheal intubation. *Middle East J Anesthesiol* 2007; 19: 5-13.
- 4 Kitahara S, Masuda Y, Kitagawa Y. Vocal fold injury following endotracheal intubation. *J Laryngol Otol* 2005; 119: 825-7.
- 5 Metzner J, Posner KL, Lam MS, Domino KB. Closed claims' analysis. *Best Pract Res Clin Anaesthesiol* 2011; 25: 263-76.
- 6 Domino KB, Posner KL, Caplan RA, Cheney FW. Airway injury during anesthesia: a closed claims analysis. *Anesthesiology* 1999; 91: 1703-11.
- 7 Colice GL. Resolution of laryngeal injury following translaryngeal intubation. *Am Rev Respir Dis* 1992; 145: 361-4.
- 8 Colice GL, Stukel TA, Dain B. Laryngeal complications of prolonged intubation. *Chest* 1989; 96: 877-84.
- 9 Donati F, Plaud B. Tracheal intubation: optimal conditions, vocal cord damage, and allergy. *Can J Anaesth* 2008; 55: 663-9.
- 10 Jesudason WV, Rothera MP. An unusual pattern of intubation injury? *J Laryngol Otol* 2003; 117: 145.
- 11 Kambic V, Radsel Z. Intubation lesions of the larynx. *Br J Anaesth* 1978; 50: 587-90.
- 12 Sariego J. Vocal fold hypomobility secondary to elective endotracheal intubation: a general surgeon's perspective. *J Voice*; 24: 110-2.
- 13 Yamanaka H, Hayashi Y, Watanabe Y, Uematu H, Mashimo T. Prolonged hoarseness and arytenoid cartilage dislocation after tracheal intubation. *Br J Anaesth* 2009; 103: 452-5.
- 14 Martins RH, Braz JR, Dias NH, Castilho EC, Braz LG, Navarro LH. [Hoarseness after tracheal intubation.]. *Rev Bras Anesthesiol* 2006; 56: 189-99.
- 15 Lesser T, Williams G. Laryngographic investigation of postoperative hoarseness. *Clin Otolaryngol Allied Sci* 1988; 13: 37-42.
- 16 Gezondheidszorg Kvd. CBO Handleiding voor werkgroepleden, nov 2007.
- 17 Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995; 273: 408-12.
- 18 Mencke T, Echternach M, Kleinschmidt S, Lux P, Barth V, Plinkert PK, Fuchs-Buder T. Laryngeal morbidity and quality of tracheal intubation: a randomized controlled trial. *Anesthesiology* 2003; 98: 1049-56.
- 19 Mencke T, Echternach M, Plinkert PK, Johann U, Afan N, Rensing H, Noeldge-Schomburg G, Knoll H, Larsen R. Does the timing of tracheal intubation based on neuromuscular monitoring decrease laryngeal injury? A randomized, prospective, controlled trial. *Anesth Analg* 2006; 102: 306-12.
- 20 Jaensson M, Olowsson LL, Nilsson U. Endotracheal tube size and sore throat following surgery: a randomized-controlled study. *Acta Anaesthesiol Scand*; 54: 147-53.
- 21 Hamdan AL, Kanazi G, Rameh C, Rifai H, Sibai A. Immediate post-operative vocal changes in patients using laryngeal mask airway versus endotracheal tube. *J Laryngol Otol* 2008; 122: 829-35.
- 22 Zimmert M, Zwirner P, Kruse E, Braun U. Effects on vocal function and incidence of laryngeal disorder when using a laryngeal mask airway in comparison with an endotracheal tube. *Eur J Anaesthesiol* 1999; 16: 511-5.
- 23 Lesser TH, Lesser PJ. Laryngeal trauma vs length of intubation. *J Laryngol Otol* 1987; 101: 1165-7.
- 24 Beckford NS, Mayo R, Wilkinson A, 3rd, Tierney M. Effects of short-term endotracheal intubation on vocal function. *Laryngoscope* 1990; 100: 331-6.
- 25 Chilla R, Gabriel P, Ilse H. [The short-time-intubation as a cause of morphological and functional disorders of the larynx (author's transl)]. *Laryngol Rhinol Otol (Stuttg)* 1976; 55: 118-23.
- 26 Friedrich T, Hansch U, Eichfeld U, Steinert M, Staemmler A, Schonfelder M. [Recurrent laryngeal nerve paralysis as intubation injury?]. *Chirurg* 2000; 71: 539-44.



- 27 Hamdan AL, Sibai A, Rameh C, Kanazeh G. Short-term effects of endotracheal intubation on voice. *J Voice* 2007; 21: 762-8.
- 28 Jones MW, Catling S, Evans E, Green DH, Green JR. Hoarseness after tracheal intubation. *Anaesthesia* 1992; 47: 213-6.
- 29 Walts LF, Calcaterra T, Cohen A. Vocal cord function following short term endotracheal intubation. *Clin Otolaryngol Allied Sci* 1980; 5: 103-5.
- 30 Schade G, Hess M. [Flexible versus rigid laryngoscopy and stroboscopy. Differential findings in voice disorders]. *HNO* 2001; 49: 562-8.
- 31 Colton RH, Woo P, Brewer DW, Griffin B, Casper J. Stroboscopic signs associated with benign lesions of the vocal folds. *J Voice* 1995; 9: 312-25.
- 32 Lee JS, Kim E, Sung MW, Kim KH, Sung MY, Park KS. A method for assessing the regional vibratory pattern of vocal folds by analysing the video recording of stroboscopy. *Med Biol Eng Comput* 2001; 39: 273-8.
- 33 Popolo PS, Titze IR. Qualification of a quantitative laryngeal imaging system using videostroboscopy and videokymography. *Ann Otol Rhinol Laryngol* 2008; 117: 404-12.
- 34 Verikas A, Uloza V, Bacauskiene M, Gelzinis A, Kelertas E. Advances in laryngeal imaging. *Eur Arch Otorhinolaryngol* 2009; 266: 1509-20.
- 35 Friedrich G, Remacle M, Birchall M, Marie JP, Arens C. Defining phonosurgery: a proposal for classification and nomenclature by the Phonosurgery Committee of the European Laryngological Society (ELS). *Eur Arch Otorhinolaryngol* 2007; 264: 1191-200.
- 36 Carding PN, Wilson JA, MacKenzie K, Deary IJ. Measuring voice outcomes: state of the science review. *J Laryngol Otol* 2009; 123: 823-9.
- 37 Kreiman J, Gerratt BR, Berke GS. The multidimensional nature of pathologic vocal quality. *J Acoust Soc Am* 1994; 96: 1291-302.
- 38 Leuchter I. [Evaluation of dysphonia--a multiparametric approach]. *Rev Med Suisse* 2010; 6: 1863-4, 1866-7.
- 39 Speyer R, Wieneke GH, Dejonckere PH. Documentation of progress in voice therapy: perceptual, acoustic, and laryngostroboscopic findings pretherapy and posttherapy. *J Voice* 2004; 18: 325-40.
- 40 Hartl DM, Hans S, Crevier Buchman L, Laccourreye O, Vaissiere J, Brasnu D. [Dysphonia: current methods of evaluation]. *Ann Otolaryngol Chir Cervicofac* 2005; 122: 163-72.
- 41 Jacobson B. The Voice Handicap Index (VHI): Development and validation. *Am J Speech Lang Pathol* 1997; 6: 66-70.
- 42 Titze IR. Workshop on Acoustic Voice Analysis: summary statement. Denver, CO: National Center for Voice and Speech, 1995; 1-36.
- 43 Mencke T, Knoll H, Schreiber JU, Echternach M, Klein S, Noeldge-Schomburg G, Silomon M. Rocuronium is not associated with more vocal cord injuries than succinylcholine after rapid-sequence induction: a randomized, prospective, controlled trial. *Anesth Analg* 2006; 102: 943-9.
- 44 Brimacombe J, Clarke G, Keller C. Lingual nerve injury associated with the ProSeal laryngeal mask airway: a case report and review of the literature. *Br J Anaesth* 2005; 95: 420-3.





Chapter 6

Voice and Vocal Fold Condition Following Short-term General Anaesthesia: a Prospective Study



Jan Wouter Brunings, Sophie Vanbelle, Ankie EW Hamaekers, Bernd Kremer,
Sonja Bašić, Gusta van Zwieten, Laura WJ Baijens

Journal of Voice 2020 Jan; Epub ahead of print

Abstract

Objective

Dysphonia, with or without laryngeal changes, has been reported as a complication following prolonged intubation. In contrast, it is unknown if laryngeal changes also occur following short-term airway instrumentation. The objectives of this study were to determine the prevalence of laryngeal changes in patients undergoing short-term routine general anaesthesia using an endotracheal tube (ETT) or supraglottic airway (SGA) and to identify predictors to these changes.

Material and Methods

Standardised voice assessments were performed preoperatively, postoperatively, and at follow up on adults undergoing general anaesthesia for an elective procedure of less than three hours requiring an ETT or an SGA. The standardised voice assessment protocol comprised a rigid videolaryngostroboscopy, the Voice Handicap Index (VHI), and acoustic voice analysis.

The effects of demographic and anaesthetic characteristics and type of airway instrumentation on the videolaryngostroboscopic variables were studied using multilevel logistic regression. Multilevel linear regression was used to reveal preoperative versus postoperative changes in VHI and acoustic voice scores.

Results

Overall, the prevalence of postoperative laryngeal changes was low. Significant postoperative laryngeal changes were found for the variables right-sided vocal fold redness in the ETT group ($P=0.048$) and right-sided vocal fold blood vessels in both groups (ETT vs. SGA). However, after adjustment for all demographic and anaesthetic characteristics in the regression model, the effect of the type of airway instrumentation (ETT versus SGA) on the variable right-sided vocal fold redness was no longer significant.

Conclusion

ETT and SGA short-term airway instrumentation are vocal fold function sparing techniques with negligible laryngeal changes.



Introduction

Laryngeal complaints are common among patients following endotracheal intubation with an incidence of dysphonia varying between 14.4 and 50%.^{1,2} Even more severe findings such as glottic haematoma, laceration, and subluxation of the cricoarytenoid joint have been reported (up to 6.2%).³ Although dysphonia is reported less frequently following a supraglottic airway (SGA), injuries, such as cricoarytenoid subluxation and recurrent laryngeal nerve palsy have been reported.^{1,4,5}

Postoperative dysphonia affects social and professional performance.⁶⁻⁸ In the American Society of Anesthesiologists Closed Claims Project database 1990-2007, 7% of the 5230 claims were related to airway injury, of which 33% were of laryngeal nature.^{9,10} Risk factors for laryngeal complaints after anaesthesia include the size of the endotracheal tube (ETT), use of an introducer or neuromuscular (NM) blocking agents, intubation and extubation conditions, type and duration of surgery, and demographic factors.^{1,2,5,11,12}

Compared with the consequences of prolonged intubation, sufficient knowledge of laryngeal changes following short-term airway instrumentation is currently lacking.^{8,13-16}

A previous systematic review showed that dysphonia and laryngeal injuries are common findings following short-term general anaesthesia. However, due to the heterogeneity in designs and poor methodological quality of the included studies, a meta-analysis to determine a reliable prevalence estimate of dysphonia or a causal relationship between laryngeal changes and short-term general anaesthesia could not be carried out.¹⁴



To address this knowledge gap, this study was conducted to determine the prevalence of laryngeal changes in patients undergoing short-term routine general anaesthesia using an ETT or SGA, and to identify predictors to these changes.

Materials and methods

Ethics approval

This single centre study protocol was approved by the Medical Research Ethics Committee of the Maastricht University Medical Centre (MUMC+) (NL33665.068.10 / 10-2-075 CCMO / METC Maastricht) in the Netherlands. The study was conducted in compliance with the provisions of the Declaration of Helsinki. Written informed

consent was obtained from all patients. The study was registered in the ‘Nederlands Trial Register’ (NTR2549).

Patient population

All adult patients planned for an elective procedure under general anaesthesia between June 2011 and December 2012 were invited to participate. The following exclusion criteria were applied: less than 18 years in age, elective surgical procedures longer than three hours, American Society of Anaesthesiologists (ASA) classification above three, upper airway or neck surgery, critical illness, emergency surgery settings, cognitive impairment, and preoperative findings suggestive of head and neck cancer.

Voice and vocal fold assessment

All patients underwent a standardised assessment protocol comprised of a structured interview, Body Mass Index (BMI) measurement, the self-report Voice Handicap Index (VHI), a videolaryngostroboscopy by trained clinicians, and acoustic voice analysis. The first measurement (T0) was performed after admission on the morning of or the day before surgery. The second measurement moment (T1) took place in the time window of two hours to four days postoperatively depending on the wellbeing of the patient. The third measurement (T2) was performed one to six weeks postoperatively (Figure 6.1).

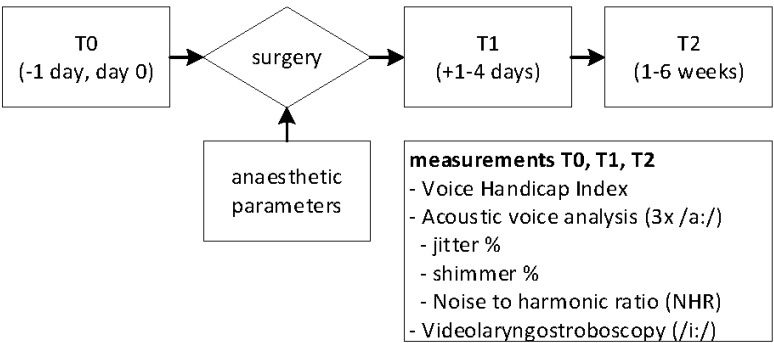


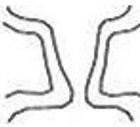
Figure 6.1 Study design

The Dutch version of the VHI was used to assess voice-related QoL. The VHI is a validated questionnaire measuring voice problems in daily life. It consists of 30 items divided into three subscales: emotional (VHI-E), functional (VHI-F), and physical (VHI-P).

Each item can be scored from 0 to 4, with 0 as 'never' and 4 as 'always'. Summing the scores on the 30 items yields a total VHI score (VHI-T) ranging from 0 to 120. The higher the score, the higher the degree of patient-perceived vocal handicap.¹⁷

All videolaryngostroboscopy videos were obtained using a 70 degree rigid Hopkins endoscope (model 8706 CA, Karl Storz GmbH & Co KG, Tuttlingen, Germany) attached to a μ -Pulsar (Karl Storz GmbH & Co KG, Tuttlingen, Germany) and a Tele Pack X (model 200450 Karl Storz GmbH & Co KG, Tuttlingen, Germany). The images were recorded on an external hard disc at 30 frames per second. If necessary, a topical anaesthetic (Xylocaine 10%) was administered to the oropharyngeal mucosa. Video recordings of vocal fold vibration were made during repeated stable phonation of a sustained vowel /i:/ at comfortable pitch and loudness. Each video contained a phonation time long enough to allow the registration of at least one complete cycle of vibration. During the examination, patients were seated upright. The field of the videolaryngostroboscopic image included the laryngeal vestibule, vocal folds, anterior and posterior commissure, and the arytenoids.

Visuoperceptual ordinal and nominal videolaryngostroboscopic variables comprised of laryngeal changes and clinical diagnosis of vocal pathology and were derived from reports of the Phonosurgery Committee of the European Laryngological Society (ELS).^{18,19-25} All of these variables were scored for each video at varying speed (slow motion, normal, up to frame-by-frame) as often as necessary using the software program Windows Movie Maker version 5.1 (Microsoft Corporation, Redmond, WA, USA). Scores were dichotomous (normal – defect) or trichotomous (none – mild – severe) (Table 6.3). Each video was initially rated by a student panel comprised of two trained observers. Subsequently, all the videos classified as abnormal by the first panel were reassessed by a second panel consisting of two expert observers. This method of panel-based consensus assessment including observer agreement levels and the training for this technique was described in previous publications.^{25,26} Both pairs of observers were blinded for the patients' identity and medical history, for the technique of airway instrumentation (ETT or SGA), for vocal sound recording, and for the preoperative versus postoperative time point (T0, T1, T2). To determine intrapanel observer agreement of the expert panel, 29% of the videos were rated twice (repeated measurements). These videos were randomly selected and again blinded for the observers. Fatigue-related observer bias was avoided by limiting the judge's rating task to two hours per session. Patients with persistent laryngeal changes following T2 underwent follow-up examination until complete remission.



All participants underwent acoustic voice analysis. Acoustic signals were collected with a computerized Speech Lab Model (4300 (CSL, Kay Elemetrics Corporation, Pine Brook, NJ, USA). The speech acoustic signal was transduced with a condenser omnidirectional microphone at a five cm mouth-to-microphone distance and 45-degree mouth-to-microphone angle. The participants were asked to produce three repeated stable phonations of a sustained vowel /a:/ at comfortable pitch and loudness. The relatively stable mid-vowel sections were analysed using the Multi-Dimensional Voice Program (MDVP, Kay Elemetrics Corporation, Pine Brook, NJ, USA). Three acoustic parameters were measured for each sample: jitter, shimmer, and noise-to-harmonic ratio (NHR). This combination appears to be successful for monitoring voice quality changes over time.²⁷ The percentage of jitter provides an indication of the variability of the pitch period within the analysed voice sample. It represents the relative period-to-period variability.²⁷ The percentage of shimmer gives an indication of the period-to-period variability of the peak-to-peak amplitude.²⁷ The evaluation of the noise present in the signal is expressed as NHR: the average ratio of energy of the harmonic components in the 1500-4500 Hz range to the harmonic components energy in the 70-4500 Hz range.²⁷



Induction and maintenance of general anaesthesia

The anaesthesia regime was left to the discretion of the anaesthesiologist who was blinded to the aim of this study. All patients received standard anaesthesia care with insertion of either an ETT or an SGA. NM blocking agents were standard in case of endotracheal intubation. However, standardised NM-monitoring was not carried out. Cuff pressure was measured in all patients. All physicians of the anaesthesiology department ($n=40$) took part in the study. The type and duration of surgery, drugs administered, type of airway management, patient position, and extubation strategy were reported on a standardised information sheet.^{1,2,11,12}

Statistics

Results were expressed as mean (standard deviation (SD)) for quantitative variables while frequencies and proportions (%) were used for categorical variables. Demographic and anaesthetic characteristics were compared between the ETT and SGA group using a Student t-test for normally distributed variables and the Wilcoxon signed-rank test otherwise. Proportions of outcome variables were compared using the chi-squared test for cross tables.

In the context of statistical dichotomisation, the videolaryngostroboscopic ordinal categories 'mild' and 'severe' of the trichotomous variables vocal fold 'redness', 'swelling', and 'blood vessels' were pooled into one category 'abnormal' to improve

statistical power. Furthermore, demographic and anaesthetic characteristics of the present study population were listed in Table 6.1. The effect of these demographic and anaesthetic characteristics and type of airway instrumentation (ETT versus SGA) on the change of the videolaryngostroboscopic variables between T0 and T1 was studied using multilevel logistic regression analyses to account for the presence of repeated measurements. In the same way, multilevel linear regression was used to reveal changes in preoperative versus postoperative VHI and acoustic voice scores. Results were considered significant at the 5% critical level ($P < .05$). All calculations were performed using SAS (version 9.1 for Windows; SAS Institute, Cary, NC, USA) statistical package. Observer agreement for the repeated measurements of categorical variables was verified using Cohen's kappa coefficient (PASW Statistic 20, SPSS Inc., Hong Kong, China).

A power analysis could not be performed before the start of this study since only heterogeneous studies on adverse laryngeal effects following short-term elective general anaesthesia have been published without any precise information on the prevalence of these adverse effects.¹⁴

Results

General results

Two hundred eighteen patients were included in the study. One hundred and one patients were lost to follow-up; gagging or excessive choking despite local oropharyngeal anaesthesia was the main reason followed by early hospital discharge and postoperative illness (Figure 6.2). Subsequently, videolaryngostroboscopic recordings of 117 patients were analysed at T0 and T1. Only 45 patients underwent measurement T2, making this group too small for further statistical analysis. The acoustic voice analysis of the first 10 patients could not be used due to a technical software problem in the equipment described above.



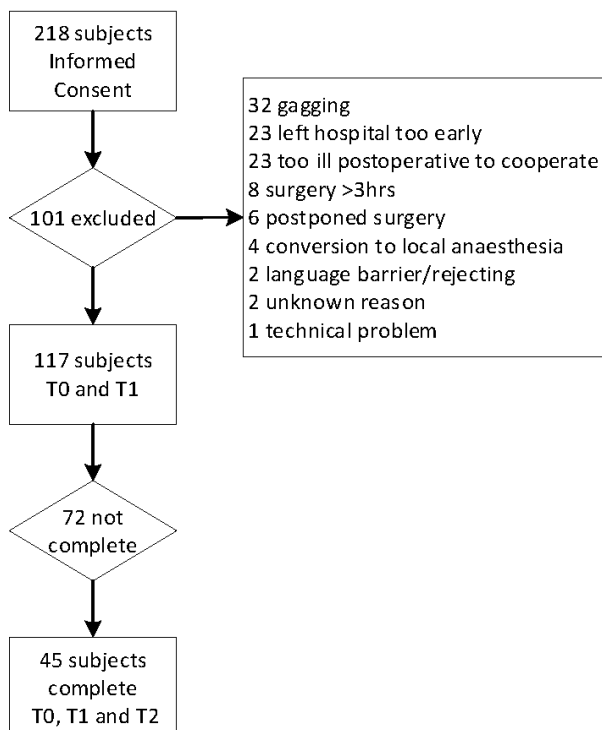


Figure 6.2 Loss to follow-up.

Patient characteristics

Sixty-seven (57.3%) of the included patients were men. The average age (SD) was 54 (14.3) years. The average BMI (SD) was 26.8 (4.8) kg m⁻². The BMI was significantly lower in the SGA group (25.8 (3.85) kg m⁻²) compared with the ETT group (27.5 (5.33) kg m⁻²) ($P=0.039$). Fifty-two (44.4%) patients were active smokers or smoked in the past, with an average (SD) of 20.7 (14.2) pack years. Thirty-two (27.4%) patients reported gastro-oesophageal reflux on a regular basis (at least once a month). Half of them ($n=16$; 16.5%) used anti-reflux medication. The surgical interventions were carried out by eight different medical disciplines (Table 6.1).

Anaesthetic characteristics

The mean duration (SD) of the general anaesthesia was 92.6 (36.2) minutes. In 65 (55.6%) patients an ETT was inserted and in the remaining 52 patients (44.4%) an SGA

was placed (Table 6.1). Twelve ETT patients (19.4%) underwent more than one intubation attempt. These patients had a significantly higher Cormack-Lehane (CL) and Mallampati score compared with patients in the ETT group undergoing just one intubation without repeated attempts (CL-score $P=0.0049$; Mallampati score $P=0.017$).

Table 6.1 Demographic and anaesthetic characteristics of the 117 patients: number (%) for qualitative variables and mean (SD) for quantitative variables. Demographic and anaesthetics characteristics were compared between the ETT and SGA group using a Student t-test for normally distributed variables and the Wilcoxon signed-rank test otherwise.

		Total <i>n</i> =117	ETT [†] <i>n</i> =65 (55.6)	SGA [‡] <i>n</i> =52 (44.4)	<i>P</i> -value
Gender	Male	67 (57.3)	37 (56.9)	30 (57.7)	0.930
	Female	50 (42.7)	28 (43.1)	22 (42.3)	
Age (yrs)		54.3 (14.3)	55.2 (15.2)	53.2 (13.2)	0.460
Length (cm)		173.4 (8.8)	172.0 (9.2)	175.2 (7.8)	0.046
Weight (kg)		80.5 (16.1)	81.5 (17.6)	79.2 (14.0)	0.450
BMI [§] (kg m ⁻²)		26.8 (4.8)	27.5 (5.3)	25.8 (3.9)	0.039
Mouth opening (cm)		4.9 (1.0)	4.9 (1.0)	4.8 (0.9)	0.640
Thyromental distance (cm)		6.2 (1.2)	6.3 (1.2)	6.1 (1.1)	0.510
Mallampati score (1-4)	1	71 (61.2)	36 (57.1)	35 (67.3)	0.370
	2	36 (31.0)	21 (33.3)	15 (28.9)	
	3	8 (6.9)	6 (9.5)	2 (3.9)	
	4	0 (0.0)	0 (0.0)	0 (0.0)	
Smoker	No	65 (55.6)	37 (56.9)	28 (53.9)	0.740
	Yes	52 (44.4)	28 (43.1)	24 (46.2)	
	PY [¶]	20.7 (14.2)	9.45 (14.4)	8.87 (13.5)	0.820
Anamnesis of reflux	No	85 (72.7)	46 (70.8)	39 (75.0)	0.610
	Yes	32 (27.4)	19 (29.2)	13 (25.0)	
Antireflux medication	No	16 (50.0)	11 (57.9)	5 (38.5)	0.280
	Yes	16 (50.0)	8 (42.1)	8 (61.5)	
Inhalation medication	No	107 (91.5)	61 (93.9)	46 (88.5)	0.300
	Yes	10 (8.6)	4 (6.2)	6 (11.5)	
Surgical discipline	ENT ^{††}	16 (13.7)	12 (18.5)	4 (7.7)	
	Gynaecology	7 (6.0)	6 (9.2)	1 (1.9)	
	Orthopaedics	38 (32.5)	23 (35.4)	15 (28.9)	
	Plastic surgery	18 (15.4)	6 (9.2)	12 (23.1)	
	Abdominal surgery	16 (13.7)	8 (12.3)	8 (15.4)	
	Ophthalmology	15 (12.8)	7 (10.8)	8 (15.4)	
	Urology	5 (4.3)	1 (1.5)	4 (7.7)	
	Vascular surgery	2 (1.7)	2 (3.1)	NA ^{‡‡}	
SGA size	LMA ^{§§} 4	22 (18.8)	NA	22 (42.3)	
	LMA 5	30 (25.6)	NA	30 (57.7)	
ETT size	ETT 6	1 (0.9)	1 (1.5)	NA	
	ETT 7	5 (4.3)	5 (7.7)	NA	
	ETT 7.5	23 (19.7)	23 (35.4)	NA	
	ETT 8	32 (27.4)	32 (49.2)	NA	
	ETT 8.5	4 (3.4)	4 (6.2)	NA	

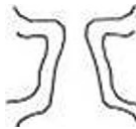


Table 6.1 (continued)

		Total n=117	ETT [†] n=65 (55.6)	SGA [‡] n=52 (44.4)	P-value
Patient position	Supine	91 (77.8)	43 (66.2)	48 (92.3)	NA
	Prone	5 (4.3)	5 (7.7)	NA	
	Recovery	20 (17.1)	16 (24.6)	4 (7.7)	
	Chair	1 (0.9)	1 (1.5)	NA	
Anaesthesia duration (min)		92.6 (36.2)	99.7 (38.2)	83.6 (31.8)	0.016
NM ^{¶¶} blocking agents	No	43 (36.8)	0 (0.0)	43 (82.7)	<0.0001
	Yes	74 (63.3)	65 (100.0)	9 (17.3)	
Cormack-Lehane score (n=60; 5 missing values)	1	48 (41.0)	48 (80.0)	NA	NA
	2	9 (7.7)	9 (15.0)	NA	
	3	3 (2.6)	3 (5.0)	NA	
Number of intubation attempts (n=62; 3 missing values)	1	50 (42.7)	50 (80.7)	NA	NA
	2	11 (9.4)	11 (17.7)	NA	
	3	1 (0.9)	1 (1.6)	NA	

[†] ETT: endotracheal tube; [‡] SGA: supraglottic airway; [§] BMI: Body Mass Index; [¶] PY: pack year; ^{††} ENT: Ear, Nose and Throat; ^{‡‡} NA: not applicable; ^{§§} LMA: laryngeal mask airway; ^{¶¶} NM: neuromuscular.

Intrapanel observer agreement

Table 6.2 shows the levels of intrapanel observer agreement of the expert panel with standard error and 95% Confidence Interval for all videolaryngostroboscopic variables. All levels of agreement were moderate to substantial (Kappa>0.50-0.80).

Table 6.2 Intrapanel observer agreement levels of the expert panel for videolaryngostroboscopic variables.

	n	% Agreement	Kappa [†]	SE [‡]	95% CI [§]
Vocal fold redness R	28	85.7	0.70	0.14	0.40 – 0.93
Vocal fold redness L	28	82.1	0.61	0.16	0.28 – 0.89
Vocal fold blood vessels R	27	81.5	0.52	0.18	0.17 – 0.84
Vocal fold blood vessels L	27	82.1	0.59	0.17	0.20 – 0.89
Vocal fold swelling R	28	82.1	0.64	0.14	0.36 – 0.88
Vocal fold swelling L	28	85.7	0.50	0.15	0.22 – 0.79
Vocal fold closure	23	82.6	0.62	0.17	0.23 – 0.91

[†] Kappa coefficient. <0 less than chance agreement; 0.01–0.20 slight agreement; 0.21– 0.40 fair agreement; 0.41–0.60 moderate agreement; 0.61–0.80 substantial agreement; 0.81–0.99 almost perfect agreement. [‡] SE: standard error; [§] CI: confidence interval.

Videolaryngostroboscopic results

Table 6.3 shows outcomes for each of the three components of the standardised voice assessment protocol at baseline (T0) and post-surgery (T1) as a function of airway instrumentation (ETT vs. SGA). Few statistically significant changes in videolaryngostroboscopic variables between T0 and T1 were found. The proportion of

abnormal right-sided vocal fold redness at T0 was similar for patients in the ETT and SGA group, the pattern changed at T1 with a greater proportion of patients in the ETT group showing abnormal right-sided vocal fold redness than in the SGA group ($P=0.048$). However, after adjustment for all demographic and anaesthetic characteristics in the multilevel logistic regression model, the effect of the type of airway instrumentation (ETT versus SGA) on right-sided vocal fold redness was no longer significant.

The next and last significant change between T0 versus T1 was found for the variable right-sided vocal fold blood vessels. The postoperative score of this variable decreased significantly ($P=0.043$) compared to the baseline score in the ETT and SGA group, independent of the type of airway instrumentation or smoking status.

Other notable findings in the present study were the significantly higher proportions of preoperative and postoperative bilateral vocal fold swelling and blood vessels in smokers compared to non-smokers independent of the type of airway instrumentation. Furthermore, a significantly higher proportion of normal vocal fold closure was seen in patients not using inhalation medication compared to users again independent of the type of airway instrumentation ($P=0.0043$). Finally, left-sided vocal fold redness significantly decreased in smokers at T1 compared to T0 ($P=0.030$).

Table 6.4 shows the clinical diagnosis of vocal pathology derived from reports of the ELS before (T0 vertical) and after (T1 horizontal) airway management for the ETT and SGA group in absolute numbers.

In the ETT group, 36 clinical diagnoses were consistent preoperative versus postoperative. Of the initial 36 normal vocal fold observations (T0), eight (22.2%) patients had a clinical diagnosis at T1 of which five (13.9%) had a clinical diagnosis of vocal fold atrophy. This atrophy was not accompanied by significant changes in other variables such as vocal fold closure and defect or vocal fold swelling.

At T1, a normal vocal fold status was observed in four (26.7%) of the 15 patients who initially had a clinical diagnosis at T0. In the SGA group, 35 clinical diagnoses were consistent preoperative versus postoperative. Of the initial 36 normal vocal fold observations (T0), three (8.3%) patients had a clinical diagnosis at T1. At T1, a normal vocal fold status was observed in five (55.6%) of the 9 patients who initially had a clinical diagnosis at T0.





Table 6.3 Descriptive statistics of the baseline data (T0) and the postsurgical data (T1) for videolaryngostroboscopy, the VHI questionnaire, acoustic voice analysis, the number of patients per airway management group, and the level of significance of the difference between postsurgical data (T1) compared to baseline data (T0) for the patients in the ETT and SGA group.

Variable	Type of airway management	n	Before intubation T0		After intubation T1		Worse	Better	Predictors to laryngeal changes	P-value
			None n(%)	Mild + severe n(%)	None n(%)	Mild + severe n(%)				
Videolaryngostroboscopic variables										
Vocal fold redness R	ETT [†]	53	44 (83.0)	9 (17.0)	38 (71.7)	15 (28.3)	9	3	Increase redness R in ETT at T1	0.048
	SGA [‡]	45	36 (80.0)	9 (20.0)	38 (84.4)	7 (15.6)	2	4		
Vocal fold redness L	ETT	53	43 (81.1)	10 (18.9)	40 (75.5)	13 (24.5)	8	5	Decrease redness L in smokers at T1	0.030
	SGA	45	36 (80.0)	9 (20.0)	39 (86.7)	6 (13.3)	1	4		
Vocal fold swelling R	ETT	53	47 (88.7)	6 (11.3)	42 (79.3)	11 (20.8)	6	1	Overall higher proportion swelling	0.0003
	SGA	45	40 (88.9)	5 (11.1)	41 (91.1)	4 (8.9)	2	3	R in smokers at T0 and T1	
Vocal fold swelling L	ETT	53	48 (90.6)	5 (9.4)	45 (84.9)	8 (15.1)	3	0	Overall higher proportion swelling	0.0018
	SGA	45	40 (88.9)	5 (11.1)	40 (88.9)	5 (11.1)	2	2	L in smokers at T0 and T1	
Vocal fold blood vessels R	ETT	52	40 (76.9)	12 (23.1)	44 (84.6)	8 (15.4)	2	6	1) Overall higher proportion vessels R in smokers at T0 and T1	1) 0.0003
	SGA	45	35 (77.8)	10 (22.2)	38 (34.4)	7 (15.6)	1	4	2) Decrease vessels R in smokers and non-smokers at T1	2) 0.043
Vocal fold blood vessels L	ETT	52	41 (78.9)	11 (21.2)	43 (82.7)	9 (17.3)	4	6	Overall higher proportion vessels	0.0003
	SGA	45	36 (80.0)	9 (20.0)	39 (86.7)	6 (13.3)	0	3	L in smokers at T0 and T1	
Vocal fold closure and Defect	ETT	50	25 (50.0)	25 (50.0)	14 (28.0)	36 (72.0)	13	2	Overall higher proportion normal closure in group without inhalation medication than in group with inhalation medication	0.0043
	SGA	45	19 (42.2)	26 (57.8)	23 (51.1)	22 (48.9)	5	9		
Redness arytenoid	ETT	18	16 (88.9)	2 (11.1)	11 (61.1)	7 (38.9)	5	0	Increase redness arytenoid in ETT at T0 and T1	0.011
	SGA	8	8 (100)	0	8 (100)	0	0	0	at T1	

Table 6.3 (continued)

Variable	Type of airway management	n	Before intubation T0		After intubation T1		Worse	Better	Predictors to laryngeal changes	P-value
			None n(%)	Mild + severe n(%)	None n(%)	Mild + severe n(%)				
VHI [§] scores										
			Before intubation T0		After intubation T1		Change			
			Mean (SD)	Median	Mean (SD)	Median	Mean (SD)	Median		
VHI total	ETT	65	6.2 (9.8)	2.0	7.9 (13.9)	2.0	1.7 (8.6)	0.0	None	0.049
	SGA	52	6.9 (12.0)	2.0	6.1 (12.6)	2.0	-0.84 (3.7)	0.0		
Acoustic voice variables										
			Before intubation T0		After intubation T1		Change			
			Mean (SD)	Median	Mean (SD)	Median	Mean (SD)	Median		
Jitter	ETT	57	0.90 (0.66)	0.6	0.9 (0.74)	0.7	0.1 (0.86)	0.0	None	0.042
(norm ≤ 1.040%)	SGA	45	1.2 (1.82)	1.0	1.2 (1.87)	0.7	0.0 (0.57)	-0.1		
Shimmer	ETT	57	3.5 (2.1)	3.1	3.8 (2.3)	3.1	0.2 (1.94)	-0.1	None	
(norm ≤ 3.810%)	SGA	45	4.3 (3.4)	3.7	4.2 (2.2)	2.9	-0.1 (2.52)	0.0		
NHR [¶]	ETT	57	0.131 (0.017)	0.129	0.134 (0.026)	0.131	0.003 (0.024)		Decrease NHR in SGA at T1	
(norm ≤0.190)							0.002			
	SGA	45	0.147 (0.092)	0.136	0.138 (0.066)	0.127	-0.010 (0.037)	-		
							0.010			

[†] ETT: endotracheal tube; [‡] SGA: supraglottic airway; [§] VHI: Voice Handicap Index; [¶] NHR: Noise-to-Harmonic Ratio. R, right-sided; L, left-sided.



Table 6.4 Clinical diagnosis of vocal pathology derived from reports of the Phonosurgery Committee of the European Laryngological Society (ELS) before (T0 vertical) and after (T1 horizontal) airway management for the ETT[†] and SGA[‡] group in absolute numbers.

		After intubation T1							
ETT group		Total	Normal	Reinke [§] 1	Reinke 3	Atrophy	Granuloma	Nodules	Consistent diagnosis
Before intubation T0	Normal	36		1		5	2		28
	Reinke 2	3			1				2
	Cyst R	1						1	
	Hematoma	2	1						1
	Atrophy	3	2						1
	Granuloma	1							0
	Laryngitis	1	1						1
	Nodules	1							0
	Cyst L	1					1		1
	Laryngitis	1							0
	Sicca								
	Leukoplakia	1							0
	Total	51	4	1	1	5	3	1	36
		After intubation T1							
SGA group		Total	Normal	Reinke 2	Polyp R	Atrophy	Laryngitis		Consistent diagnosis
Before intubation T0	Normal	36		1		1	1		33
	Reinke 1	2	2						2
	Reinke 2	1					1		1
	Cyst R	2	1		1				2
	Atrophy	2	2						2
	Laryngitis	1						1	0
	Cyst L	1						1	0
	Total	45	5	1	1	1	2		35
									10

[†] ETT: endotracheal tube; [‡] SGA: supraglottic airway; [§] Reinke: Reinke space oedema grade 1, 2 and 3. R, right-sided; L, left-sided.

Results of the VHI and acoustic voice analysis

The VHI questionnaire was reviewed for possible floor and ceiling effects, noting the number of respondents who obtained the lowest or highest possible scores. The floor effect was significant because 36.8% (n=43 out of 117) of the respondents got the lowest possible score for the VHI-T at T0 and 37.1% (n=43 out of 116) at T1. Only 22.2% (n=26) of the patients at T0 versus 19% (n=22) at T1 scored ten or more points on the VHI-T, indicating a pathologic VHI score (cut-off score of ten).^{28,29} The threshold values of the measured acoustic voice variables were $\leq 1.040\%$ for jitter, $\leq 3.810\%$ for shimmer, and ≤ 0.190 for the NHR (Table 6.3).

Multilevel linear regression did not reveal any significant preoperative versus postoperative differences for the VHI-T and subscale scores, nor for the measured

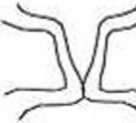
acoustic voice outcome scores in either group (ETT versus SGA). Therefore, only the VHI-T scores were presented (Table 6.3).

Despite the fact that for the SGA group the mean (SD) T0 value for shimmer was higher than the threshold value (4.3% (3.4)) and that the NHR scores decreased at T1, multilevel linear regression did not reveal any significant differences between the preoperative versus postoperative measured acoustic voice outcome scores in both groups (ETT versus SGA).

Discussion

This is the first and largest prospective study that systematically investigated the prevalence of laryngeal changes following short-term routine anaesthesia using an ETT or SGA in adults and using high quality validated assessment tools. Laryngeal changes following prolonged intubation have been described in the literature, but the prevalence of laryngeal changes due to short-term airway instrumentation remained unclear because of poor methodological quality of previous studies and heterogeneity in the design of these studies.²⁶

All patients in the current study underwent short-term airway instrumentation in the same hospital. The use of strict and clearly defined inclusion and exclusion criteria monitors the external validity of the current target sample. Moreover, the sample size consisted of 117 patients whose demographics are seen in the larger population. For example, the average BMI of the study population was 26.8 kg m^{-2} which is similar to the average BMI of 26.2 kg m^{-2} for Dutch men and 25.3 kg m^{-2} for women.³⁰ Finally, by using a large number of anaesthesiologists, possible therapist effects on group performance or treatment outcome were minimised.



In the present study, a multidimensional validated voice assessment protocol was applied to capture different aspects of the vocal function as recommended by the Phonosurgery Committee of the ELS. In addition to vocal fold imaging and patient self-assessment of voice, the analysis of vocal sound production is an indispensable dimension of the voice assessment protocol.

The prevalence of laryngeal changes between T1 and T0 using videolaryngostroboscopic variables was very low. Previous studies described higher prevalence values for dysphonia and vocal fold problems following short-term airway instrumentation. For example, Böttcher and colleagues reported vocal fold lesions in 30.2 % ($n=16$) of their participants following short-term intubation.³¹ Similar findings

were described by Mencke and colleagues with a frequency of 27% ($n=14$) vocal fold problems in their study population.³² Several other studies found a prevalence of vocal fold injuries ranging from 14.5 to 50%.¹⁴ However, these studies were small and frequently showing diverse methodological shortcomings. The requirements to obtain reliable study results are among others blinding of the observers combined with consensus training for the visuoperceptual variables applied and independent scoring to determine observer agreement levels. The three studies of Mencke and colleagues did not apply blinding of observers and reliability analysis, as only one observer scored the videolaryngostroboscopic videos.^{2,12,14,30} The same methodological observation was done for the study of Beckford and colleagues not providing any information on the videolaryngostroboscopic assessment protocol making it impossible to repeat the study.³³ Another methodological issue in previous studies on this topic were small sample sizes such as in the study of Lesser and Lesser just including six patients.³⁴



In the present study, changes to some videolaryngostroboscopic outcomes were observed. Specifically, the score of the variable right-sided vocal fold redness was significantly increased following general anaesthesia in the ETT group but the same pattern was not seen with the patients in the SGA group ($P=0.048$). However, after adjustment for all demographic and anaesthetic characteristics in the regression model, the type of airway instrumentation was no longer significant. Counter-intuitive was the finding of decreased right-sided vocal fold blood vessels following general anaesthesia (T1 versus T0; $P=0.043$) independent of the type of airway instrumentation. A decreased score on that variable following general anaesthesia with an ETT carried out by a population of anaesthesiologists who are usually right-handed, can be expected as a result of local contact between the ETT and the right vocal fold compressing the blood vessels. However, a decreased score on this same variable following SGA placement is more difficult to explain as there is no contact with the vocal folds during SGA airway instrumentation.

In addition to an instrumental assessment of the voice, attention was also paid to the patient's perception of his/her own voice. Because of "no vocal fold contact", we expected an advantage of SGA insertion compared to ETT on VHI and acoustic voice analysis outcomes. However, robust multilevel linear regression did not reveal any significant differences between the preoperative and postoperative (T0 versus T1) scores of the VHI-T and of the measured acoustic voice outcome variables in either group (ETT versus SGA). The latter results compare with two previous small studies ($n=35$ and 10 patients) examining short-term airway instrumentation using ETT; there

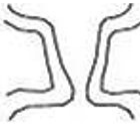
too none of the acoustic voice variables was associated with ETT-related parameters.^{16,33}

Currently, the VHI is one of the most reliable self-report tools to assess voice-related QoL.^{35,36} Originally the Dutch version of the VHI was validated for mixed aetiologies of dysphonia.²⁸ This validated tool uses a cut-off point to determine the relevant impact of dysphonia on the health-related QoL. The current patient population had a very low prevalence of voice complaints at T0, reflecting a VHI-T floor effect. However, no significant increase in VHI-T score was observed at T1 for the total group (ETT and SGA) showing that patients did not experience any adverse effects from short-term airway instrumentation.

Limitations of the study

A significant number of patients was lost to follow up after completing the written informed consent.

However, a sample size of 117 patients was believed to be sufficient to reveal laryngeal changes between T0 versus T1 following short-term anaesthesia. The majority of the patients cancelled the measurements at T2 because they did not experience any vocal or laryngeal complaints. It remains difficult to motivate participants to return to the hospital for follow-up measurements in the context of scientific research if they do not experience any complaints.



Conclusion

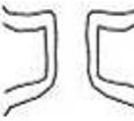
ETT and SGA short-term airway instrumentation are vocal fold function sparing techniques with negligible laryngeal changes.

References

1. Reber A, Hauenstein L, Echternach M. [Laryngopharyngeal morbidity following general anaesthesia. Anaesthesiological and laryngological aspects]. *Anaesthesist* 2007; 56:177-89; quiz 190-171.
2. Mencke T, Echternach M, Kleinschmidt S et al. Laryngeal morbidity and quality of tracheal intubation: a randomized controlled trial. *Anesthesiology* 2003; 98:1049-56.
3. Kambic V, Radsel Z. Intubation lesions of the larynx. *British journal of anaesthesia* 1978; 50:587-90.
4. Higgins PP, Chung F, Mezei G. Postoperative sore throat after ambulatory surgery. *Br J Anaesthesia* 2002; 88:582-4.
5. Hamdan AL, Kanazi G, Rameh C, Rifai H, Sibai A. Immediate post-operative vocal changes in patients using laryngeal mask airway versus endotracheal tube. *J Laryngol Otol* 2008; 122:829-35.
6. Sario J. Vocal fold hypomobility secondary to elective endotracheal intubation: a general surgeon's perspective. *J Voice* 2010;24:110-2.
7. Colice GL, Stukel TA, Dain B. Laryngeal complications of prolonged intubation. *Chest* 1989; 96:877-84.
8. Kitahara S, Masuda Y, Kitagawa Y. Vocal fold injury following endotracheal intubation. *J Laryngol Otol* 2005; 119:825-7.
9. Metzner J, Posner KL, Lam MS, Domino KB. Closed claims' analysis. *Best Pract Res Clin Anaesthesiol* 2011; 25:263-76.
10. Domino KB, Posner KL, Caplan RA, Cheney FW. Airway injury during anesthesia: a closed claims analysis. *Anesthesiology* 1999; 91:1703-11.
11. Friedrich T, Hansch U, Eichfeld U, Steinert M, Staemmler A, Schonfelder M. [Recurrent laryngeal nerve paralysis as intubation injury?]. *Chirurg* 2000; 71:539-44.
12. Mencke T, Knoll H, Schreiber JU et al. Rocuronium is not associated with more vocal cord injuries than succinylcholine after rapid-sequence induction: a randomized, prospective, controlled trial. *Anesth Analg* 2006; 102:943-9.
13. Abouleish AE, Dexter F, Whitten CW, Zavaleta JR, Prough DS. Quantifying net staffing costs due to longer-than-average surgical case durations. *Anesthesiology* 2004; 100:403-12.
14. Mendels EJ, Brunings JW, Hamaekers AE, Stokroos RJ, Kremer B, Baijens LW. Adverse laryngeal effects following short-term general anesthesia: a systematic review. *Arch Otolaryngol Head Neck Surg* 2012; 138:257-64.
15. Ellis SF, Pollak AC, Hanson DG, Jiang JJ. Videolaryngoscopic evaluation of laryngeal intubation injury: incidence and predictive factors. *Otolaryngol Head Neck Surg* 1996; 114:729-31.
16. Hamdan AL, Sabra O, Rameh C, El-Khatib M. Persistent dysphonia following endotracheal intubation. *Middle East J Anesthesiol* 2007; 19:5-13.
17. Hakkesteegt MM, Wieringa MH, Gerritsma EJ, Feenstra L. Reproducibility of the Dutch version of the Voice Handicap Index. *Folia Phoniatr logopaedica* 2006; 58:132-8.
18. Verdonschot RJ, Baijens L, Vanbelle S, Florie M, Kremer B, Leue C. The relationship between fiberoptic endoscopic evaluation of swallowing outcome and symptoms of anxiety and depression in dysphagic patients. *Laryngoscope* 2016; 126:E199-207.
19. Friedrich G, Remacle M, Birchall M, Marie JP, Arens C. Defining phonosurgery: a proposal for classification and nomenclature by the Phonosurgery Committee of the European Laryngological Society (ELS). *Eur Arch Otorhinolaryngol* 2007; 264:1191-200.
20. Baijens LW, Pilz W, Kremer B, Passos VL. Identifying Patterns of FEES-Derived Swallowing Trajectories Using Group-Based Trajectory Model. *Dysphagia* 2015; 30:529-39.
21. Baijens LW, Speyer R, Passos VL, Pilz W, Roodenburg N, Clave P. Swallowing in Parkinson Patients versus Healthy Controls: Reliability of Measurements in Videofluoroscopy. *Gastroenterol Res Pract* 2011; 2011:380682.
22. Baijens LW, Speyer R, Passos VL et al. Surface electrical stimulation in dysphagic Parkinson patients: a randomized clinical trial. *Laryngoscope* 2013; 123:E38-44.
23. Baijens LW, Speyer R, Pilz W, Roodenburg N. FEES protocol derived estimates of sensitivity: aspiration in dysphagic patients. *Dysphagia* 2014; 29:583-90.



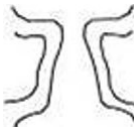
24. Pilz W, Baijens LW, Passos VL et al. Swallowing assessment in myotonic dystrophy type 1 using fiberoptic endoscopic evaluation of swallowing (FEES). *Neuromuscul Disord* 2014; 24:1054-62.
25. Pilz W, Vanbelle S, Kremer B et al. Observers' Agreement on Measurements in Fiberoptic Endoscopic Evaluation of Swallowing. *Dysphagia* 2016; 31:180-7.
26. Brunings JW, Vanbelle S, Akkermans A, et al. Observer Agreement for Measurements in Videolaryngostroboscopy. *J Voice* 2018;32(6):756-62.
27. Dejonckere PH, Bradley P, Clemente P, et al. 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.
28. Verdonck-de Leeuw I. M. B-KB, van Gogh C.D.L., Greven A.J, Peeters JAJGE, Mahieu HF. Een multidimensioneel stemonderzoeksprotocol. Het meten van stemkarakteristieken van patiënten met een klein larynxcarcinoom. *Tijdschrift voor logopedie en foniatrie* 2002; 74:340-5.
29. Van Gogh CD, Mahieu HF, Kuik DJ, Rinkel RN, Langendijk JA, Verdonck-de Leeuw IM. Voice in early glottic cancer compared to benign voice pathology. *Eur Arch Otorhinolaryngol* 2007; 264:1033-8.
30. Blokstra A VP, Venmans LMAJ, Holleman P, van der Schouw YT, Smit HA, Verschuren WMM. Measuring the Netherlands: A monitoring study of risk factors in the general population, 2009-2010. *RIVM Rapport* 260152001 2011.
31. Bottcher A, Mencke T, Zitzmann A et al. Laryngeal injuries following endotracheal intubation in ENT surgery: predictive value of anatomical scores. *Eur Arch Otorhinolaryngol* 2014; 271:345-52.
32. Mencke T, Echternach M, Plinkert PK, et al. Does the timing of tracheal intubation based on neuromuscular monitoring decrease laryngeal injury? A randomized, prospective, controlled trial. *Anesth Analg* 2006; 102:306-12.
33. Beckford NS, Mayo R, Wilkinson A, 3rd, Tierney M. Effects of short-term endotracheal intubation on vocal function. *Laryngoscope* 1990; 100:331-6.
34. Lesser TH, Lesser PJ. Laryngeal trauma vs length of intubation. *J Laryngol Otol* 1987; 101:1165-7.
35. Jacobson B. The Voice Handicap Index (VHI): Development and validation. *Am J Speech Lang Pathol* 1997; 6:66-70.
36. Carding PN, Wilson JA, MacKenzie K, Deary IJ. Measuring voice outcomes: state of the science review. *J Laryngol Otol* 2009; 123:823-9.





PART III

General Discussion, Valorisation Addendum and Summary

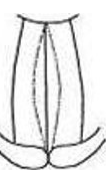




Chapter 7

General Discussion





General Discussion

Assumptions in laryngology were the basis for this dissertation. An assumption is an unexamined belief. An assumption is something that one accepts to be accurate, even without proof. In clinical practice, doctors and other medical professionals base their conclusions often on implicit assumptions. However, we should be alert to these assumptions because they can be incorrect or misleading.

Laryngology is the medical (sub)speciality focussing, among other things, on diagnosis and treatment of voice disorders. Preservation or restoration of voice function and thus human expression and communication is the core business of laryngology. Although often underestimated in comparison with other medical conditions like cancer or stroke, voice disorders severely affect the quality of life.¹ The societal burden of voice disorders is huge; in the U.S. the overall annual costs caused by voice disorders are 5 billion US dollar.^{2,3}

Every day, thousands of patients all over the world are medically examined because of impaired voice function or dysphonia.⁴ Every region or continent has laryngological societies that develop gold standards and guidelines for diagnostics and therapy in laryngology.^{5,6}

If possible guidelines are developed using evidence-based medicine. However, levels of evidence are often unsatisfactory and expert-based opinions are incorporated in guidelines. We have identified several expert opinion-based assumptions in laryngology, partly concerning the assessment of vocal function, and partly concerning perioperative vocal function.

This dissertation aimed at improving the body of evidence in laryngology by supporting or disproving the hereafter mentioned assumptions concerning vocal function.

Assessment of Voice and Vocal Function

According to the guidelines of the European Laryngological Society (ELS),⁵ the following multidimensional measuring instruments of voice function are preferred in daily practice: audioperceptual voice evaluation, videolaryngostroboscopy, acoustics, aerodynamics and subjective self-rating by the patient, usually via the Voice Handicap Index (VHI).⁷ One of the aerodynamic measurement tools is the maximum phonation time (MPT)⁸ which is considered to be an objective measurement tool. Many studies have described maximum phonation time in diverse populations and under various testing conditions,⁹ however, limited information was available on the reliability of the



data over time.¹⁰ The purpose of the study in chapter 2 was to determine the reliability of the MPT in dysphonic and control subjects as a function of the number of trials, days, and observers. The results of our study were in line with those of other studies in literature,¹¹ showing a significantly shorter MPT in dysphonic patients compared to that of healthy subjects. MPT was found to be a reliable instrument for voice assessment, intraobserver and interobserver reliability was high and therefore a single observer was sufficient to provide highly reliable measurements. Nevertheless, it has still not been proven to be a useful discriminative tool resulting in changed MPT after therapy.¹² Voice therapy may not necessarily result in a prolonged MPT. Bearing the above in mind, it remains to be discussed whether the MPT is a useful measurement tool for the assessment of the voice in daily clinical practice. Further research is needed to prove the clinical value of MPT in the various laryngological diagnoses and in relation to the therapy used.



According to the guidelines of the ELS⁵ and the American Speech-Language-Hearing Association (ASHA)¹³ videolaryngostroboscopy is the gold standard diagnostic tool to examine vocal fold movement disorders. The guidelines define variables to be assessed, but no objective, measurable parameters are defined. Therefore it is not clear whether all visuoperceptual variables are assessed uniformly by different medical stakeholders. For this reason, the purpose of the study in chapter 3 was to evaluate the intraobserver, interobserver, and intrapanel observer agreement for measurements in videolaryngostroboscopy and to compare the observers' behaviour during independent versus consensus panel rating. We have shown that the value of a single clinician's observation is of poor methodological quality.¹⁴ The intraobserver agreement level of several visuoperceptual variables was fair to moderate. Therefore we recommend using an experienced panel composed of, for example, a laryngologist and a speech and language pathologist to improve the observer agreement level during the videolaryngostroboscopic examination in daily clinical practice. The results of this study and the resulting recommendations are in line with other studies of our laryngological research group. These studies on the assessment of the visuoperceptual variables of fiberoptic/flexible endoscopic evaluation of swallowing and videofluoroscopy of swallowing showed comparable results.¹⁵⁻¹⁷

A possible reason for bias in the interpretation of videolaryngostroboscopy is the sound of the patient's voice during the judgement, which along with the patient's anamnesis already leads to a differential diagnosis. For example, the sound of a typical heavy smoker's voice is recognisable even to a non-trained clinician. Therefore in chapter 4, the videolaryngostroboscopy was judged without the sound of the patient's voice, which differentiates our study from studies in literature.^{18,19}

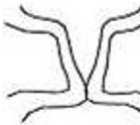
It is known that sex hormones influence the human voice. For example, the voice of boys changes during puberty, the voice of females changes during postmenopausal transformation, and in women changes of voice are dependent on their hormonal cycle.^{20,21} The exact mechanism of these sex hormone-dependent voice changes is not known yet. In the study described in chapter 4, we hypothesised that sex hormone receptors are present in vocal fold tissues, although in a previous study, Schneider et al. did not find sex hormone receptors in human vocal fold cadaver specimens.²² We, therefore, performed a similar study with slight adjustments in the methodology.²³ During phonosurgery in female patients, we took vocal fold biopsies, which were fixated immediately and stained for oestrogen and progesterone receptor antibodies. We found receptors for both oestrogen and progesterone. A new study of Kirgezen et al.,²⁴ published four years after our publication, also demonstrated the existence of sex hormone receptors within different subunits of the vocal fold, mainly in the macula flava and vocal ligaments.

Of course, the existence of hormone receptors in vocal folds does not explain the exact mechanism of sex hormone effect. Therefore a much larger study including tissue from healthy vocal folds of people across all ages would be necessary. However, the feasibility of such a study is limited by ethical concerns.

Perioperative Voice and Vocal Fold Situation

Long-term endotracheal intubation is associated with postoperative hoarseness or laryngeal morbidity.^{25,26} Short-term airway instrumentation is assumed to cause less injury, but noteworthy side-effects have been reported.²⁷⁻²⁹ An alternative to endotracheal intubation is a supraglottic airway (SGA) insertion which is presumed to result in less laryngeal damage compared to endotracheal intubation.^{26,30,31} We performed a systematic review of the literature (chapter 5) and concluded that in literature hoarseness and vocal fold injuries are reported to be clinically relevant complications following both endotracheal tube (ETT) and SGA placement even after short-term general anaesthesia. However, the quality and quantity of the available literature were limited. Only 13 articles of the initially generated 5079 hits reported both on preoperative and postoperative measurements of vocal function. Only these studies were included for analysis. The heterogeneous study designs using different assessment tools, different outcome variables, and a wide range of postoperative measurement moments, made the included studies difficult to compare.³²

For that reason, we conducted a prospective study (chapter 6) to evaluate voice and vocal fold condition following short-term general anaesthesia using a standardised



preoperative and postoperative assessment protocol including videolaryngostroboscopy, acoustic analysis and the VHI.

The ELS-guideline⁵ endorses to assess the visuoperceptual variables amplitude of the vocal fold, phase symmetry, periodicity, mucosal wave and opening during the glottic cycle. Because the level of observer agreement of the variables amplitude, symmetry, and periodicity was only fair (kappa coefficient 0.21-0.40), as studied in chapter 4, we recommended observing modified videolaryngostroboscopic variables. We proposed a more clinical classification of lesions: redness of the vocal fold, swelling, existence of hypervascularisation, closing patterns and closing defects.

The results of this study showed that ETT and SGA short-term airway instrumentation are vocal fold function sparing techniques with negligible laryngeal changes. Perhaps a different choice of assessment tools would have led to a different outcome? However, the assessment tools used in the current study are worldwide very frequently used instruments to assess vocal function in daily clinical practice.

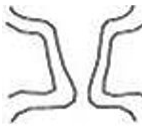


During the evaluation of the data in chapter 6, a methodological limitation became clear being the use of the VHI in a study population with a low prevalence of voice complaints. The VHI is an instrument to measure the level of voice impairment experienced by the patient. The current patient population had a very low prevalence of voice complaints at the different measurement moments, reflecting a floor effect of the VHI. Therefore it had no added value for the measurement in our study. However, besides the Voice-Related Quality of Life (V-RQOL) measure, there is no other validated questionnaire for voice complaints concerning short-term differences.³³ The question is whether very slight voice changes can be measured at all in an initial population without voice complaints using a patient self-report questionnaire?

Despite the results of this study, we cannot exclude that some vocal discomfort is caused during airway instrumentation, but statistically, the current findings were negligible. However, the translation of this data into daily clinical practice is a challenge: can we guarantee that the voice of a professional singer will not be harmed during airway instrumentation? Therefore more data are necessary which can be generated in a large, prospective multicentre study.

References

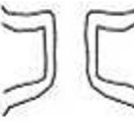
- 1 Cohen SM, Dupont WD, Courey MS. Quality-of-life impact of non-neoplastic voice disorders: a meta-analysis. *Ann Otol Rhinol Laryngol* 2006; 115: 128-34
- 2 Cohen SM, Kim J, Roy N, Asche C, Courey M. The impact of laryngeal disorders on work-related dysfunction. *Laryngoscope* 2012; 122: 1589-94
- 3 Cohen SM, Kim J, Roy N, Asche C, Courey M. Direct health care costs of laryngeal diseases and disorders. *Laryngoscope* 2012; 122: 1582-8
- 4 Mehta DD, Hillman RE. Current role of stroboscopy in laryngeal imaging. *Curr Opin Otolaryngol Head Neck Surg* 2012; 20: 429-36
- 5 Dejonckere PH, 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 Otorhinolaryngol* 2001; 258: 77-82
- 6 Stachler RJ, Francis DO, Schwartz SR, Damask CC, Digoy GP, Krouse HJ, McCoy SJ, Ouellette DR, Patel RR, Reavis CCW, Smith LJ, Smith M, Strode SW, Woo P, Nnacheta LC. Clinical Practice Guideline: Hoarseness (Dysphonia) (Update). *Otolaryngol Head Neck Surg* 2018; 158: S1-S42
- 7 Verdonck-de Leeuw I. M. B-KB, van Gogh C.D.L., Greven A.J, Peeters J.A.J.G.E., Mahieu H.F. Een multidimensioneel stemonderzoeksprotocol. Het meten van stemkarakteristieken van patiënten met een klein larynxcarcinoom. *Tijdschrift voor logopedie en foniatrie* 2002; 74: 340-345
- 8 Speyer R, Bogaardt HC, Passos VL, Roodenburg NP, Zumach A, Heijnen MA, Baijens LW, Fleskens SJ, Brunings JW. Maximum phonation time: variability and reliability. *J Voice* 2010; 24: 281-4
- 9 Baken RJO, R. F. . Clinical Measurement of Speech and Voice. Wien: Springer-Verlag, 2000
- 10 Lewis K, Casteel R, McMahon J. Duration of sustained /a/ related to the number of trials. *Folia Phoniatr (Basel)* 1982; 34: 41-8
- 11 Raes JP, Clement PA. Aerodynamic measurements of voice production. *Acta Otorhinolaryngol Belg* 1996; 50: 293-8
- 12 Treole K, Trudeau MD. Changes in sustained production tasks among women with bilateral vocal nodules before and after voice therapy. *J Voice* 1997; 11: 462-9
- 13 Association AS-L-H. Vocal tract visualization and imaging: technical report, 2004
- 14 Brunings JW, Vanbelle S, Akkermans A, Heemskerk NMM, Kremer B, Stokroos RJ, Baijens LWJ. Observer Agreement for Measurements in Videolaryngostroboscopy. *J Voice* 2017
- 15 Baijens L, Barikroo A, Pilz W. Intrarater and interrater reliability for measurements in videofluoroscopy of swallowing. *Eur J Radiol* 2013; 82: 1683-95
- 16 Baijens LW, Speyer R, Passos VL, Pilz W, Roodenburg N, Clave P. Swallowing in Parkinson Patients versus Healthy Controls: Reliability of Measurements in Videofluoroscopy. *Gastroenterol Res Pract* 2011; 2011: 380682
- 17 Pilz W, Vanbelle S, Kremer B, van Hooren MR, van Becelaere T, Roodenburg N, Baijens LW. Observers' Agreement on Measurements in Fiberoptic Endoscopic Evaluation of Swallowing. *Dysphagia* 2016; 31: 180-7
- 18 Dejonckere PH, Crevier L, Elbaz E, Marraco M, Millet B, Remacle M, Woisard V. Quantitative rating of video-laryngostroboscopy: a reliability study. *Rev Laryngol Otol Rhinol (Bord)* 1998; 119: 259-60
- 19 Nawka T, Konerding U. The interrater reliability of stroboscopy evaluations. *J Voice* 2012; 26: 812.e1-10
- 20 Abitbol J, Abitbol P, Abitbol B. Sex hormones and the female voice. *J Voice* 1999; 13: 424-46
- 21 Abitbol J, Brux, J., Millot, G., Masson, M.F., Mimoun, O.L., Pau, H., & Abitbol, B. . Does a hormonal vocal cord cycle exist in women? Study of vocal premenstrual syndrome in voice performers by videostroboscopy-glottography and cytology on 38 women. *J Voice* 1989; 3: 157-162
- 22 Schneider B, Cohen E, Stani J, Kolbus A, Rudas M, Horvat R, van Trotsenburg M. Towards the expression of sex hormone receptors in the human vocal fold. *J Voice* 2007; 21: 502-7
- 23 Brunings JW, Schepens JJ, Peutz-Kootstra CJ, Kross KW. The expression of estrogen and progesterone receptors in the human larynx. *J Voice* 2013; 27: 376-80



- 24 Kirgezen T, Sunter AV, Yigit O, Huq GE. Sex Hormone Receptor Expression in the Human Vocal Fold Subunits. *J Voice* 2017; 31: 476-482
- 25 Mencke T, Echternach M, Kleinschmidt S, Lux P, Barth V, Plinkert PK, Fuchs-Buder T. Laryngeal morbidity and quality of tracheal intubation: a randomized controlled trial. *Anesthesiology* 2003; 98: 1049-56
- 26 Reber A, Hauenstein L, Echternach M. [Laryngopharyngeal morbidity following general anaesthesia. Anaesthesiological and laryngological aspects]. *Anaesthesist* 2007; 56: 177-89; quiz 190-1
- 27 Kitahara S, Masuda Y, Kitagawa Y. Vocal fold injury following endotracheal intubation. *J Laryngol Otol* 2005; 119: 825-7
- 28 Ellis SF, Pollak AC, Hanson DG, Jiang JJ. Videolaryngoscopic evaluation of laryngeal intubation injury: incidence and predictive factors. *Otolaryngol Head Neck Surg* 1996; 114: 729-31
- 29 Hamdan AL, Sabra O, Rameh C, El-Khatib M. Persistent dysphonia following endotracheal intubation. *Middle East J Anesthesiol* 2007; 19: 5-13
- 30 Hamdan AL, Kanazi G, Rameh C, Rifai H, Sibai A. Immediate post-operative vocal changes in patients using laryngeal mask airway versus endotracheal tube. *J Laryngol Otol* 2008; 122: 829-35
- 31 Higgins PP, Chung F, Mezei G. Postoperative sore throat after ambulatory surgery. *Br J Anaesth* 2002; 88: 582-4
- 32 Mendels EJ, Brunings JW, Hamaekers AE, Stokroos RJ, Kremer B, Baijens LW. Adverse laryngeal effects following short-term general anesthesia: a systematic review. *Arch Otolaryngol Head Neck Surg* 2012; 138: 257-64
- 33 Hogikyan ND, Sethuraman G. Validation of an instrument to measure voice-related quality of life (V-RQOL). *J Voice* 1999; 13: 557-69



Valorisation Addendum



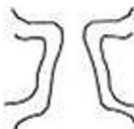


Valorisation Addendum

The social value of this thesis is reflected mostly in the counselling of patients. We now know that there are hormone receptors in the vocal folds. Patients' questions about the influence of their hormonal status on their voice can be answered better. This applies, for instance, to the mutation voice among boys, and the ageing voice among ladies. It also shines a light on the assumed impact of the hormone balance on the singing voice of professional singers. Although the study shows that hormone receptors exist in the vocal folds, we cannot yet assess how these receptors relate to a change of voice.

As a result of the chapters on vocal fold damage after short-term intubation, we can now give patients better information about the risks of vocal damage after placement of an endotracheal tube or laryngeal mask. According to our research, short term tube placement doesn't increase the risk of perioperative vocal injury. Previously published data on claims regarding airway damage due to intubation didn't take the duration into account.¹

The results of the studies on observer agreement of both videolaryngostroboscopic images and maximum phonation time (MPT) provide direction on the clinical assessment of patient data. In particular, regarding videolaryngostroboscopy, one has to become more nuanced. At least two people (laryngologist and speech and language pathologist) should observe the videolaryngostroboscopy to improve the assessment. The MPT, on the other hand, can be measured by only one person.



These insights are primarily interesting for care providers in the field of laryngology (laryngologists and speech and language pathologists). The Dutch Guideline for Voice Complaints now recommends that laryngologists and speech and language pathologists work together within one institute.² The new insights suggest collaboration should even take place within one room during the consultation. The patient has the advantage of being assessed multidisciplinary.

A multidisciplinary out-patient voice clinic is standard in an academic care setting in the Netherlands but is often not organised in a peripheral hospital.

Setting up a multidisciplinary out-patient voice clinic requires commitment and enthusiasm from the team and, unfortunately, a conviction within the group stakeholders. This kind of consultation takes more time. However, experience has shown that a repeat factor (patients visiting for a second or third time) can decrease

with correct, reliable assessment during a consultation. Long-term, this will result in less healthcare consumption.

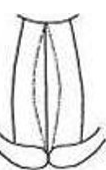
The results of this dissertation cannot be converted into market opportunities, let alone a cost-benefit analysis. However, education of the various professional groups, supported by well-founded research, can provide direction.



References

- 1 Metzner J, Posner KL, Lam MS, Domino KB. Closed claims' analysis. Best Pract Res Clin Anaesthesiol 2011; 25: 263-76
- 2 (NVKNO) NVvK-N-OeHvhH-H. Richtlijn Stemklachten volume 2015





Summary





Summary

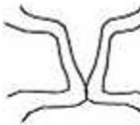
This dissertation deals with assumptions in laryngology. In the prologue, a short introduction is given about assumptions and how they are intertwined in our daily life. There are many assumptions within laryngology, some of which encouraged further investigation.

The first assumption, described in **chapter 2**, deals with the maximum phonation time (MPT). The MPT is used in daily practice as a non-invasive, fast and inexpensive measurement tool. The ability to hold a vowel for a long time after maximum inspiration is considered a measure of the efficiency of the respiratory mechanism during phonation. In the literature, the MPT has been used to objectify the severity of dysphonia and to determine the effects of voice therapy. But is MPT a reliable measurement tool as a function of the number of tests, different measurement days and assessors for dysphonic and healthy persons?

Two groups of adult subjects participated in this reliability study: a group of patients with functional or organic dysphonia and a control group of healthy individuals. Over a period of 6 weeks, three video recordings were made of five subjects' MPT trials. A panel of five experts assessed all measurements, including a repeated measurement of the subjects' first recordings. The group of dysphonic persons showed a significantly shorter MPT compared to the healthy controls (on average 6.6 seconds shorter). There was a high intrarater reliability of the measurements.

The MPT has proven to be a very reliable measure in voice assessment. A single rater is sufficient, making a panel of raters superfluous. The results of the study suggest that neither a learning effect nor a fatigue effect occurred in either group.

Clinicians around the world apply videolaryngostroboscopy in daily clinical practice, using many variables to describe vocal fold functioning and the vibratory pattern. According to the literature en guidelines, amplitude, periodicity, symmetry, closure, closure defect and anatomical integrity of the vocal folds are examined. The assumption is that laryngologists can all properly assess these images and that this assessment is also reliable. **Chapter 3** states that the visuoperceptual assessment of videolaryngostroboscopy is subject to errors. Factors such as experience, fatigue of the observer, the severity of the deviation, etc. can play a role here. There is still no good standardisation and validation of measurement criteria. The validity of a measurement scale, such as a three-point scale "good, better, best" starts with a good evaluator agreement. Previous studies have been done where the methodology can be questioned; usually, the interobserver agreement varied considerably; the expert level



was not clear, nor the size of the study group; it was not clear whether the sound was part of the videos; whether or not assessors were trained, and so on.

The study included 64 patients with dysphonia of different aetiology. They all received standardised videolaryngostroboscopy, and 2 trained experts each assessed the same videos individually and in a panel. After 6 months, another 32 images were reassessed, the images were played without sound and the raters were blinded to the anamnesis.

In particular for periodicity and to a certain extent also for amplitude and symmetry, the rater agreement is not good. In general, the intra-rater agreement scored better than the inter-rater agreement for certain measurement variables. The intra-panel agreement levels scored slightly better than the intra-rater agreement during the individual assessments. In consensus, the observation-behaviour of the raters deviated considerably from the scores during the individual assessments

Observer agreement for videolaryngostroboscopic measurements has important implications, not only on the diagnosis and treatment of dysphonic patients but also on the interpretation of results of videolaryngostroboscopic outcome parameters used in scientific research. Identification of factors that influence the observer agreement can provide a better understanding of the rating process and its limitations. The results of this study suggest that better observer agreements could be achieved by assessing visuoperceptual variables in a panel situation of the laryngologist and speech and language pathologist who work together at the outpatient voice clinic.

The larynx responds to sex hormones and is said to be an androgen-sensitive organ. This is mainly due to the changes in the male voice during puberty. The absence of testosterone before puberty in a castrate leads to a female voice. The quality of the female voice changes during the menstrual cycle, pregnancy or menopause. In menopause, these changes are explained by a reduced concentration of oestrogens and progesterone and an increase in androgenic dominance in the serum. The underlying pathophysiology is unknown. Is it true that there is an expression of oestrogen and progesterone receptors in the human larynx? The study in **chapter 4** aims to demonstrate the presence of oestrogen and progesterone receptors in the female larynx by examining the expression of these receptors in biopsies of benign vocal fold lesions.

In a first group, 34 sections were viewed of 25 women who underwent surgery for benign vocal fold lesions. These lesions included polyps, cysts, nodules, laryngeal oedema, laryngocele, and granulomas. A second group of 25 sections of 12 women with only laryngeal oedema or laryngoceles was then added. Immunohistochemical staining according to the protocol from the pathological department were done for oestrogen receptor (ER) and progesterone receptor (PR). The positive control came from breast tissue.



The sections were viewed independently by 2 pathologists, and a separate session followed later about the sections that needed conference. The staining of ER and PR was particularly evident in fibroblasts that lay next to small vessels and more prominently in laryngeal oedema.

At first glance, more laryngocele sections seemed positive for ER. However, after the conference session among the pathologists, these were all rated as negative.

In the search for sex hormone expression in human vocal folds, the presence of oestrogen and progesterone receptors was found by chance in laryngeal oedema and laryngoceles. There is no explanation for the voice changes in women, but the first step towards more knowledge about oestrogens in laryngeal oedema of human vocal folds has been made.

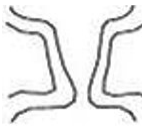
Every year millions of patients undergo airway manipulation or instrumentation as part of routine anaesthetic care. Although both endotracheal intubation and the use of a laryngeal mask are associated with postoperative laryngeal morbidity, the incidence of hoarseness and vocal fold injury is not clear. Postoperative hoarseness is an important clinical sign of laryngeal injury or dysfunction.

Patients usually consult a laryngologist if postoperative hoarseness persists because it is considered, and assumed, as a typical side effect of general anaesthesia and is expected to recover spontaneously.

In **chapter 5** a systematic review is presented focusing on the incidence and type of vocal fold pathology, and postoperative hoarseness due to general anaesthesia with endotracheal intubation or insertion of a laryngeal mask in adults.

Two independent authors carried out a search strategy with the help of PubMed, EMBASE and Cochrane Library. The search strategy was limited to English, German, French and Dutch publications. Reference lists were viewed for additional articles. Studies describing vocal fold pathology and/or hoarseness due to general anaesthesia (<8 hours) with endotracheal intubation or insertion of a type of laryngeal mask were included. To obtain a reliable incidence of anaesthesia-induced laryngeal morbidity, only studies were reported that reported both pre- and post-operative findings. 4,119 articles were found of which thirteen were included. There was heterogeneity, making it difficult to compare the articles. However, hoarseness and vocal fold pathology were regularly occurring findings in the majority of studies. To generate reliable data and to develop techniques to reduce laryngeal complications, more well-designed prospective studies are needed.

This finding was the advance to the development of the study presented in **chapter 6**; a prospective study of vocal fold abnormalities after short-term intubation.

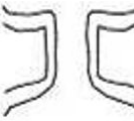


Preoperatively and postoperatively, standardised measurements were performed on 117 patients who underwent general anaesthesia for less than 3 hours before an elective procedure using an endotracheal tube (ETT) or laryngeal mask, also called supraglottic airway (SGA). The standard measuring protocol consisted of a rigid videolaryngostroboscopy, the Voice Handicap Index (VHI) and performing acoustic analysis. The effects of demographic and anaesthetic characteristics and type of airway management on the videolaryngostroboscopic variables were studied using multilevel logistic regression. Multilevel linear regression was used to reveal preoperative versus postoperative changes in VHI and acoustic voice scores.

Overall, the prevalence of postoperative laryngeal changes was low. Significant postoperative laryngeal changes were found for the variables right-sided vocal fold redness in the ETT group ($P=0.048$) and right-sided vocal fold blood vessels in both groups (ETT vs. SGA). However, after adjustment for all demographic and anaesthetic characteristics in the regression model, the effect of the type of airway instrumentation (ETT versus SGA) on the variable right-sided vocal fold redness was no longer significant. From this, it is concluded that short-term intubation with ETT or SGA is a vocal fold function sparing technique resulting in negligible laryngeal changes.



Dutch Summary





Samenvatting

Dit proefschrift handelt over aannames in de laryngologie. In de proloog wordt een korte inleiding gegeven over aannames en hoe die verweven zijn in ons dagelijkse leven.

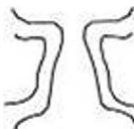
Binnen de laryngologie zijn er vele aannames waarvan een aantal stimuleerde om nader onderzocht te worden.

De eerste aanname, beschreven in **hoofdstuk 2**, handelt over de maximale fonatietijd (maximum phonation time, MPT). De MPT wordt in de praktijk gebruikt als een niet-invasief, snel en goedkoop meetinstrument. Het vermogen om een klinker maximaal lang aan te houden na maximale inspiratie, wordt beschouwd als maat voor de efficiëntie van het ademhalingsmechanisme tijdens fonatie. In de literatuur is de MPT gebruikt om de ernst van dysfonie te objectiveren en de effecten van stemtherapie te bepalen. Maar, is MPT een betrouwbaar meetinstrument als functie van het aantal testen, verschillende meetdagen en beoordelaars bij dysfone en gezonde personen?

Twee groepen volwassen proefpersonen namen deel aan deze betrouwbaarheidsstudie: een groep poliklinische patiënten met functionele of organische dysfonie en een controle groep van gezonde personen. In een periode van 6 weken zijn drie video-opnames gemaakt van de uitgevoerde MPT van vijf proefpersonen. Een panel van vijf experts beoordeelde alle metingen, inclusief een herhaalde meting van de eerste opnames. De groep dysfone personen vertoonde een significant kortere MPT in vergelijking met de gezonde controles (gemiddeld 6,6 seconden korter). Er was een hoge intraraterbetrouwbaarheid van de metingen.

De MPT is een zeer betrouwbare maat gebleken voor stemevaluatie. Een enkele beoordelaar is voldoende waardoor een panel van beoordelaars overbodig is. De resultaten van de studie suggereren dat noch een leereffect noch een vermoeidheidseffect zich voordeed in beide groepen.

In de hele wereld worden videolaryngostroboscopieën gedaan ter beoordeling van de stemplooien. Volgens de literatuur en richtlijnen wordt er gekeken naar amplitude, periodiciteit, symmetrie, sluiting, sluitingsdefect en anatomische integriteit van de stemplooi. De aanname is dat laryngologen deze beelden allemaal goed kunnen beoordelen en dat deze beoordeling ook betrouwbaar is. In **hoofdstuk 3** wordt gesteld dat de visuoperceptieve beoordeling van videolaryngostroboscopie onderhevig is aan fouten. Hierbij kunnen factoren spelen als ervaring, vermoeidheid van de observator, de ernst van de afwijking etc. Er is nog geen goede standaardisatie en validatie van



meetcriteria. De validiteit van een meetschaal, zoals een driepuntschaal 'goed, beter, best' begint bij een goede beoordelaarovereenkomst.

Er zijn eerder studies gedaan waar vraagtekens kunnen worden gesteld bij de methodologie. Meestal varieerde de interobserverovereenkomst aanzienlijk; het expertniveau was niet duidelijk, de grootte van studiegroep, het wel of geen geluid bij de beelden horen, het wel of niet getraind zijn van beoordelaars, enzovoort.

In de studie werden 64 patiënten met dysfonie van verschillende etiologie geïnccludeerd. Ze kregen allemaal een gestandaardiseerde stroboscopie en 2 getrainde experts scoorden ieder afzonderlijk en in panel dezelfde video's. Na 6 maanden werden nog eens 32 beelden herbeoordeeld. De beelden werden zonder geluid afgespeeld en de beoordelaars waren geblindeerd voor de anamnese.

Met name voor periodiciteit en in zekere mate ook voor amplitude en symmetrie is de beoordelaarovereenstemming niet goed. In het algemeen scoorde de intrabeoordelaarovereenkomst beter dan de interbeoordelaarovereenkomst voor bepaalde meetvariabelen. De intrapanelovereenkomstniveaus scoorden in lichte mate beter dan de intrabeoordelaarovereenkomst tijdens de individuele beoordelingen. In consensus devieerde het observatiegedrag van de beoordelaars aanzienlijk ten opzichte van de scores tijdens de individuele beoordelingen. Beoordelaarovereenkomst bij videolaryngostroboscopische metingen heeft behoorlijke implicaties, niet alleen op de diagnose en behandeling van dysfone patiënten, maar ook op de interpretatie van resultaten van videolaryngostroboscopische uitkomst parameters die in wetenschappelijk onderzoek worden gebruikt. Identificatie van factoren die de beoordelaarovereenkomst beïnvloeden geven beter begrip van het beoordelingsproces en de limitatie die daar aan verbonden is.

De resultaten van deze studie suggereren dat betere beoordelaarovereenkomsten zouden kunnen worden behaald door het beoordelen van visuoperceptieve variabelen in een panel situatie zoals bijvoorbeeld door een laryngoloog/KNO-arts samen met een logopedist.

De larynx reageert op geslachtshormonen en heet een androgeen gevoelig orgaan te zijn. Dit met name vanwege de veranderingen van de mannelijke stem in de puberteit. De afwezigheid van testosteron voor de puberteit in een castraat leidt tot een vrouwelijke stem. De kwaliteit van de vrouwelijke stem verandert tijdens de menstruele cyclus, een zwangerschap of menopauze.

In de menopauze worden deze veranderingen verklaard door een verlaagde concentratie oestrogenen en progesteron en een verhoging van androgene dominantie in het serum.



De onderliggende pathofysiologie is onbekend. Is het nu zo dat er expressie is van oestrogeen en progesteronreceptoren in de humane larynx? Het doel van de studie in **hoofdstuk 4** is de aanwezigheid van oestrogeen en progesteronreceptoren in de vrouwelijke larynx aan te tonen door expressie van deze receptoren te onderzoeken in bipten van benigne stemplooi afwijkingen.

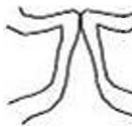
In een eerste groep werden 34 coupes bekeken van 25 vrouwen die werden geopereerd aan benigne stemplooi afwijkingen. De afwijkingen behelsde poliepen, cysten, noduli, Reinke oedeem, laryngocele en granulomen. Een tweede groep van 25 coupes van 12 vrouwen met alleen Reinke oedeem of laryngoceles werd vervolgens toegevoegd. Immunohistochemische kleuringen volgens protocol van de pathologische afdeling werden gedaan voor oestrogeenreceptor (ER) en progesteronreceptor (PR). De positieve controle kwam van mamma weefsel.

De coupes werden door 2 pathologen onafhankelijk van elkaar bekeken, en later volgde een aparte sessie over de coupes waar verdeeldheid over was.

De aankleuring van ER en PR was met name te zien in fibroblasten die naast kleine vaatjes lagen en meer prominent in laryngeaal oedeem.

Bij de eerste blik leken er veel laryngocele coupes positief voor ER. Echter, na consensusdiscussie tussen de pathologen werden deze allen als negatief afgegeven.

In de zoektocht naar expressie van geslachtshormonen in de humane stemplooien werd bij toeval de aanwezigheid van oestrogeen- en progesteronreceptoren gevonden in Reinke-oedeem en laryngoceles. Een uitleg voor de stemveranderingen bij vrouwen is er niet, maar de eerste stap naar meer kennis over oestrogenen in Reinke-oedeem van humane stemplooien is gelegd.



Elk jaar ondergaan miljoenen patiënten manipulatie van de luchtwegen als onderdeel van routinematige anesthesie. Hoewel zowel endotracheale intubatie als het gebruik van een larynxmasker geassocieerd zijn met postoperatieve laryngeale morbiditeit, is de incidentie van heesheid en stemplooi schade niet duidelijk. Postoperatieve heesheid is een belangrijk klinisch teken van larynxletsel of dysfunctie.

Patiënten raadplegen meestal een laryngoloog als postoperatieve heesheid langer aanhoudt, omdat het wordt beschouwd, en aangenomen, als een typische bijwerking van algemene anesthesie en naar verwachting spontaan zal herstellen.

In **hoofdstuk 5** wordt een systematisch literatuuronderzoek gepresenteerd gericht op de incidentie en type stemploopathologie, en postoperatieve heesheid als gevolg van algemene anesthesie met endotracheale intubatie of insertie van een larynxmasker bij volwassenen.

Twee onafhankelijke auteurs voerden een zoekstrategie uit met behulp van PubMed, EMBASE en Cochrane Library. De zoekstrategie werd beperkt tot Engels-, Duits-, Frans-

en Nederlandstalige publicaties. Referentielijsten werden bekeken voor aanvullende artikelen. Studies die stemplooiopathologie en/of heesheid als gevolg van algemene anesthesie (<8 uur) met endotracheale intubatie of insertie van een type larynxmasker beschreven, werden geïnccludeerd. Om een betrouwbare incidentie van anesthesie geïnduceerde laryngeale morbiditeit te verkrijgen, werden alleen studies geïnccludeerd die zowel pre- als postoperatieve bevindingen rapporteerden. 4119 artikelen werden gevonden waarvan dertien werden geïnccludeerd. Er was sprake van heterogeniteit, waardoor de artikelen moeilijk vergelijkbaar waren. Toch waren heesheid en stemplooi pathologie regelmatig voorkomende bevindingen in de meerderheid van de studies. Om betrouwbare data te genereren en technieken te ontwikkelen om laryngeale complicaties te reduceren, zijn meer goed ontworpen prospectieve studies noodzakelijk.

Deze bevinding was de opmars naar de ontwikkeling van de studie gepresenteerd in **hoofdstuk 6**; een prospectieve studie naar stemplooi-afwijkingen na kortdurende intubatie.

Pre- en postoperatief werden gestandaardiseerde metingen verricht bij 117 patiënten die voor een electieve ingreep minder dan 3 uur onder algehele anesthesie werden gebracht met gebruikmaking van een endotracheale tube (ETT) of larynxmasker, ook wel supraglottic airway (SGA). Het standaard meetprotocol bestond uit een starre videolaryngostroboscopia, het invullen van de Voice Handicap Index (VHI) en het verrichten van akoestische analyse.

De effecten van demografische en anesthesische karakteristieken en type luchtweg management op de videolaryngostroboscopische variabelen werden bekeken door middel van multilevel logistische regressie. Multilevel lineaire regressie werd gebruikt om pre- versus postoperatieve veranderingen in VHI en akoestische analyse aan te tonen.

In het geheel was de prevalentie van postoperatieve laryngeale veranderingen laag. Significante veranderingen werden gevonden bij de variabele rechtszijdige stemplooiroodheid in de ETT groep ($P=0.048$) en rechtszijdige stemplooiwaatinjectie in beide groepen (ETT vs. SGA). Echter na aanpassing voor alle demografische en anesthesische karakteristieken in het regressiemodel, was het effect van het type luchtwegmanagement op de variabele rechtszijdige stemplooiroodheid niet meer significant. Hieruit wordt geconcludeerd dat kortdurende intubatie met ETT of SGA een stemplooisparende techniek is met verwaarloosbare laryngeale veranderingen tot gevolg.



Epilogue





Een terugblik

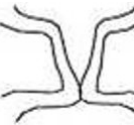
Na 19 jaar Maastrichtse tijd en een proefschrift als mijlpaal is het tijd om eens achterom te kijken, niet alleen naar het promotietraject, al dekt die tijd ook alweer een aantal jaren, maar een era op zich.

Het gebruik te starten met de promotor dient gevolgd te worden. Mijn promotor: **Professor Bernd Kremer**. Bernd was nog 1 maand mijn opleider in 2005 maar is vooral mijn collega en leidend afdelingshoofd. Mijn promotie bekeek hij vanaf de zijlijn, daarin de ruimte latend voor de copromotoren. Naast een zeer kundig en alwetend arts is hij voor mij persoonlijk vooral een groot voorbeeld op het leidende vlak. Door zijn ondersteuning en niet nalatende adviezen ben ik inmiddels gekomen waar ik ben. Tegenwoordig naast hem te mogen staan betekent zeer veel voor mijn ontwikkeling als beroepsbeoefenaar en als mens. Mijn volgende doel: streven naar de helft van zijn politieke kwaliteiten.

Dankzij mijn copromotoren **Laura Baijens** en **Ankie Hamaekers** is dit boekje gevuld geraakt. Met Ankie heb ik tijdens een cursuslunch het onderwerp van hoofdstuk 6 bedacht. Zuiver gestuurd door nieuwsgier en het proberen aan te tonen van aannames kwam van het één het ander. Zonder de medewerking van alle anesthesiologen en hun medewerkers was hoofdstuk 6 er nooit geweest. Ankie is een anesthesioloog met wie ik dolgraag samen op de operatiekamer een patiënt met een moeilijke luchtweg wil helpen. Haar ervaring en charisma geven een KNO-arts blind vertrouwen.

Laura is mijn collega en maatje sinds jaren. Door Laura werd ik nog opgeleid in de laryngologie en samen hebben wij als collegae jarenlang ‘alles behalve oren en oncologie’ gedaan. Kort na het ontstaan van de blauwprint voor deze promotie werd ik een van haar eerste promovendi. Dat is inhoudelijk goed, maar voor het schrijfplezier minder: het moet perfect! (ondanks haar fameuze stopzin: “Perfectie bestaat niet, kijk maar in de spiegel”). Het was een lange weg, waar steeds nieuwe promovendi mij ruimschoots inhaalden, maar zij heeft mij de eindstreep laten halen, waarvoor mijn dankbaarheid naar haar zeer groot is.

De bovengenoemde blauwprint kwam tot stand doordat oud-collega **Professor Robert Stokroos** mij op de gang tegenkwam, met zijn jas al aan, maar door mijn enthousiasme deze weer uit deed en in een half uur de volgorde van dit boekje optekende. Met een feilloos gevoel voor het simplificeren van iets groots en engs heeft hij mij gedeeltelijk over de promotie-streep getrokken, het leek allemaal appeltje-eitje...



De andere stimulans om toch een keer de stoute promotieschoenen aan te trekken kwam van mijn opleider en collega **Professor Hans Manni†**. Hij sprak ooit over mij uit dat ik nooit zou promoveren maar beloofde op mijn laatste dag van de opleiding, bij het overhandigen van het boek van “Brünings und Denker”, dat als ik ooit zou promoveren, hij nog een Brünings-relikwie voor mij zou hebben. Helaas heb ik te lang over mijn promotie gedaan en kan hij dit heugelijk feit niet meer meemaken. Desalniettemin ben ik hem dankbaar voor de wijze lessen en opleiding, tijdens mijn assistententijd en daarna.

De goede lezer vindt ook een vreemde eend in dit boekje; het onderzoek naar de hormoonreceptoren in de stemplooien; een onderzoek dat ik zelf had bedacht als beoogd promotieonderzoek. Patholoog **Carine Peutz – Kootstra** puzzelde mee en stak er vele uren in. Dankzij haar medewerkers op het laboratorium van de afdeling pathologie, waar ik leerde coupes te snijden en kleuren, is daar een mooi artikel uitgekomen dat ik niet zonder de steun van **Kenneth Kross** had af gekregen. Hij hielp het stuk ‘even’ leesbaarder te maken dankzij zijn eigen kennis van celbiologie. Tegenwoordig is Kenneth KNO-arts in Zuid-Frankrijk, een carrièrestap waar ik best jaloers op ben, omdat ik deze niet zou durven nemen.

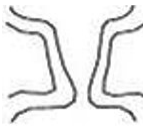


Voor promoveren heb je zitvlees nodig en die kreeg ik elk trimester in de vorm van 1 week ‘RS’ zoals dat in ons rooster heet. Dat betekent dat collegae alles die week moeten overnemen en dat voelt niet fijn. Bij dezen dank ik alleen daarom al **Janny Hof, Martin Lacko, Joost van Tongeren, Raymond van de Berg, Bing Tan, Josine Widdershoven, Katja Hellingman, Dirk Kunst, Jérôme Waterval, Femke Verhees, Mark van Tilburg en Stijn van Weert**, naast de reeds genoemden. Zij zijn allen niet alleen waarnemers tijdens mijn RS-weken maar ook stuk voor stuk collegae met wie ik met zeer veel plezier samen werk, of heb gewerkt. Dit geldt natuurlijk ook voor de niet-medische stafleden **Lucien Anteunis†, Jan Brokx, Erwin George, Gijs Hoskam, Denise Breukels – van Barneveld, Herman Kingma, Ellen Gerrits, Renée Speyer en Edith Berry**.

Niet alleen de KNO-stafleden nemen waar, ook alle **assistenten in opleiding**, die ik allen in mijn hart heb gesloten, zijn er in goede en slechte tijden, zowel in mijn opleidingstijd als tegenwoordig in mijn hoedanigheid als stafid. Mijn werk heeft dagelijkse waarde door het feit dat er jonge artsen zijn die ik (ook ongevraagd) kennis mag overdragen maar die mij ook scherp houden, op medisch als op niet-medisch vlak. De waardering en voorgehouden spiegel die ik van hen krijg zijn een groot onderdeel van mijn levensgeluk. Na bijna 20 jaar beslaan hun namen een flink deel van de pagina; bij

dezen: **Janny Hof, Harriët Hafkamp, Rico Rinkel, Susan Balter, Martijn Alles, Gertjan Kamps, Noortje Puts, Ewa Bergshoeff, Ivette Gooskens, Stijn van Weert, Willeke Lok, Isabelle Lutgens, Sjors van Pinxteren, Luke Schoenmakers, Jos Straetmans, Job Postelmans, Jérôme Waterval, Diane Smit, Jeroen Mooren, Guido Dees, Josine Widdershoven, Marieke Lips, Raymond van de Berg, Rick van de Langenberg, Sietze Reitsma, Nicoline Leunis, Maarten Borgemeester, Darius Henatsch, Simone Schaëfer, Sanne Wagemakers, Stijn van der Heijden, Marijn Stuut, Rens van de Goor, Anna Jasmijn Nobel – Hoff, Jasper Smit, Mark van Tilburg, Femke Verhees, Michel van Hooren, Glen Kempers, Michelle Florie, Sophie Kraaijenga, Tessa van Dooren, Milou van Orsouw, Floris Bettink, Marc van Hoof, Iris Krebbers, Frans Mulder, Max Gerdson, Joost Stultiens, Lisa van Stiphout; Mitch Hoogstraten; 51 dierbare collegae!**

De keuze voor mijn promotieonderwerp komt natuurlijk niet uit de lucht vallen; al sinds mijn 2^e studiejaar is de menselijke (zang)stem mijn drijfveer in de geneeskunde. Daar kwam in 2001 meteen extra kleur bij toen ik logopedisten **Joyce Uiterwijk Winkel** en **Hellen Moes** ontmoette die mij de logopedische kant van de laryngologie bij brachten. Met Joyce begin ik al 15 jaar het leukste spreekuur van de week op maandagmorgen, de 'PLA'. Samen met alle assistenten in opleiding die 'SLA' doen is dat de kers op de weektaart. Ik mag nog een paar jaar van haar genieten!



Promoveren doe je niet alleen, laat staan dat je tijdens een gewone werkweek ook nog patiënten kunt includeren voor het onderzoek of alle stemplooi-beelden kunt beoordelen. **Elodie Mendels** was mijn eerste heldin voor 'de grote studie', al in 2010, die de grote klus van de review op zich nam en de pilot voor 'de grote studie' begon. Een hardwerkende, flamboyante studente die ook nog eens in opleiding kon bij ons, maar haar hart volgde en nu kinderdermatoloog in Rotterdam is. Ik ben blij dat wij nog steeds contact hebben. Zij werd opgevolgd door **Loes van Kempen, Annemarie Akkermans** en **Nienke Heemskerk** die allen de inclusies verzorgden en participeerden in de 'observatiestudie', en **Sonja Bašić** en **Gusta van Zwieten** die hun tanden in de beoordelingen zetten van 'de grote studie'. Al deze dames hebben, terecht, een plekje in dit boekje gekregen. Dit geldt ook voor **Janneke Schepens** die mij al hielp bij het 'hormoonartikel' en statistische heldin **Sophie Vanbelle** die de onmogelijke vragen van mij wist om te zetten in, voor mij nog onmogelijker te begrijpen, statistische antwoorden.

De op een na laatste hobbel van de promotie is het aanbieden aan en laten beoordelen door de leescommissie; **Professor Wolfgang Buhre, Professor Freek Dikkers, Professor Ernst-Jan Spiel en Elisabeth Sjögren** hebben deze kluif aanvaard en positief

beoordeeld. Aangevuld met de coronaleden **Josine Widdershoven en Eva Schaffrath**, is de laatste hobbel hopelijk een mooie, leuke en enerverende verdediging, bijgestaan door mijn paranimfen, directe collega en vriendin **Janny Hof**, en oud-assistent en vriendin **Diane Smit**, bij wie ik allebei ook paranimf mocht zijn.

Een KNO-bedrijf waarin een promotie en een leuk spreekuur doen mogelijk is, wordt draaiende gehouden door een keur aan medewerkers die de artsengrillen kunnen trotseren: **Marèse, Ester, Nelly, Elly, Nicole, Judith** (die mij nog fantastisch heeft geholpen met de sponsoring van dit geheel), **Luana, Audrey, Antje, Rita, Rowan, Cécile, Alice** en alle stagiaires van het secretariaat; **Anton, Irene, Ruud, Suzanne, Mirjam, Karin, Doris, Phil, Bianca, Wendy, Ingrid, Jamilla, Nicole, Monique, Annemie, Kim, Birgit, Jacqueline** en alle stagiaires op de polikliniek.

Zij kennen allen mijn OCD, en mijn 'hoe het heurt' houding, en ik op mijn beurt ken ook hun eigenaardigheden. Samen houden we het werken op de afdeling KNO zo leuk. Ik hoop nog jaren met hen te mogen werken.



Inmiddels kom ik aan bij wie ik eigenlijk allemaal zou willen bedanken, maar na bijna 20 jaar wordt het dankwoord dan meer pagina's dan de promotie zelf. Alle medewerkers van het **Audiologisch Centrum, het Evenwichtslab, de afdeling Logopedie, de operatie afdeling, de Expertise Eenheid Transmurale Zorg** en natuurlijk alle artsen en medewerkers van aanpalende vakken, leden van werkgroepen, commissies, besturen en Avicenna-genoten hebben allen hun aandeel gehad in de ontwikkeling tot wie ik nu ben, maar ook nog wie ik ga worden.

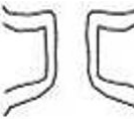
Hoeveel heeft mijn privéleven eigenlijk te lijden gehad onder de promotie? Collega **Piet Mirck** nam aan dat ik thuis er veel aan deed. Niets is minder waar. In mijn vrije tijd moet ik muziek maken, anders word ik niet gelukkig. **Vocal Group Odeon** is de vereniging waar ik kan ontladen in zang, spel en scherts. Een heerlijke groep vrienden met allen dezelfde (valse) humor. Mèt dee lui kin iech afsjakele.

Elke dag wordt begonnen en regelmatig beëindigd met een bericht van de Holletjes; **Kenneth, Raymond, Jérôme, Jeroen, Guido en Rick**, ik zal hun koosnamen hier maar niet zetten. Ik ben niet de "Dr-hekkensluter" zoals ik had verwacht, die eer is aan Guido. Ik gun iedereen een Holletjesgroep; lief en leed, anonieme patiëntencasuïstiek en ranzige grappen; een dagelijks genoeg al sinds 2 januari 2012.

Ik heb er geen spijt van dat ik (zeer) lang over mijn promotie heb gedaan, al was het maar omdat mijn moeder, **Huups van der Veen Vonk – Brunings**, de heugelijke dag wil

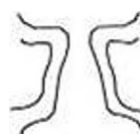
meemaken. Als de promotie van je zoon (en examens van je kleinkinderen **Filip** en **Marijke**) een goed streven is om niet weer ziek te worden, dan stel ik met liefde de einddatum nog wat uit. Het jarenlang vragen door **Arnold van der Veen Vonk** of ik al een zaaltje gehuurd heb (voor het feestje) geeft ook weinig stimulans. Edoch, zij hebben mij wel de lading meegegeven om zelf van het leven te maken wat er in zit. Niets is cadeau of komt vanzelf. Met dat gedachtengoed ben ik er alleen maar rijker op geworden. Mijn zussen **Madeline** en **Marlies** bewandelen datzelfde pad, ieder op hun eigen manier.

Ik werk om te leven, om een leven te delen met de mensen van wie ik houd. **Toon van Baars** heeft voor mij een leven in 'het Westen' opgegeven, zodat wij samen ons leven kunnen vieren in Maastricht, maar ook vaak met **Arthur** in Heilbach, en hopelijk op vele plekken op deze wereld. Ik ben daar dankbaar voor, al laat ik dat waarschijnlijk nooit genoeg blijken. C'est la vie! Ik houd van hem!





List of publications





List of publications

van Hoof M, Wigren S, Blechert JI, Molin M, Andersson H, Mateijsen DJM, Bom SJH, Calmels MN, van der Rijt AJM, Flynn MC, van Tongeren J, Hof JR, **Brunings JW**, Anteunis LJ, Algarra JM, Stokroos RJ, Joore MA. A Multinational Cost-Consequence Analysis of a Bone Conduction Hearing Implant System—A Randomized Trial of a Conventional vs. a Less Invasive Treatment With New Abutment Technology. *Front. Neurol.*, 13 March 2020 Volume 11, Article 106, Open Access

Brunings JW, Vanbelle S, Hamaekers AEW, Kremer B, Bašić S, van Zwieten G, Baijens LWJ. Voice and Vocal Fold Condition Following Short-Term General Anesthesia: A Prospective Study. *J Voice*. 2020 Jan 2. [Epub ahead of print]

Calon TGA, Johansson ML, de Bruijn AJG, van den Berge H, Wagenaar M, Eichhorn E, Janssen MML, Hof JR, **Brunings JW**, Joore MA, Jonhede S, van Tongeren J, Holmberg M, Stokroos RJ. Minimally Invasive Ponto Surgery Versus the Linear Incision Technique With Soft Tissue Preservation for Bone Conduction Hearing Implants: A Multicenter Randomized Controlled Trial. *Otol Neurotol*. 2018 Aug;39(7):882-893.

Calon TGA, van Tongeren J, Heuft AME, **Brunings JW**, Bollen D, Hof JR, Stokroos RJ. Percutaneous bone-anchored hearing system implant survival after 550 primary implant surgeries. *Clin Otolaryngol*. 2018 Apr;43(2):735-739

Brunings JW, Vanbelle S, Akkermans A, Heemskerk NMM, Kremer B, Stokroos RJ, Baijens LWJ. Observer Agreement for Measurements in Videolaryngostroboscopy. *J Voice*. 2018 Nov;32(6):756-762.

Johansson ML, Stokroos RJ, Banga R, Hol MK, Mylanus EA, Savage Jones H, Tysome JR, Vannucchi P, Hof JR, **Brunings JW**, van Tongeren J, Lutgert RW, Banerjee A, Windfuhr JP, Caruso A, Giannuzzi AL, Bordin S, Hanif J, Scharf-Morén N, Singam S, Jonhede S, Holmberg M, Cremers CW, Hultcrantz M. Short-term results from seventy-six patients receiving a bone-anchored hearing implant installed with a novel minimally invasive surgery technique. *Clin Otolaryngol*. 2017 Oct;42(5):1043-1048.



Calon TG, van Hoof M, van den Berge H, de Bruijn AJ, van Tongeren J, Hof JR, **Brunings JW**, Jonhede S, Anteunis LJ, Janssen M, Joore MA, Holmberg M, Johansson ML, Stokroos RJ. Minimally Invasive Ponto Surgery compared to the linear incision technique without soft tissue reduction for bone conduction hearing implants: study protocol for a randomized controlled trial. *Trials*. 2016 Nov 9;17(1):540.

van Leijen-Zeelenberg JE, **Brunings JW**, Houkes I, van Raak AJ, Ruwaard D, Vrijhoef HJ, Kremer B. Using Lean Thinking at an otorhinolaryngology outpatient clinic to improve quality of care. *Laryngoscope*. 2016 Apr;126(4):839-46.

San Giorgi MR, van den Heuvel ER, Tjon Pian Gi RE, **Brunings JW**, Chirila M, Friedrich G, Golusinski W, Graupp M, Horcasitas Pous RA, Ilmarinen T, Jackowska J, Koelmel JC, Ferran Vilà F, Weichbold V, Wierzbicka M, Dikkers FG. Age of onset of recurrent respiratory papillomatosis: a distribution analysis. *Clin Otolaryngol*. 2016 Oct;41(5):448-53.

van Leijen-Zeelenberg JE, Huismans GW, Bisschop JA, **Brunings JW**, van Raak AJ, Ruwaard D, Vrijhoef HJ, Kremer B. Experiences and preferences of patients visiting an otorhinolaryngology outpatient clinic: a qualitative study. *Health Expect*. 2016 Apr;19(2):275-87

Brunings JW, Schepens JJ, Peutz-Kootstra CJ, Kross KW. The expression of estrogen and progesterone receptors in the human larynx. *J Voice*. 2013 May;27(3):376-80.

Tjon Pian Gi RE, Ilmarinen T, van den Heuvel ER, Aaltonen LM, Andersen J, **Brunings JW**, Chirila M, Dietz A, Ferran Vilà F, Friedrich G, de Gier HH, Golusinski W, Graupp M, Hantzakos A, Horcasitas R, Jackowska J, Koelmel JC, Lawson G, Lindner F, Remacle M, Sittel C, Weichbold V, Wierzbicka M, Dikkers FG. Safety of intralesional cidofovir in patients with recurrent respiratory papillomatosis: an international retrospective study on 635 RRP patients. *Eur Arch Otorhinolaryngol*. 2013 May;270(5):1679-87.

Mendels EJ, **Brunings JW**, Hamaekers AE, Stokroos RJ, Kremer B, Baijens LW. Adverse laryngeal effects following short-term general anesthesia: a systematic review. *Arch Otolaryngol Head Neck Surg*. 2012 Mar;138(3):257-64.

Speyer R, Pilz W, Van Der Kruis J, **Brunings JW**. Reliability and validity of student peer assessment in medical education: a systematic review. *Med Teach*. 2011;33(11):e572-85.



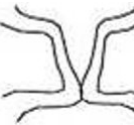
Speyer R, Bogaardt HC, Passos VL, Roodenburg NP, Zumach A, Heijnen MA, Baijens LW, Fleskens SJ, **Brunings JW**. Maximum phonation time: variability and reliability. *J Voice*. 2010 May;24(3):281-4.

Brunings JW, Driessen A, De Jong JM, Manni JJ. A 15-year-old girl with a spindle cell carcinoma of the parotid gland. A diagnostic challenge solved by immunohistochemistry and cytogenetic analysis. *Acta Otolaryngol*. 2006 Jan;126(1):104-7.





Curriculum Vitae





Curriculum Vitae

Jan Wouter Brunings werd geboren op 19 november 1973 te Gorssel. Hij rondde in 1992 het VWO af in Zutphen aan de Stedelijke Scholengemeenschap (voorheen Stedelijk Lyceum) in de destijds genoemde gymnasiumstroom. Wegens uitloting studeerde hij eerst een jaar farmacie aan de Rijksuniversiteit Groningen waarna hij in 1993 kon starten met de studie medicijnen aan diezelfde universiteit. De wetenschappelijke stage verrichtte hij in 1997 aan de Medical University of South Africa (MEDUNSA, thans Sefako Makgatho Health Sciences University) op de afdeling Obstetrie & Gynaecologie in Ga-Rankuwa, nabij Pretoria in Zuid-Afrika. De coschappen doorliep hij in de Deventer Ziekenhuizen (thans Deventer Ziekenhuis) van 1998 tot in 2000 met een uitstap in 1999 naar het Sint-Elisabeth Hospitaal in Willemstad op Curaçao voor het coschap Obstetrie & Gynaecologie. Het keuze coschap volgde hij bij de afdeling Keel-, Neus- en Oorheelkunde van de Deventer Ziekenhuizen.

In augustus 2000 behaalde hij het artsexamen waarna hij in oktober 2000 begon als arts-assistent Keel-, Neus- en Oorheelkunde niet in opleiding in het academisch ziekenhuis Maastricht (thans het Maastricht UMC+).

Op 1 januari 2001 startte zijn opleiding Keel-, Neus- en Oorheelkunde aldaar onder wijlen Professor dr J.J. Manni en later Professor dr B. Kremer. Het niet-academische deel van de opleiding werd gevolgd in het Atrium Medisch Centrum (thans Zuyderland Medisch Centrum) onder Dr T.D. Zylker. Aan het einde van de opleidingstijd in 2005 heeft hij nog 3 maanden een buitenlandse stage gevolgd aan het Landeskrankenhaus – Universitäts Klinikum Graz, Oostenrijk, bij de afdeling Foniatrie onder Professor dr G. Friedrich. De gehele opleiding heeft hij extra aandacht gehad voor de menselijke (zang)stem.

In januari 2006 werd hij ingeschreven als KNO-arts en bleef hij werkzaam als staflid in het Maastricht UMC+ met laryngologie, rhinologie en voorste schedelbasis chirurgie als aandachtsgebieden. In deze fase is de basis gelegd voor dit proefschrift. Tevens verrichtte hij hiernaast diverse organisatorische taken binnen en buiten de afdeling, welke in 2006 begonnen als Medisch Hoofd van de polikliniek, in 2014 interim afdelingshoofd van het Audiologisch Centrum, sinds 2016 als Waarnemend Hoofd van de afdeling KNO en sinds 2020 als Directeur Transmurale Zorg.

De opleiding van de arts-assistenten ligt hem na aan het hart.

Jan Wouter Brunings woont samen met zijn man Toon van Baars (en hond Arthur) in Maastricht en in de weekenden in hun 'Ausflug vom Alltag' in Heilbach in de Duitse Eifel.

