Aan Margaux en Marie-Julie

Marcel Vanhooren†

Faculteit Geneeskunde en Gezondheidswetenschappen

A MULTI-CENTRE PROSPECTIVE STUDY

ON FUNCTIONAL OUTCOMES AND

POST-STROKE OROPHARYNGEAL DYSPHAGIA

EEN MULTICENTRISCHE PROSPECTIEVE STUDIE NAAR

FUNCTIONELE OUTCOMES EN

OROFARYNGEALE DYSFAGIE NA CVA

Proefschrift voorgelegd tot het behalen van de graad van doctor in de medische wetenschappen aan de Universiteit van Antwerpen te verdedigen door

Ingeborg SIMPELAERE

Promotoren

Prof. dr. Marc De Bodt

Prof. dr. Gwen Van Nuffelen

Antwerpen, 2018

A multi-centre prospective study on functional outcomes and post-stroke oropharyngeal dysphagia

Ingeborg Simpelaere

Faculteit Geneeskunde en Gezondheidswetenschappen, Universiteit Antwerpen, Antwerpen, 2018

Thesis Universiteit Antwerpen – with summary in Dutch

TABLE OF CONTENTS

MEMBERS OF THE JURY

Prof. dr. De Bodt Marc – Supervisor (PhD, MSc)

Head of the Rehabilitation Center for Communication Disorders, Department of Otorhinolaryngology and Head & Neck Surgery, Antwerp University Hospital, Antwerp, Belgium

Department of Translational Neurosciences, Faculty of Medicine and Health Sciences, University of Antwerp, Antwerp, Belgium

Department of Speech, Language and Hearing Sciences, Faculty of Medicine and Health Sciences, Ghent University, Ghent, Belgium

Prof. dr. Van Nuffelen Gwen – Supervisor (PhD, MSc)

Rehabilitation Center for Communication Disorders, Department of Otorhinolaryngology and Head & Neck Surgery, Antwerp University Hospital, Antwerp, Belgium

Department of Translational Neurosciences, Faculty of Medicine and Health Sciences, University of Antwerp, Antwerp, Belgium

Department of Speech, Language and Hearing Sciences, Faculty of Medicine and Health Sciences, Ghent University, Ghent, Belgium

Prof. dr. Vanderveken Olivier – President of the jury (PhD, MD)

Vice-Chair of the Department of Otorhinolaryngology and Head & Neck Surgery, Antwerp University Hospital, Antwerp, Belgium

Department of Translational Neurosciences, Faculty of Medicine and Health Sciences, University of Antwerp, Antwerp, Belgium

Dr. Boey Ronny – Member of the jury (PhD, MSc)

Post-doc researcher, Antwerp Centre of Stuttering Therapy, University of Antwerp, Belgium

President of the Centre for Information, Education and Research on Stuttering (CIOOS), Ypres, Belgium

Prof. dr. Desuter Gauthier – Member of the jury (PhD, MD, MSc)

Director of the Voice & Swallowing Clinic, ENT Head & Neck Surgery, Saint-Luc University Hospital, Catholic University of Louvain-la-Neuve, Brussels, Louvain-la-Neuve, Belgium

Prof. dr. Goeleven Ann – Member of the jury (PhD, MSc)

Research Group Experimental Oto-rhino-laryngology, University Hospitals Leuven, Catholic University of Leuven, Leuven, Belgium

ABBREVIATIONS AND SYMBOLS

α	Alpha
AAT	Akense Afasie Test
aDSWAL-QoL	adjusted DSWAL-QoL
ANOVA	analysis of variance
Assist.	assistance
AUC	area under the (receiver operating characteristic) curve
AZ	academisch ziekenhuis
BOHSE	Kayser-Jones Brief Oral Health Status Examination
BSSD	Belgian Society for Swallowing Disorders
bv.	bijvoorbeeld
CEBAM	Centre for Evidence-Based Medicine
CI	confidence interval
CINAHL	Cumulative Index to Nursing and Allied Health Literature
COM	Communication
COSMIN	COnsensus-based standards for the selection of health status Measurement INstruments
CSE	clinical swallowing examination
CT	computed tomography
CTT	classical test theory
CVA	cerebrovasculair accident
Dent	Dentures
df	degrees of freedom
DHI	Deglutition Handicap Index
d.i.	dit is
DIF	differential item functioning
doi	digital object identifier
DP	dental pain
DRS	Dysphagia Research Society
DSWAL-QoL	Dutch version of the Swallowing Quality-of-Life
Dys	patients suffering from dysphagia
DysLC	patients with dysphagia accompanied by language impairment and/or cognitive disorders
EATDES	Eating desire
EATDUR	Eating duration
e.g.	for example
Embase	Excerpta Medica Database
EORTC QLQ- PR25	European Organization for Research and Treatment of Cancer Quality of Life Questionnaire-prostate specific 25-item
ESSD	European Society for Swallowing Disorders
Ext	extreme
Ext (+ext/÷ext)	extreme (with extremes/without extremes)
FAT	Fatigue
FDA-USHHS	Food and Drug Administration of the United States Department of Health and Human Services
FDI	Fédération Dentaire Internationale
FEAR	Fear of eating

FEES	Fiberoptic Endoscopic Evaluation of Swallowing
FOIS	Functional Oral Intake Scale
FR	fit residual
FS	Food selection
GandT	gums and tissues
GB	General burden
Google Scholar	is the name of a search engine for scholarly literature
H.Hart Hospital	Heilig Hart Hospital (Note that this institution is now called AZ Delta)
HoGent	Hogeschool Gent
HRQoL	health-related quality of life
HR-PRO	health-related patient-reported outcome
HRQoL-PRO	health-related quality of life patient-reported outcome
ICC	intraclass correlation coefficient
i.e.	that is
Ind.	independent
INTEGO	geïntegreerd computernetwerk
IQR	interquartile range
IRT	item response theory
IS	initial scale
ISOQOL	International Society for Quality of Life Research
JBI	Joanna Briggs Institute
JBISRIR	JBI Database of Systematic Reviews and Implementation Reports
JBI COnNECT+	The Joanna Briggs Institute Clinical Online Network of Evidence for Care and Therapeutics Plus
κ	Kappa statistic value
KHBO	Catholic University College of Bruges (Note that this institution is now called VIVES)
KR-20	Kuder-Richardson Formule 20
LC	patients suffering from language impairment and/or cognitive disorders without the presence of dysphagia
LCI	lower confidence interval
LID	local item dependence or local item dependency
LOA or LoA	limits of agreement
LOS	length of hospital stay
LSVT	Lee Silverman Voice Treatment
MASA	Mann Assessment of Swallowing Ability
Max	maximum
MEDLINE	Medical Literature Analysis and Retrieval System Online
MedNar	is a consumer health search engine
MeSH	Medical Subject Headings
MH	Mental health
MIC	minimal important change
MID	minimal important difference
Min	minimum
MMSE	Mini Mental State Examination
MRI	magnetic resonance imaging
N or n	number of patients/subjects
NA	not applicable

NCSS LLC	NCSS LLC is the company that produces and distributes PASS
NG	nasogastric feeding
NHANES	National Health and Nutrition Examination Survey
NIHSS	National Institutes of Health Stroke Scale
No. or No	number
NT	natural teeth
OC	oral cleanliness
OD	orofaryngeale dysfagie
OHAT	Oral Health Assessment Tool
OHP	oral health professional
OR	odds ratio
p or P or p-value	probability or probability value
PASS	Power Analysis and Sample Size System
PCA	principal component analysis
PEG	percutaneous gastrostomy feeding or percutaneous endoscopic gastrostomy-feeding
PH	Cox proportional hazard
PNI	prognostic nutritional index
PREE	Patient-Rated Elbow Evaluation
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analysis
PRO	patient-reported outcome
PROM	patient-reported outcome measure
PROQOLID	Patient-Reported Outcome and Quality of Life Instruments Database
ProQuest	is the name of a database (full name is not known)
PSI	person separation index
PSOD	post-stroke oropharyngeal dysphagia
PubMed	Public MEDLINE or Publisher MEDLINE
Q1	first quartile
Q2	second quartile
Q3	third quartile
QoL	quality of life
QQ-plot or Q-Q plot	quantile-quantile plot
R 3.0.1, R 3.3.2	is a language and environment for statistical computing and graphics
r_s	Spearman's rank correlation coefficient
R-squared	coefficient of determination (regression analysis)
ROC	receiver operating characteristic
RR	responsiveness ratio
RS	rescaled scale based on the suggestions from the Rasch methodology
RUMM	Rasch Unidimensional Measurement Models
SAS	Statistical Analysis Software
SD	standard deviation
SDC	smallest detectable change
SD_{change}	standard deviation of change
SDC_{group}	smallest detectable change for one group
SDC_{ind}	smallest detectable change in one individual
SE	standard error
SEM	standard error of measurement

SF	Social functioning
SGRQ	St. George's Respiratory Questionnaire
Sig.	significant
SL	Sleep
SLP or SP	speech-language pathologist or speech pathologist (same meaning)
Spearman's rho	Spearman's rank correlation coefficient
SPSS	Statistical Package for the Social Sciences
Std.dev.	standard deviation
SWAL-QoL	Swallowing Quality-of-Life
SYMPT	Symptoms
TIA	transient ischaemic attack
Tukey HSD test	Tukey Honestly Significant Difference test
VIVES	is part of the name of the University College Bruges
VVL	Vlaamse Verening voor Logopedisten
WHO	World Health Organization
χ^2	chi-square
%	percentage
#	number
&	and

PUBLICATIONS

International first author publications

Simpelaere IS, Van Nuffelen G, Vanderwegen J, Wouters K, De Bodt M. Oral health screening: feasibility and reliability of the oral health assessment tool as used by speech pathologists. International Dental Journal. 2016;66(3):178-89.

Simpelaere I, White A, Bekkering GE, Geurden B, Van Nuffelen G, De Bodt M. Patient-reported and proxy-reported outcome measures for the assessment of health-related quality of life among patients receiving enteral feeding: a systematic review protocol. JBI Database of Systematic Reviews and Implementation Reports. 2016;14(7):45-75. doi:10.11124/JBISRIR-2016-002982.

Simpelaere IS, Vanderwegen J, Wouters K, De Bodt M, Van Nuffelen G. Feasibility and Psychometric Properties of the Adjusted DSWAL-QoL Questionnaire for Dysphagic Patients with Additional Language and/or Cognitive Impairment: Part I. Dysphagia. 2017;32(3):401-19.

Simpelaere IS, Van Nuffelen G, De Bodt M, Vanderwegen J, Hansen T. Validation of the Dutch version of the Swallowing Quality-of-Life Questionnaire (DSWAL-QoL) and the adjusted DSWAL-QoL (aDSWAL-QoL) using item analysis with the Rasch model: a pilot study. Health and Quality of Life Outcomes. 2017;15(1):66. doi:10.1186/s12955-017-0639-3.

Simpelaere IS, De Bodt M, Roelant E, Vanderwegen J, Van Nuffelen G. The role of oropharyngeal dysphagia in predicting place of residence and functional outcomes after stroke: a one-year prospective study. In review.

THESIS AT A GLANCE

Objective	Methods
STUDY I: ***Oral health screening: feasibility and reliability of the oral health assessment tool as used by speech pathologists*** (Chapter II)	
To evaluate the feasibility and reliability of the OHAT.	A multi-centre study in 132 elderly subjects was conducted by 3 SPs. The feasibility and inter-rater, test-retest and intra-rater reliabilities of the OHAT were assessed.
STUDY II: ***The role of oropharyngeal dysphagia in predicting place of residence and functional outcomes after stroke: a one-year prospective follow-up study*** (Chapter III)	
To determine whether the presence of PSOD at baseline (i.e., during the first swallowing examination after the onset of stroke) could predict place of residence post-discharge. To evaluate whether the MASA and/or the OHAT could predict short- and long-term patient outcomes (i.e., pneumonia, feeding status, functional independence, residence and survival)	A multi-centre prospective study with a one-year follow-up (i.e., baseline, 1 month, 3 months, 6 months, and 1 year) was conducted in 151 acute stroke patients.
STUDY III: ***Feasibility and psychometric properties of the adjusted DSWAL-QoL questionnaire for dysphagic patients with additional language and/or cognitive impairment: part I*** (Chapter IV)	
To develop the aDSWAL-QoL. To evaluate its feasibility, internal consistency, test-retest reliability, and criterion validity.	Development of the aDSWAL-QoL based on aphasia- and cognitive-friendly recommendations. Cross-sectional study: 78 dysphagic patients, among 43 had additional language and/or cognitive impairments (DysLC).
STUDY IV: ***Validation of the Dutch version of the Swallowing Quality-of-Life Questionnaire (DSWAL-QoL) and the adjusted DSWAL-QoL (aDSWAL-QoL) using item analysis with the Rasch model: a pilot study*** (Chapter V)	
To examine the structural validity and objectivity of the DSWAL-QoL and aDSWAL-QoL total scales and subscales and to evaluate the statistical sufficiency of the total scores and subscale scores.	Item analysis with the Rasch model was performed with previously collected data from a validation study of 108 patients.
STUDY V: ***Patient-reported and proxy-reported outcome measures for the assessment of health-related quality of life among patients receiving enteral feeding: a systematic review protocol*** (Chapter VI)	
To develop a systematic review protocol for PROs and proxy-reported outcome measures that assess HRQoL among patients receiving enteral feeding.	The systematic review protocol encompassed a description of the methods to evaluate the clinical utility and psychometric properties of included PROs and proxy-reported outcome measures.

Conclusions
STUDY I: ***Oral health screening: feasibility and reliability of the oral health assessment tool as used by speech pathologists*** (Chapter II)
This is the first study to examine the feasibility and reliability of the OHAT as used by SPs. As the results showed both good feasibility and reliability, the OHAT has the potential to add to the clinical swallowing examination. However, future research investigating actual referral strategies and adaptation of care strategies following assessment with OHAT is needed.
STUDY II: ***The role of oropharyngeal dysphagia in predicting place of residence and functional outcomes after stroke: a one-year prospective follow-up study*** (Chapter III)
PSOD at baseline could predict failure to return home after discharge. Baseline MASA and OHAT scores were both significant predictors of pneumonia during hospitalization and follow-up. The baseline MASA score was a significant predictor of feeding status, functional independence and place of residence during follow-up, although more information regarding feeding status and functional independence could be obtained by repeating the MASA at the particular time points. The MASA repeated at specific time points was not a significant predictor of place of residence during follow-up. The baseline MASA score was a significant predictor of survival following stroke, and severe dysphagia was associated with poor survival.
STUDY III: ***Feasibility and psychometric properties of the adjusted DSWAL-QoL questionnaire for dysphagic patients with additional language and/or cognitive impairment: part I*** (Chapter IV)
The aDSWAL-QoL is a feasible, reliable and valid tool for use with DysLC patients. The aDSWAL-QoL was more feasible than the DSWAL-QoL for the DysLC group. We obtained high internal consistency for total scale and for almost all subscales. Total aDSWAL-QoL scores showed excellent test-retest agreement and good criterion validity with respect to the DSWAL-QoL. Almost all subscales showed significantly moderate to good test-retest agreement and criterion validity. The psychometric properties of the 'Food selection' and – to a lesser extent – the Fear of eating and Eating desire subscales were inadequate. Conversion of the aDSWAL-QoL into a computer-assisted self-administered format should be investigated.
STUDY IV: ***Validation of the Dutch version of the Swallowing Quality-of-Life Questionnaire (DSWAL-QoL) and the adjusted DSWAL-QoL (aDSWAL-QoL) using item analysis with the Rasch model: a pilot study*** (Chapter V)
This study represents the first analyses of the DSWAL-QoL and aDSWAL-QoL with the Rasch model. The analysis could not establish the psychometric properties of either of the scales or their subscales because they did not fit the Rasch model, and multidimensionality, disordered thresholds, DIF, and/or LID were found. The reliability and power of fit were high for the total scales (PSI = 0.93) but low for most of the subscales (PSI < 0.70). The targeting of persons and items was suboptimal. Despite the implementation of adjustments suggested by the Rasch model, not all weaknesses could be resolved. Therefore, relying on the DSWAL-QoL and aDSWAL-QoL total scale and subscale scores to make conclusions about a person's dysphagia-related HRQoL should be undertaken with caution before the psychometric requirements have been established. A larger and well-targeted sample is recommended to derive definitive conclusions about the items and the scales.
STUDY V: ***Patient-reported and proxy-reported outcome measures for the assessment of health-related quality of life among patients receiving enteral feeding: a systematic review protocol*** (Chapter VI)
Ongoing.

CHAPTER I: GENERAL INTRODUCTION

1.1 Post-stroke oropharyngeal dysphagia

Stroke

Stroke has been recognized worldwide as a leading cause of disability and death [1, 2]. The incidence of stroke, often stated as the number of newly diagnosed cases (i.e., first-ever stroke) per 100.000 (or per 1000) inhabitants over a period of time (e.g., 'per year') [3, 4, 5, 6] or expressed as 'per 1000 person-years' [7, 8], can be measured by population-based stroke registers [3]. Comparing data on stroke incidence based on previous population-based stroke registers in Europe is limited because of methodological differences (e.g., implementing age restrictions to define the population of interest) in data collection [3, 8], and therefore, it is difficult to estimate the true impact of stroke for the whole population [3]. A study in Belgium covering data from 178 Belgian sentinel general practitioners recording all cerebrovascular events for the period 1998-1999 revealed that the yearly age-and-gender-adjusted stroke attack rates were estimated at 185 cases per 100.000 inhabitants [9, 10]. It should be noted that the estimation of the incidence in this study was not limited to only first-ever stroke; patients with a history of stroke that showed a new event of stroke were also included [9]. The incidence rate of stroke in Flanders for 2015, using the database from INTEGO that collects data from 46 general practitioners, was estimated on 1.64 per 1000 inhabitants [6]. Knowledge of the key factors that affect the outcomes is important for predicting stroke prognosis [11]. The factors that have improved outcomes after stroke include the increased number of dedicated stroke units that have been developed during previous decades [11, 12] and the current guidelines for stroke prevention [11, 13]. The factors that can negatively affect outcomes after stroke include increasing age, stroke characteristics (i.e., lesion location, clinical features and stroke severity) and related complications [11, 14, 15]. The estimated yearly Belgian stroke mortality rate based on the study of Devroey et al. [10] was about 88 per 100 000 inhabitants.

Post-stroke oropharyngeal dysphagia

A common complication after stroke is post-stroke oropharyngeal dysphagia (PSOD) [1], defined as the presence of an abnormal swallowing physiology of the upper aerodigestive tract [16]. The reported incidence – which can also be expressed as a percentage (%) – of PSOD varies greatly (i.e., ranges from 14% to 94%) [17]. The latter can be explained by variations in the dysphagia identification methods used,

the timing of the initial evaluation after stroke, and the lesion location [18]. The presence of dysphagia after stroke is associated with poor outcomes as it may lead to institutionalization [15], but also to dehydration [19], malnutrition [2, 19], pneumonia [2, 20], decreased functional independence [15], increased mortality and morbidity [2, 10], increased length of hospital stay (LOS) [17], and increased health care costs [21].

Pneumonia

Pneumonia is the most feared complication following stroke [20]. Post-stroke pneumonia is mostly caused by dysphagia, leading to aspiration of colonized oropharyngeal material (i.e., liquid, foods, and oral secretions) [2, 20, 22, 23]. Depending on the operational definition for pneumonia and the timing of the pneumonia diagnosis, the reported incidence of pneumonia in acute stroke patients with dysphagia ranges from 16.2% to 33% [18]. Furthermore, these patients have a 3.17-fold higher risk of developing pneumonia than stroke patients without dysphagia [18]. As post-stroke pneumonia is a multifactorial phenomenon [20], tube feeding and a poor oral health status are also considered contributing factors [17, 20]. Tube-fed patients show a higher rate of pneumonia than patients who receive food or liquid by mouth [22]. Poor oral health status increases the risk of pneumonia, particularly among patients who have dysphagia and those receiving enteral feeding [2]. The risk of aspiration pneumonia is the highest when poor oral health and oral diseases are combined with the presence of swallowing and feeding problems, poor functional status, underlying diseases and an increasing age [24, 25].

Oral health

Poor oral health is an important contributing factor to the development of aspiration pneumonia [22-24, 26-28]. Therefore, oral care in dysphagic patients is essential, and oral health-care interventions have the potential to diminish the risk of aspiration pneumonia [28, 29] and its associated mortality risk [29, 30]. Oral health and dentition are both essential for swallowing and speech production [31]. In particular, they are essential during the preparatory phase of swallowing (i.e., for mastication) [32]. Maintaining functional units (i.e. pairs of opposing mandibular and maxillary teeth, particularly natural teeth) is crucial for masticatory function [32, 33]. The degree of masticatory function not only determines food selection [34, 35] but also influences nutritional status [34–37].

1.2 Health-related quality of life, patient-reported outcome measures and oropharyngeal dysphagia.

Quality of life

Quality of life (QoL) is defined by the World Health Organization (WHO) as follows: ‘‘individuals’ perceptions of their position in life in the context of the culture and value systems in which they live and in relation to their goals, expectations, standards and concerns. It is a broad-ranging concept incorporating in a complex way the person’s physical health, psychological state, level of independence, social relationships, personal beliefs and their relationships to salient features of the environment’’[38]. There is still a lack of consensus among researchers about the definition of QoL that can explain the different constructs or concepts that are selected for their measurement instruments [39, 40]. Quality of life varies between populations, time periods and diseases [41–43]. Also, the concept is subjective as it is based on a perception by the individual around his or her wellbeing [44]. Individuals from the same environment seem to have a similar conception about QoL (e.g. ‘‘reaching happiness’’ and ‘‘satisfaction in life’’ are important constructs of QoL in individuals living in the Western countries) [39].

Health-related quality of life

Health-related quality of life (HRQoL) is distinct from QoL as it is directly associated to a health state of an individual [39, 45]. Various definitions of HRQoL exist [42, 46]. However, agreement is reached in that HRQoL is a multidimensional concept (i.e. consists of physical, psychological and social domains of health), it is based on the individual’s perception of his or her wellbeing, encompasses the current health situation and the future perspective and is not restricted to the disease [39, 42]. Assessment of HRQoL is useful for measuring the outcome of disease and treatment effectiveness [47].

Patient-reported outcome measures and oropharyngeal dysphagia

The most common way to measure HRQoL issues is by patient-reported outcome (PRO) instruments, using a standardized questionnaire or an interview [48-52]. These PROs are defined by the Food and Drug Administration of the United States Department of Health and Human Services (FDA-USHHS) as ‘‘any report of the status of a patient’s health condition that comes directly from the patient, without interpretation of the patient’s response by a clinician or anyone else’’ [53]. These measures are important because they allow patients to communicate about their disease, describe the impact of the disease on

their HRQoL [45], and evaluate treatment outcomes [45, 54] from their own perspective [45, 54, 55]. Many PROs have been developed to measure HRQoL in patients with oropharyngeal dysphagia [52]. These PROs assess the presence and severity of dysphagia symptoms, and measure the influence of dysphagia on a person's HRQoL [48, 52, 56, 57]. These PROs add useful information to the clinical swallowing examination and instrumental investigations [49] and can be used as outcome measures of therapeutic interventions [48, 52, 56, 57]. The selection of the most suitable PRO for assessing HRQoL among a specific patient population group depends on the quality of its psychometric properties and its clinical utility [58]. Clinical utility is defined in terms of feasibility (i.e., the facilitation in completing the PRO, the way of administration, time duration and the patient's burden to complete the PRO) and interpretability (i.e., the extent to which a qualitative meaning can be assigned to quantitative scores) [59, 60]. The Swallowing Quality-of-Life Questionnaire (SWAL-QoL) [61-63] is a multidimensional, disease-specific PRO with good psychometric properties that has been considered the gold standard for assessing HRQoL in individuals with oropharyngeal dysphagia [64]. The SWAL-QOL has been translated into Dutch (DSWAL-QoL) and validated for a Flemish population [48]. According to classical test theory (CTT), the DSWAL-QoL shows good psychometric properties [48]. The examination of the psychometric properties of the DSWAL-QoL using modern item response theory (IRT) has not been previously performed.

Feasibility of patient-reported outcome measures

Patient involvement is crucial in the development and validation of these self-reported scales [65]. Use of PROs may therefore be impeded by certain functional impairments, such as communication disorders, often faced by individuals with dysphagia [66, 67]. Dysphagic patients may present with language impairments (i.e., aphasia) and/or cognitive disorders resulting from underlying diseases such as stroke and dementia [68, 69]. Such patients with language or cognition difficulties often struggle with PRO scales and may be unable to complete them [54, 70, 71] owing to comprehension difficulties [72] related to aphasia [73] or poor vision [74]. Indeed, they may face difficulties providing a response [75] that truly reflects their thoughts. Consequently, this group is at times excluded from HRQoL measurement studies [54, 71, 76]. This likely affects the validation of such scales [66], as the results cannot be generalized to the entire population. This problem may be alleviated by the use of interviewer-administered versions of HRQoL scales, which can yield the following advantages: (1) the interviewer can ask additional questions if inconsistent responses are provided; (2) further assistance, such as

explanations about the questions, can be given; and (3) the patient's comprehension of the questions can be assessed [71, 77]. However, self-administered questionnaires have an advantage in that they often reveal more delicate health information and can control for interviewer bias [75, 77]. Furthermore, caution is necessary when interpreting HRQoL responses given in a face-to-face interview, because interview responses may be affected by social desirability and/or acquiescent bias [75, 77].

Psychometric properties of patient-reported outcome measures

In this section, we will define the most relevant measurement properties for this thesis [58, 59, 78]. For a detailed description, we refer to Chapter VI.

Validity:

- Content validity: refers to the extent to which the different domains of the PRO are comprehensively sampled by the items in the questionnaire/scale.
- Criterion validity: measures the extent to which scores on the PRO are related to a gold standard.
- Construct validity: refers to the degree to which the scores of the PRO are consistent with hypotheses (e.g., with regard to relationships to scores of other instruments or to internal relationships) based on the assumption that the PRO validly measures the construct to be measured. Construct validity involves structural validity, hypotheses testing and cross-cultural validity. Structural validity refers to the extent to which the scores of a PRO instrument are an adequate reflection of the dimensionality of the construct to be measured. To examine the structural validity of the instrument, the PRO scale should be tested against the Rasch model (cf. below).

Reliability: refers to the degree to which the scale is free from measurement error. Reliability is defined as the extent to which scores for patients who have not changed are the same for repeated measurements under several conditions: 1) using different sets of items from the same PRO (i.e. internal consistency), or 2) over time (i.e., test-retest reliability):

- Internal consistency: refers to the extent to which items in a (sub)scale are inter-correlated and thus measuring the same concept.
- Test-retest reliability: refers to the stability of the scores over time.

Floor and ceiling effects: refer to the number of respondents who achieved the lowest or highest possible score. Floor or ceiling effects are present if more than 15% of the respondents achieved the lowest or highest possible score, respectively.

The Rasch model

The Rasch model within IRT is a mathematical measurement model, intended for the development and psychometric examination of measurement scales [79]. It is considered the gold standard against which scales summarizing item responses must be tested [80]. Item analysis using the Rasch model involves formal testing of a scale against a mathematic measurement model that specifies what should be expected in the item responses to provide interval-based measures instead of ordinal values [81, 82]. The Rasch model is based on a probabilistic Guttman pattern [81, 83]. Guttman scaling is deterministic as it expects a strict hierarchical ordering of items, which means that if a patient has affirmed an item representing a task of average difficulty, then all the items below that task (i.e., easier items) on the scale should also be affirmed [81]. The Rasch model is not deterministic, but probabilistic. This means that for the same person ability, the probability to affirm an easy item has to be higher than the probability to affirm more difficult items and for the same item difficulty, a person with a higher ability is expected to affirm all items endorsed by a person with lower ability and additionally one or more severe items [79, 81].

Proxy reported outcome measures

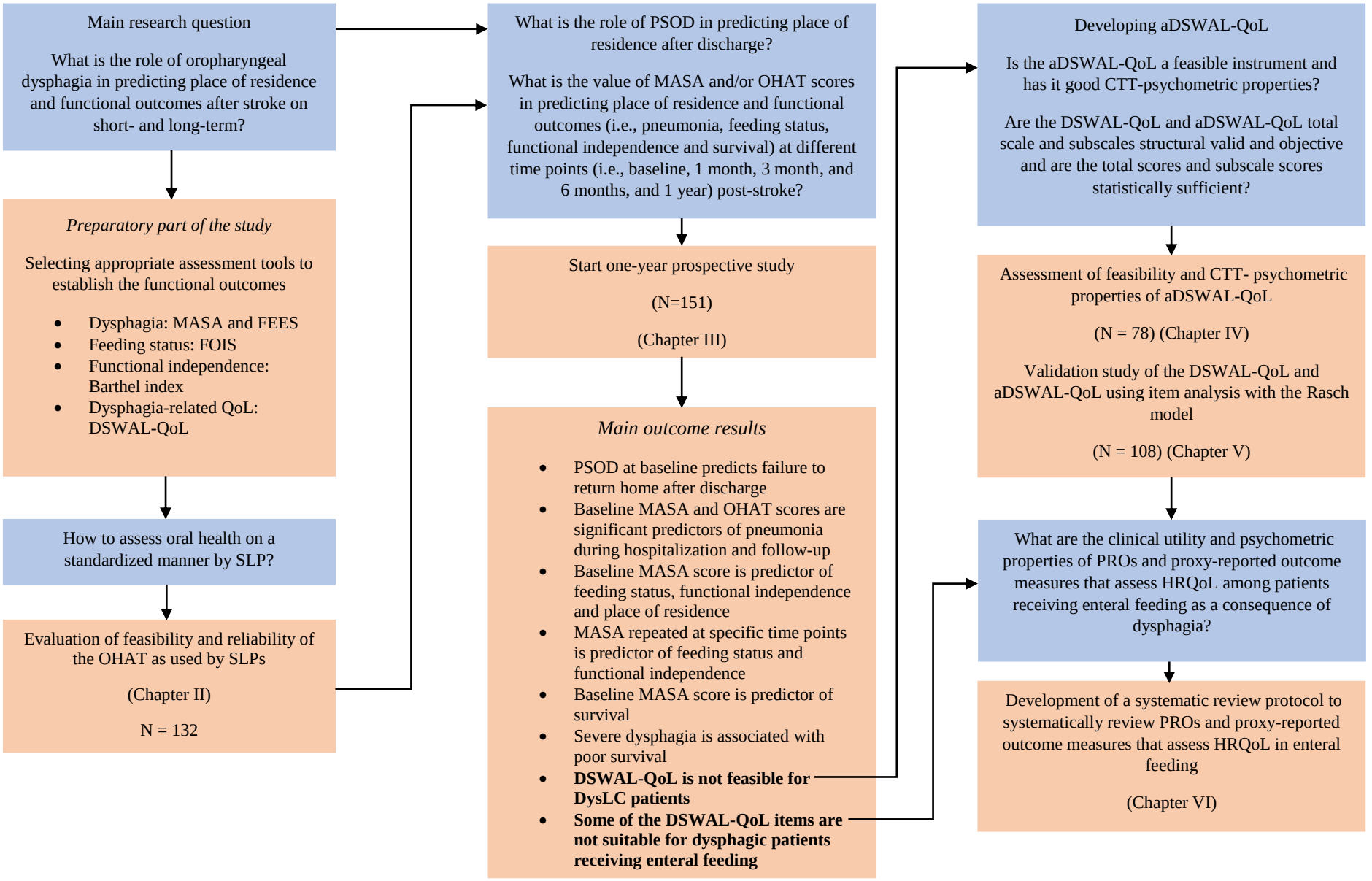
Among patients with severe language and cognitive impairment, self-reporting can be impossible. In such situations, proxy-reported outcome measures may be used to assess the patient's HRQoL: proxies can provide relevant information about how they think the patient would report on their HRQOL. Defined as health-care providers, spouses, parents, or relatives closely involved with the patient [84], proxies may be an alternative source of information about a patient's HRQoL that could otherwise be lost [47]. However, relying on proxy-provided information is only useful if proxies are able to report on the different HRQoL domains and if their ratings are sufficiently reliable [85]. Patient-proxy agreement should be examined to evaluate the reliability of the proxy's ratings [47]. This requires that the PRO and the accompanying proxy-reported outcome instrument measure the same constructs with identical items and that these instruments are valid and reliable [86].

1.3 Research aims

The main aim of this thesis was to investigate the role of oropharyngeal dysphagia in predicting place of residence and functional outcomes (i.e., feeding status, functional independence, pneumonia, survival and HRQoL) on short- and long-term after stroke (cf. study II, described in Chapter III). An important methodological aspect to be considered in the preparatory phase of this prospective long-term study was to select appropriate standardized clinical assessment tools that enable to assess the presence and severity of dysphagia, feeding status, oral health, functional independence and dysphagia-related QoL of the involved stroke patients. We were particularly interested in the value of the Mann Assessment of Swallowing Ability (MASA) and of an oral health assessment instrument to predict functional outcomes after stroke. The MASA is a standardized clinical swallowing examination (CSE) that has been specifically validated for assessing PSOD [87-89]. However, there were no standardized oral health assessment tools available for use by SLPs. The Oral Health Assessment Tool (OHAT) has been shown a valid and reliable tool that can be used by non-dental professionals to screen the oral health status of elderly people – even those with cognitive impairment – and to make appropriate and timely referrals to a dentist or a dental hygienist [90]. We therefore established the feasibility and reliability of the OHAT to be used by SLPs (cf. study I, described in Chapter II), prior to the start of the prospective long-term study (study II). For this second study, we investigated more variables than those that will be reported in this thesis. For example, we investigated oral health-related QoL, the presence and risk of pressure ulcers and the association between premorbid factors and the presence of PSOD. We also investigated dysphagia-related QoL, using the DSWAL-QoL. The DSWAL-QoL was completed by patients having PSOD at different time points during the one-year follow-up study (i.e., 1 month, 3 months, 6 months and 1 year post-stroke). However, one of the outcomes resulting from this prospective longitudinal study was that dysphagic patients suffering from additional language and/or cognitive impairment (DysLC) often struggled to complete this PRO. Therefore, we decided to develop an adjusted DSWAL-QoL questionnaire (aDSWAL-QoL) for use with DysLC patients and to establish its feasibility and CTT-psychometric properties (i.e., reliability and validity) (cf. study III, reported in Chapter IV). Since item scores of both the DSWAL-QoL and aDSWAL-QoL are summated into subscale and total scale scores, it was important to test both the DSWAL-QoL and aDSWAL-QoL scales against the Rasch model. The purpose of study IV (reported in Chapter V) was therefore to assess the structural validity and objectivity of both the DSWAL-QoL and aDSWAL-QoL total scales and subscales and the statistical sufficiency of

the total scores and subscale scores using item analysis with the Rasch model. Other outcomes of the prospective one-year follow-up study (study II) and the aDSWAL-QoL and DSWAL-QoL validation studies (study III and IV) were the observations that (1) not all items of the DSWAL-QoL and aDSWAL-QoL were suitable for patients receiving enteral feeding and (2) self-reporting was impossible in patients with very severe language impairment and/or cognitive disorders. We were therefore interested in available PROs and proxy-reported outcome measures that assess HRQoL among patients receiving enteral feeding as a consequence of dysphagia. As a means to select the most suitable PRO and proxy-reported outcome measure for use with this population group, it is important to evaluate the psychometric properties and the clinical utility of these measurement instruments to make recommendations for use in clinical practice and research [58]. We therefore developed a systematic review protocol (study V) that will be used to systematically review PROs and proxy-reported outcome measures that assess HRQoL among patients receiving enteral feeding, (1) regardless the cause of receiving enteral feeding and (2) as a consequence of dysphagia (cf. Chapter VI). The flow chart illustrates the research process of this thesis regarding its research questions and methods/data collection (Fig. 1.1). The rationale and specific research aims/questions for each study are summarized separately in the subsequent sections.

Fig. 1.1 Flow chart, showing the research process with its research questions and methods/data collection



Study I ***Oral health screening: feasibility and reliability of the Oral Health Assessment Tool as used by speech pathologists.***

Daily oral health care in long-term institutions is often viewed as a nursing task [27]. However, the importance of maintaining adequate oral health has long been recognized by speech pathologists (SPs) (i.e., SLPs), who primarily evaluate the motor and sensory functioning of the oral cavity structures in speech and swallowing [91]. Based on the scope of practice [92, 93] and the educational background [94] of SLPs, they play a supplementary role in oral health promotion [27] and in the evaluation of oral health and dentition [95]. Screening, assessment and diagnosis of swallowing disorders are activities within the scope of practice for SLPs. Typical CSEs include an evaluation of oral motor function [88,95, 96] and/or dentition [95]; however, despite the importance of oral health in the prevention of aspiration pneumonia, especially in individuals with dysphagia [26], there is no standardized method for screening the oral health of patients as part of such an examination. Furthermore, SLPs have expertise in communication management [92, 93] and are trained to be effective managers of communication difficulties experienced by patients who may have cognitive impairment. Such impairment may lead to behavioral responses that are often deemed by care staff to be uncooperative and seen as refusal of oral assessment and care. The dental literature reports a variety of communication strategies that may be utilized to assist in the completion of an oral assessment or dental examination [97], and SLPs who are experienced in the domain of communication management [92, 93] are also suitable professionals for performing oral health screenings.

As reported in previous studies [90, 97], reliable and valid oral health assessment tools have been developed for use by non-dental professionals such as nurses, personal care attendants and allied health or medical professionals [90, 98-102]. Although they are called ‘assessment tools’, they should actually be considered ‘screening instruments’, because they differ from dental examinations performed by qualified dentists [97]. The OHAT was initially adapted from the Kayser-Jones Brief Oral Health Status Examination (BOHSE) [98, 100] by Chalmers et al. [90] and then subsequently modified by the Halton Region’s Health Department [103, 104]. The reliability and validity of the OHAT have been established for use by caregivers (i.e., nurses and personal care attendants) for residents in residential care facilities [90]; however, no data regarding its use by SLPs are available.

The purpose of this study (reported in Chapter II) was therefore to investigate (1) the feasibility and (2) the inter-rater reliability, test-retest reliability and intra-rater reliability of the OHAT as used by SLPs.

Study II ***The role of oropharyngeal dysphagia in predicting place of residence and functional outcomes after stroke: a one-year prospective follow-up study***

Studies have investigated the outcomes of PSOD during hospitalization and post-discharge [12, 14, 105-107], and during follow-up [1, 14, 15, 19, 108-110]. However, the best practices for the management (including diagnosis and investigation) of PSOD remain undefined [2]. Furthermore, the reported short- and long-term outcomes associated with PSOD are often not consistent among studies [18], which can be explained by variations in the identification methods of dysphagia, the timing of the initial evaluation after stroke, the lesion location [18], and the manner in which long-term outcomes are assessed. One of the methods for evaluating swallowing capability in stroke patients is the use of bedside screening tests, of which many of them are available [111]. However, there is no consensus regarding the preferred screening protocol, and few of these tests have sufficient psychometric properties [87]. Furthermore, there is limited evidence regarding the benefits of using dysphagia screening tools on health outcomes [112]. We therefore decided to use the MASA for this study, as this is a CSE with good psychometric properties [88, 89, 113] and specifically validated for assessing PSOD [15, 87, 89]. Furthermore, we were particularly interested in the value of the MASA in predicting functional outcomes. The MASA provides a thorough examination of the oral, pharyngeal, and laryngeal anatomy and physiology, establishes a foundation for the construction of a treatment plan, and allows the determination of the appropriate diet for each patient [114, 115]. The manner in which long-term outcomes are assessed can also influence outcome predictions. Most studies have used retrospective methods [108-110], chart reviews [108-110] and/or interview methods [15, 19, 109] to document the outcomes of PSOD and/or feeding status. Furthermore, the interval between successive measurements was often too long [15, 116]. Consistency in assessing swallowing and evaluating functional independence at regular intervals during follow-up [15] enables better descriptions of and predictions regarding who will have a persistent swallowing disorder, who will require a modified oral intake, who will have decreased functional independence, and who is at risk of long-term institutionalization. Furthermore, these improvements will help patients and their families to better understand the course of the disease [11]. Therefore, longitudinal datasets with

comprehensive measurements performed at different time points are required to build models that allow accurate predictions of mortality and short- and long-term outcomes [116] to enable the planning of future care and support [15]. The predictive value of the OHAT for functional outcomes post-stroke was also a key issue. As poor oral health is a contributing factor to the development of post-stroke pneumonia [17, 20], it was important to screen patients for oral health. The feasibility and the reliability of the OHAT as used by SLPs was established in a previous study (reported in Chapter II) [117].

The primary purpose of this study (cf. Chapter III) was to determine whether the presence of PSOD at baseline (i.e., during the first swallowing examination after the onset of stroke) could predict place of residence after discharge. A secondary aim of this study was to evaluate whether baseline MASA and OHAT scores could predict the occurrence of pneumonia during hospitalization and follow-up. A third purpose of this study was to examine whether the baseline MASA score could predict patient outcomes in terms of feeding status, functional independence and place of residence during a one-year follow-up (i.e., at 1 month, 3 months, 6 months and 1 year). Because the MASA was performed at different time points during the subsequent year, we also investigated whether the relationships among the MASA score and feeding status, functional independence and place of residence changed over time by studying the associations among these factors at each time point. A fifth aim of the study was to investigate survival as a function of baseline MASA and/or OHAT scores while correcting for the other variables (e.g., the occurrence of pneumonia during the one-year follow-up).

Study III ***Feasibility and Psychometric Properties of the Adjusted DSWAL-QoL Questionnaire for Dysphagic Patients with Additional Language and/or Cognitive Impairment: Part I***

As previously written, one of the outcomes of the prospective, long-term study (cf. study II, reported in Chapter III) was the observation that many stroke patients struggled with completing the DSWAL-QoL or were unable to complete them due to comprehension difficulties related to aphasia or poor vision. Extant literature also revealed that up to 50% of dysphagic patients frequently needed assistance to complete this questionnaire [48, 118, 119]. Although not specified in these SWAL-QoL studies [48, 118, 119], the need for assistance could have been due to the presence of language disorders [48, 118, 119] or cognitive impairment [118, 119]: patients with neurological disorders were involved in the validation process. Those studies [48, 118, 119] provided limited details about the assistance required despite the fact that this information could allow improvement of the scale's feasibility for a specific population group,

namely DysLC patients. To increase the feasibility of the SWAL-QoL, Lemmens et al. [51] developed an interview-administered version. Although this interview version can be considered an important step forward, the format deviates from a self-administered questionnaire, which is conventionally preferred [77]. Such an adjusted DSWAL-QoL-questionnaire (aDSWAL-QoL) – based on aphasia- and cognitive-friendly recommendations from the literature [55, 70, 71, 120-123] – could facilitate self-reporting in DysLC patients and, consequently, increase the reliability of the ratings.

The primary purpose of this study (reported in Chapter IV) was to develop the aDSWAL-QoL and to examine its feasibility and psychometric properties (i.e., internal consistency, test-retest reliability, and criterion validity) to be used by DysLC patients. A second aim of this study was to predict the need for assistance to complete the aDSWAL-QoL based on cognitive impairment, language comprehension, age, group, and functional dependency.

Study IV ***Validation of the Dutch version of the Swallowing Quality-of-Life Questionnaire (DSWAL-QoL) and the adjusted DSWAL-QoL (aDSWAL-QoL) using item analysis with the Rasch model: a pilot study***

The CTT-psychometric properties of the SWAL-QoL, the DSWAL-QoL, and aDSWAL-QoL have been demonstrated to be sufficient [48, 63, 124]. The CTT-psychometric assessment included an examination of internal consistency based on Cronbach's alpha, test-retest reliability and criterion validity via the intraclass correlation coefficient (ICC), and/or construct validity based on principal component analysis (PCA) techniques [48, 63, 124]. Some drawbacks related to CTT methods are recognized [125, 126], such as test and sample dependence [125-127] and the assumption of equal weight for all of the items even if there is a difference in the level of difficulty [127]. The scale's total sum score is based on ordinal values and the standard error of measurement is assumed to be constant [83, 126], in contrast to the Rasch methodology.

The Rasch model has been considered the gold standard against which scales summarizing item responses must be tested [80]. If the observed data fit the model, the following can be concluded: interval data have been generated, the measurement scale demonstrates structural validity and objectivity, and the total score is statistically sufficient [80]. In addition to identifying measurement weaknesses, analysis with the Rasch model provides potential solutions for scale improvement [127-129]. The SWAL-QoL nor

its translated (e.g., DSWAL-QoL) or adapted (e.g., aDSWAL-QoL) versions have been tested against this measurement model.

The purpose of study IV (described in Chapter V) was to assess the structural validity and objectivity of both the DSWAL-QoL and aDSWAL-QoL total scales and subscales and the statistical sufficiency of the total scores and subscale scores using item analysis (*note:* for ease of readability, we will sometimes use the term 'Rasch analysis' in this thesis) with the Rasch model.

Study V ***Patient-reported and proxy-reported outcome measures for the assessment of health-related quality of life among patients receiving enteral feeding: a systematic review protocol***

An important observation resulting from studies II, III and/or IV, was that some of the items of the DSWAL-QoL and aDSWAL-QoL were not suitable for dysphagic patients receiving enteral feeding. Those patients could not provide a response on items such as 'food sticking in your throut' or 'choking when you take liquids' because they were not able to take food or liquid by mouth and had to rely on tubefeeding for their nutritional support. Caution should therefore be undertaken in the use of these scales with patients receiving enteral feeding. A disease-specific PRO that is specifically developed and validated for this population group is recommended. Furthermore, it is important that this PRO measures the relevant constructs or concepts that are crucial for those patients receiving enteral feeding as a consequence of oropharyngeal dysphagia [130]. Among some of the patients in study II and III, self-reporting was impossible because of severe cognitive and/or language disorders. In these situations, the use of proxy-reported outcome measures allows to reveal information about the patient's HRQoL that may otherwise be lost [47]. As a means to select the most suitable PRO and proxy-reported outcome measure for assessing HRQoL among patients receiving enteral feeding as a consequence of oropharyngeal dysphagia, it is important to evaluate the psychometric properties and the clinical utility of these measurement instruments [58]. Based upon an initially limited search in MEDLINE, the Cochrane Database of Systematic Reviews and JBI COnNECT+ in February 2016, to our knowledge there are no systematic reviews that have reviewed the psychometric properties and the clinical utility of PROs and proxy-reported outcome measures assessing HRQoL among patients receiving enteral feeding.

We therefore developed a systematic review protocol (outlined in Chapter VI) that will be used to systematically review the psychometric properties and the clinical utility of PROs and proxy-reported

outcome measures (i.e., questionnaires or structured interviews) that assess HRQoL among patients receiving enteral feeding to make recommendations for its use in clinical practice and research.

In summary, the research aims of this thesis were:

1. To evaluate the feasibility and reliability of the OHAT as used by SPs (Chapter II).
2. To determine whether the presence of PSOD at baseline could predict place of residence after discharge (Chapter III).
3. To evaluate whether baseline MASA and OHAT scores could predict the occurrence of pneumonia during hospitalization and follow-up (i.e., 1 month, 3 months, 6 months, and 1 year) (Chapter III).
4. To evaluate whether the baseline MASA score could predict patient outcomes in terms of feeding status, functional independence and place of residence during the one-year prospective study (Chapter III).
5. To evaluate whether the baseline MASA score and the MASA performed at different time points during the subsequent year could predict patient outcomes in term of feeding status, functional independence and place of residence during the one-year follow-up period (Chapter III).
6. To investigate survival following stroke as a function of baseline MASA and/or OHAT scores (Chapter III).
7. To develop an adjusted DSWAL-QoL to be used by DysLC patients (Chapter IV).
8. To examine the psychometric properties of the aDSWAL-QoL using CTT-approaches (Chapter IV).
9. To examine the structural validity and objectivity of both the DSWAL-QoL and aDSWAL-QoL total scales and subscales, and the statistical sufficiency of the total scores and subscale scores using item analysis with the Rasch model (Chapter V).
10. To develop a systematic review protocol in order to systematically review the psychometric properties and the clinical utility of PROs and proxy-reported outcome measures that assess HRQoL among patients receiving enteral feeding to make recommendations for use in clinical practice and research (Chapter VI).

CHAPTER II

ORAL HEALTH SCREENING: FEASIBILITY AND RELIABILITY OF THE ORAL HEALTH AS USED BY SPEECH PATHOLOGISTS

Study I has been published in:

Simpelaere IS, Van Nuffelen G, Vanderwegen J, Wouters K, De Bodt M. Oral health screening: feasibility and reliability of the oral health assessment tool as used by speech pathologists. International Dental Journal. 2016;66(3):178-89.

ABSTRACT

The aim of this study was to investigate the feasibility and reliability of the Oral Health Assessment Tool (OHAT) as used by speech pathologists, to become part of a comprehensive clinical swallowing examination. A multicentre study in 132 elderly subjects was conducted by speech pathologists. The inter-rater, test-retest and intra-rater reliabilities of the OHAT were assessed in R statistics, version 3.0.1. Intraclass correlation coefficients (ICCs) were used for the total OHAT, and Kappa statistics were used for the individual categories. Total OHAT scores showed good inter-rater (ICC = 0.96), intra-rater (ICC ≥ 0.95) and test-retest (ICC ≥ 0.78) agreement. The inter-rater Kappa statistics were almost perfect ($\kappa \geq$ 0.83) for seven of the eight individual categories of the OHAT and perfect for 'dental pain' (κ = 1.00). The test-retest Kappa statistics indicated excellent agreement for 'natural teeth' and 'dentures' ($\kappa \geq 0.86$). The intra-rater per cent agreement was excellent for all categories except 'gums and tissues'. This is the first study to examine the feasibility and reliability of the OHAT as used by speech pathologists. As the results showed both good feasibility and reliability, the OHAT has the potential to add to the clinical swallowing examination. However, future research investigating actual referral strategies and adaptation of care strategies following assessment with OHAT is needed.

Keywords: Dental care, Elderly, Feasibility, Oral health screening, Reliability

2.1 INTRODUCTION

2.1.1 Oral health care and speech pathologists

The oral health of patients with dysphagia is concerning, particularly in elderly patients [24, 32], because when poor oral health and oral diseases are combined with the presence of swallowing and feeding problems, poor functional status, underlying diseases and an increasing age, the risk of aspiration pneumonia is highest [24, 25]. The importance of maintaining adequate oral health has long been recognised by speech pathologists (SPs), who primarily evaluate the motor and sensory functioning of the oral cavity structures involved in speech and swallowing [91]. Based on their professional knowledge of oral anatomy and physiology [94], they pay particular attention to oral health and dentition [95], which are essential for swallowing and speech production [31].

Screening, assessment, and diagnosis of swallowing disorders are activities within the scope of practice for SPs [92, 93]. Swallowing disorders resulting from oropharyngeal dysfunction [131], known as oropharyngeal dysphagia, can be caused by oral abnormalities, such as dental malocclusion, as well as oral-motor dysfunction [93]. Typical clinical swallowing examinations include an evaluation of dentition [95] and oral-motor function [88, 95, 96]; however, despite the importance of oral health in the prevention of aspiration pneumonia, especially in individuals with dysphagia [26], there is no standardised method for screening the oral health of patients as part of such an examination.

Aspiration pneumonia is defined as the development of pneumonia after the aspiration of colonised oropharyngeal material into the larynx and lower respiratory airways [22, 23], and it occurs in dysphagic patients, who are at increased risk for oropharyngeal aspiration [22]. Poor oral health is an important contributing factor to the development of aspiration pneumonia [22-24, 26-28]. Therefore, oral care in dysphagic patients is essential, and oral health care interventions have the potential to diminish the risk of aspiration pneumonia [28, 29] and its associated elevated mortality risk [29, 30].

Oral health and dental status are also critical during the preparatory phase of swallowing, particularly for mastication [32]. Maintaining functional units (i.e. pairs of opposing mandibular and maxillary teeth, particularly natural teeth) is crucial for masticatory function [32, 33]. The degree of masticatory function not only determines food selection [34, 35] but also influences nutritional status [34-37].

Although the provision of daily oral hygiene support is primarily considered the responsibility of nursing staff [27, 98], SPs are also in a position to detect oral ailments during their routine assessments [27, 95]. The literature reports that declines in oral health status often go unnoticed until oral health becomes visibly poor [27]. Oral care is often poor in dysphagic patients, whether these patients reside in a hospital or in rehabilitation or residential facilities [32]. Therefore, a multidisciplinary approach to enhancing the quality of oral health has been suggested [132], and SPs can provide valuable contributions because of their existing scope of practice [92, 93].

SPs have expertise in communication management [92, 93] and are trained to be effective managers of communication difficulties experienced by patients who may have cognitive impairment. Such impairment may lead to behavioural responses that are often deemed by care staff to be uncooperative and seen as refusal of oral assessment and care. The dental literature reports a variety of communication strategies that may be utilised to assist in the completion of an oral assessment or dental examination [97], and SPs who are experienced in the domain of communication management [92, 93] are also suitable professionals for performing oral health screenings.

2.1.2. Oral Health Assessment

As reported in previous studies [90, 97], reliable and valid Oral Health Assessment Tools (OHATs) have been developed for use by non-dental professionals such as nurses, personal care attendants, and allied health or medical professionals [90, 98-102]. Although they are called ‘assessment tools’, they should actually be considered ‘screening instruments’ because they differ from dental examinations performed by qualified dentists [97]. These tools are meant to screen the oral health status of a patient to make appropriate and timely referrals to a dentist or a dental hygienist [90]. To be in accordance with the terminology used in the literature [90, 97, 100], in this study, we use the term ‘assessment’; however, this term refers to an oral health screening that addresses patients’ dental needs [97]. The Kayser-Jones Brief Oral Health Status Examination (BOHSE) [98, 100] and the subsequently developed OHAT [90] are two instruments with proven validity and reliability that can be used by residential care staff for various patients, including those with cognitive impairment. The OHAT was initially adapted from the BOHSE by Chalmers et al. [90] and then subsequently modified by the Halton Region’s Health Department [103, 104]. The success of the OHAT is that it requires minimal training [90, 98, 133, 134] and it is therefore a

feasible instrument for SPs to use for oral health screening. Furthermore, the psychometric properties of the modified OHAT have not been investigated previously; no data regarding its use by SPs are available. The purpose of this study was to investigate (1) the feasibility and (2) the inter-rater reliability, test-retest reliability and intra-rater reliability of the modified OHAT as used by SPs.

2.2 METHODS

2.2.1 Oral Health Assessment Tool

As previously reported in the initial study by Chalmers et al. [90], the OHAT consists of eight categories ('lips', 'tongue', 'gums and tissues', 'saliva', 'natural teeth', 'dentures', 'oral cleanliness', and 'dental pain') with three possible scores (0: healthy, 1: some changes present and 2: unhealthy condition) [90]. Scoring of each category is based on structured observation with clear operational definitions [90, 103, 104, 135-137]. A score of 1 or 2 for any of the specifically marked categories (starred and underlined) mandates referral to an oral health professional (dentist, dental hygienist, or denturist) [103, 104]. The total score is the sum of the various subscores. Based on the screening results, staff members can determine whether patient needs can be met by daily oral care based on the development of an Oral Hygiene Care Plan [138, 139], or if referral to an oral health professional [103, 104] should be instituted. The materials required to perform the screening include only clean gloves and an adequate light source (daylight or artificial) [90]. In this study, the modified OHAT tool was used [103, 104] (Appendix 2.1).

2.2.2 Subject recruitment

Subjects were recruited based on specific criteria. The inclusion criteria were: (1) staying in residential care settings (assisted living facilities and nursing homes); or (2) being hospitalised in an acute geriatric department. A consecutive sample was used to recruit subjects in the acute geriatric department and nursing homes. To recruit subjects in the assisted living facilities, a convenience sample was used due to practical reasons. A large variation in dental status in all these settings was expected [140, 141]. Institutionalised and home-bound elderly are among the most dentally neglected subjects and have poorer oral health compared with elderly individuals living independently [98, 142]. Age and cognitive ability were not used as exclusion criterion, as the OHAT was specifically developed for use in elderly patients

with varying degrees of cognitive impairment. The level of cognitive impairment was determined from a participant's chart review based on their medical diagnosis or their Mini Mental State Examination (MMSE) score [143]. If the MMSE score was unavailable, the principle investigator, who is experienced in cognitive disorders, determined the level of cognitive impairment through extensive observation of language comprehension, executive functioning, attention and consciousness. Three nursing homes, two assisted living facilities and one acute geriatric department in a general hospital participated in the study. There were some differences in the provision of oral care between these institutions. Subjects residing at nursing homes were dependent on a delegated nurse for oral health care or from any other available nurse. Subjects in the acute geriatric department also received oral care from any available nurse, whereas subjects residing at the assisted living facilities were mainly responsible for their own oral health care.

Ethical approval was granted by two independent ethics boards, namely the Committee for Medical Ethics of the University Hospital of Antwerp and the Ethics Committee of the H. Hart Hospital of Roeselare-Menen (B300201215080). The study was conducted in full accordance with the World Medical Association Declaration of Helsinki. Prior to the start of the study, verbal and written consent was obtained from all subjects or from the legal representative or the appropriate directors of nursing if the patient could no longer provide written consent. The consent procedure for the study and the test-retest reliability evaluation was approved by the two previously mentioned ethics committees. Subjects with incomplete baseline data were not withheld from the test-retest evaluation.

2.2.3 Procedures and measures

The study was divided into two major parts: the preparatory study and the actual investigation. The flow chart presented in Appendix 2.2 shows the different steps of the study.

Part 1: Preparatory study

The preparatory study consisted of two phases.

Phase 1

A 3-hour training session with visual instruction was delivered to three SPs by the principle investigator using publicly available visual training resources [90, 103, 104, 135-137]. The SPs had extensive experience in dysphagia management. The training was followed by trial assessments of three subjects in the acute geriatric department. The scoring of each category was discussed until agreement was reached in accordance with the visual training resources of the OHAT [90, 103, 104, 135-137]. To facilitate OHAT use, a manual with descriptors of the different scores for each category was provided.

Phase 2

Trial assessments of 17 subjects in a nursing home were performed to determine whether the SPs experienced difficulties in assigning a score for a certain category, to determine whether the SPs felt confident in completing the screening by means of the manual and the publicly available visual training resources [90, 103, 104, 135-137], and to facilitate time registration while performing the screening. Each subject was simultaneously screened by the three SPs who independently completed the OHAT. After the completion of all trial assessments, the results were compared. Between Phase 2 of the preparatory study and the start of the actual investigation, there was an interval of 2 weeks to allow further practise with the OHAT tool. The SPs were required to perform the screening on their family members or acquaintances to become familiar with the scoring.

Part 2: Actual investigation

Two weeks after the preparatory part of the study, the actual study was performed over the following 14 weeks (Appendix 2.2). All three SPs went simultaneously to all facilities. During the actual investigation, the feasibility, inter-rater reliability and test-retest reliability were assessed. Following the actual study, the intra-rater reliability was assessed using videotapes.

Feasibility of the OHAT

The feasibility of the OHAT was defined based on the time required to complete the OHAT, the ability to score the categories of the OHAT and possible problems in administering the OHAT. All SPs completed a semi-structured questionnaire at three time points, namely, at the end of Phase 2 of the preparatory study, at baseline, and at the end of the actual study; this questionnaire was similar to the original questionnaire from the study by Chalmers et al. [90] with the addition of a few open-ended questions. As in the original questionnaire [90], a four-point Likert scale, ranging from 'strongly disagree' to 'strongly agree', was applied to rate the statements. The questionnaire is presented in Appendix 2.3. For each participant in the actual study, the SPs were asked to register the time taken to complete the screening evaluation.

Reliability of the OHAT

Reliability of the OHAT was measured by evaluating the inter-rater reliability, test-retest reliability and intra-rater reliability.

Inter-rater reliability

One-hundred and thirty-five subjects were screened by three SPs, simultaneously, but independently, to evaluate the inter-rater reliability of the OHAT. As in Phase 2 of the preparatory study, each SP was blinded to the scores assigned by the other SPs. The subjects were screened in a sitting or supine position.

Test-retest reliability

Based on advice from a medical statistician, test-retest reliability was assessed in 46 subjects. These subjects were randomly selected from the two nursing homes and one assisted living facility because individuals at these facilities tend to have a more stable health status. Two SPs re-evaluated the selected subjects during a second screening, 2 weeks later. This interval was sufficient to avoid memory effects and the occurrence of genuine oral health status changes in the subjects [144].

Intra-rater reliability based on videotapes

Intra-rater reliability was also investigated based on independent videotape ratings by two SPs at three different time points, with at least 14-day intervals. Ten subjects were randomly selected to be videotaped at a frontal angle with the subject's head and mouth in the frame to enable observation of the oral cavity. For every subject, a 10-minute recording was performed. The subjects were asked to open their mouth. The camera zoomed in on particular parts of the mouth (i.e., the lips, tissues, tongue, dentures and natural teeth), moving from one side to another. Afterward, dentures were removed and the camera was zoomed in on the upper and lower side of the dentures to obtain a clear view of oral hygiene. Videos were chosen over photographs, because videotapes have proven to be useful for allowing multiple raters to observe the same performance [144]. The rating of the videotapes was performed in a random order, 14 days after completing the actual study.

2.2.4 Data analysis

Population characteristics (i.e., age and gender) were assessed for normality using the Shapiro–Wilk test and QQ-plots. Differences in age, gender and the presence of cognitive impairment according to place of residence were assessed using the Kruskal–Wallis test (age) and the chi-square test (gender and cognitive impairment). The distribution of the scores for individual categories at baseline was assessed for the three different settings, and the Kruskal–Wallis test was used to detect significant differences between these settings. The frequency distribution of the total OHAT score was evaluated at baseline. Floor and ceiling effects associated with the total OHAT score were considered to be present if more than 15% of the subjects achieved the lowest or highest possible score [59]. To evaluate the inter-rater reliability of the OHAT, the intraclass correlation coefficient (ICC) and a two-way random-effects model with measures of absolute agreement ($ICC_{absolute\ agreement}$) [59, 145] were used for the total OHAT scores. The inter-rater reliability of the individual OHAT categories was assessed using Fleiss Kappa. The ICC with a one-way random-effects model with measures of absolute agreement was used to assess the test-retest stability, an evaluation of intra-rater reliability [144], and the intra-rater reliability based on individual evaluations of the videotapes at different times. For individual categories, Cohen's Kappa and Fleiss Kappa were, respectively, used to assess the test-retest and intra-rater reliabilities based on videotapes. The associated 95% confidence intervals were calculated using a bootstrap of 1,000 samples. As suggested in previous studies [144], ICC values higher than 0.75 are indicative of good reliability, whereas values lower than

0.75 represent poor-to-moderate reliability. Kappa statistic values < 0.00 were interpreted as indicating poor agreement, 0.00–0.20 as indicating slight agreement, 0.21–0.40 as indicating fair agreement, 0.41–0.60 as indicating moderate agreement, 0.61–0.80 as indicating substantial agreement and > 0.80 indicating almost perfect [144] or excellent agreement [146]. Because Fleiss and Cohen's Kappa depend heavily on the observed marginal frequencies, these scores can be misleading and should be treated with caution. To aid interpretation, we therefore additionally reported the per cent agreement for each individual category. The statistical analyses were performed using R 3.0.1 (R Foundation for Statistical Computing, Vienna, Austria) and SPSS 20.0 (IBM, SPSS, Inc., Chicago, IL, USA), with $P < 0.05$ considered significant. Only complete screenings were included in the data analysis.

2.3 RESULTS

Of the 135 subjects, 132 completed the screening. Three subjects (two in nursing homes and one in the acute geriatric department) were excluded because of dementia-associated behavioural problems. In total, 70 subjects were recruited from nursing homes, 30 from assisted living facilities and 32 from the acute geriatric department. The demographic characteristics of the subjects are presented in Table 2.1.

Table 2.1 Demographic characteristics of the subjects (n = 132)

Characteristics		Nursing home (n = 70)	Acute geriatric department (n = 32)	Assisted living places (n = 30)
Age*	Mean (SD)	83.4 (7.2)	84.3 (7.3)	86.2 (7.1)
	Median (min–max)	84.5 (63–101)	85.5 (62–100)	87.5 (63–101)
Gender*	Male, *n* (%)	18 (25.7)	11 (34.4)	8 (26.7)
	Female, *n* (%)	52 (74.3)	21 (65.6)	22 (73.3)
Cognitive impairment**	Presence, *n* (%)	54 (77.1)	14 (43.8)	7 (23.3)
	Absence, *n* (%)	16 (22.9)	18 (56.3)	23 (76.7)

Values are given as mean and standard deviation (SD) and median (min–max) for age and as *n* (%) for gender and for cognitive impairment. Analyses were performed using the Kruskal–Wallis test for equality of means for age and the chi-square test for gender and cognitive impairment.
*$P > 0.05$
** $P < 0.001$

No significant differences ($P > 0.05$) in age and gender were observed with regard to the place of residence. Significant differences ($P < 0.001$) in cognitive status were found between the three settings.

Significantly more nursing home residents had cognitive impairment. Among the subjects from the nursing homes and the acute geriatric department, the presence of cognitive impairment was based on a medical diagnosis. However, for seven (23.3%) of the 30 subjects residing at the assisted living facilities, the presence of cognitive impairment was based only on a comprehensive evaluation by the principle investigator. The OHAT score distribution at baseline for the individual categories is shown in Table 2.2.

Table 2.2 Distribution of the scores at baseline for the individual categories of the Oral Health Assessment Tool (OHAT) for all subjects (n = 132)

Category	Score 0		Score 1		Score 2	
	n	%	*n*	%	*n*	%
Lips	112	84.8	16	12.1	4	3.0
Tongue	37	28.0	94	71.2	1	0.8
Gums and tissues	69	52.3	55	41.7	8	6.1
Saliva	104	78.8	28	21.2	0	0
Natural teeth	44	33.3	21	15.9	67	50.8
Dentures	56	42.4	2	1.5	74	56.1
Oral cleanliness	17	12.9	68	51.5	47	35.6
Dental pain	121	91.7	10	7.6	1	0.8

The majority of the subjects scored 0 in the categories ‘lips’, ‘saliva’, ‘dental pain’ and ‘gums and tissues’. With regard to ‘oral cleanliness’ and ‘tongue’, more than half of the subjects scored 1, and most subjects scored 2 in the categories ‘natural teeth’ and ‘dentures’. Irrespective of the place of residence, the score distribution was similar between the categories ‘lips’, ‘tongue’, ‘gums and tissues’ and ‘dental pain’, as shown in Table 2.3. The scores for ‘saliva’, ‘natural teeth’ and ‘dentures’ were different between subjects in the acute geriatric department compared with subjects in the other settings. More than half of the hospitalised subjects scored 2 for ‘natural teeth’ and ‘dentures’, and more subjects (37.5%) scored 1 for ‘saliva’ compared with the subjects in the other settings. Half of the subjects residing in nursing homes scored 2 for ‘oral cleanliness’, in contrast with a majority score of 1 for subjects at the other facilities. Differences in the score distribution were only significant ($P < 0.05$) for ‘saliva’ between the hospitalised patients and the subjects residing in nursing homes and for ‘oral cleanliness’ between the subjects in nursing homes and the subjects residing in assisted living facilities. Figure 2.1 shows the frequency distribution and corresponding mean and median of the subjects’ total OHAT scores. None of the subjects scored higher than 10, and no floor and ceiling effects were present.

Table 2.3 Percentage distribution of Oral Health Assessment Tool (OHAT) scores at baseline for the individual categories regarding the place of residence (n = 132)

Category	Nursing home			Acute geriatric department			Assisted living places		
	Score 0	Score 1	Score 2	Score 0	Score 1	Score 2	Score 0	Score 1	Score 2
Lips	85.7	12.9	1.4	78.1	15.6	6.3	90.0	6.7	3.3
Tongue	31.4	67.1	1.4	28.1	71.9	0.0	20.0	80.0	0.0
GandT	55.7	40.0	4.3	50.0	46.9	3.1	46.7	40.0	13.3
Saliva*	88.6	11.4	0.0	62.5	37.5	0.0	73.3	26.7	0.0
NT	37.1	12.9	50.0	18.8	18.8	62.5	40.0	20.0	40.0
Dent	44.3	2.9	52.9	28.1	0.0	71.9	53.3	0.0	46.7
OC*	11.4	38.6	50.0	12.5	59.4	28.1	16.7	73.3	10.0
DP	90.0	8.6	1.4	90.6	9.4	0.0	96.7	3.3	0.0

Analyses were based on the Kruskal–Wallis test for the score distribution, irrespective of the place of residence.
Dent, dentures; DP, dental pain; GandT, gums and tissues; NT, natural teeth; OC, oral cleanliness.
*P < 0.05.

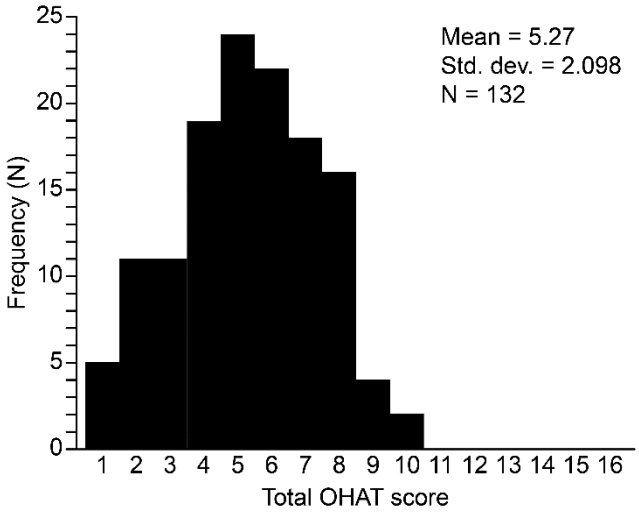


Figure 2.1 Frequency distribution of the total OHAT scores

Feasibility of the OHAT

At baseline, the mean time (standard deviation; range) to completion of the OHAT was 2.45 (1.05; 0.42–6.20) minutes. All SPs thought that the OHAT was simple to use and quick to administer. One SP reported that the category ‘oral cleanliness’ was difficult to score at times. Subjects with dentures were easier to score than subjects with natural teeth. The questionnaire results showed that at both baseline and the end of the actual study, all three SPs ‘strongly agreed’ with nearly all the formulated statements. However, at baseline, one SP only ‘agreed’ with the statement related to the ‘dentures’ category. Comparing the questionnaire results from baseline with those from the end of the preparatory study,

greater diversity was found in the self-perceived ability to score the various categories of the OHAT. During the preparatory part of the study, the SPs 'agreed' with most of the formulated statements, and two SPs 'strongly agreed' with the statements related to 'natural teeth', 'dentures' and 'dental pain'. However, one SP 'disagreed' with the statement "able to complete the 'dentures' category" and another SP 'disagreed' with the statement related to 'oral cleanliness'. None of the statements was rated as 'strongly disagree'. An analysis of the answers to the open-ended questions revealed the following results: (1) there was a lack of information on pairs of teeth in the chewing position; (2) sufficient visual resources were available to complete the OHAT; and (3) the manual was seen as an important contribution because it leads to higher consensus in score assignment. At the end of the actual study, the manual was no longer necessary because of familiarity with the scoring. The presence of referral possibilities at the bottom of the scoring sheet was judged as useful.

Inter-rater reliability

The ICC value for the total OHAT score was 0.96 [95% confidence interval (95% CI) = 0.95–0.97], indicating very good inter-rater reliability [144]. The inter-rater reliabilities of the individual OHAT categories and the advice for referral to an oral health professional are shown in Table 2.4. Kappa statistics neared perfect values ($\kappa \geq 0.83$) for seven of the eight individual categories and the need for referral, and achieved perfect agreement for 'dental pain' ($\kappa = 1.00$).

Table 2.4 Inter-rater reliability data [per cent agreement and Fleiss Kappa (three raters)] for individual categories and for referral to an oral health professional ($n = 132$)

Category	Per cent agreement (95% CI[†])	Fleiss Kappa (95% CI[†])
Lips	0.97 (0.94–0.99)	0.88 (0.80–0.96)
Tongue	0.95 (0.92–0.98)	0.89 (0.81–0.96)
GandT	0.91 (0.87–0.94)	0.83 (0.76–0.90)
Saliva	0.99 (0.96–1.00)	0.95 (0.90–1.00)
NT	0.98 (0.96–1.00)	0.97 (0.94–1.00)
Dent	0.98 (0.96–1.00)	0.97 (0.93–1.00)
OC	0.92 (0.89–0.96)	0.87 (0.81–0.93)
DP	1.00 (1.00–1.00)	1.00 (1.00–1.00)
Refer OHP	0.98 (0.96–1.00)	0.93 (0.84–1.00)

Dent, dentures; DP, dental pain; GandT, gums and tissues; NT, natural teeth; OC, oral cleanliness; Refer OHP, referral to an oral health professional (i.e. dentist, dental hygienist or denturist).
[†]95% confidence interval.

Test-retest reliability

The stability of scores over time, the intra-rater reliability (represented by Cohen's Kappa statistic) and the test-retest per cent agreement for the various categories and for the decision to refer in 46 subjects are shown in Table 2.5. The reliability data are provided separately for each rater. The ICC for the total OHAT score were 0.81 (95% CI = 0.68–0.89) and 0.78 (95% CI = 0.64–0.87) for raters 1 and 2, respectively, indicating good reliability [144]. The test-retest and intra-rater Kappa statistics indicated almost perfect agreement for the categories 'natural teeth' and 'dentures'. A substantial level of agreement was reached for 'oral cleanliness' for rater 1, whereas the agreement was only moderate for rater 2 ($\kappa = 0.55$). The test-retest and intra-rater Kappa statistics were moderate for tongue and saliva for both raters. With regard to 'gums and tissues', moderate scores were obtained by one rater, and the other rater only achieved slight agreement ($\kappa = 0.15$). 'Lips' showed fair agreement ($\kappa = 0.38$) and 'dental pain' showed slight agreement ($\kappa = 0.14$) for both raters. Regarding referral, the agreement was substantial for rater 1 ($\kappa = 0.69$) and was fair for rater 2 ($\kappa = 0.39$).

Table 2.5 Test-retest reliability (per cent agreement and Kappa statistics for individual categories and for referral to an oral health professional; *n* = 46)

Category	Rater 1		Rater 2	
	Per cent agreement (95% CI†)	Cohen's Kappa (95% CI†)	Per cent agreement (95% CI†)	Cohen's Kappa (95% CI†)
Lips	0.89 (0.80–0.98)	0.38 (-0.10–0.79)	0.89 (0.78–0.96)	0.38 (-0.10–0.78)
Tongue	0.78 (0.67–0.89)	0.44 (0.14–0.69)	0.80 (0.70–0.91)	0.51 (0.23–0.76)
GandT	0.72 (0.59–0.83)	0.42 (0.15–0.65)	0.59 (0.46–0.74)	0.15 (-0.10–0.42)
Saliva	0.91 (0.83–0.98)	0.52 (-0.05–0.90)	0.91 (0.83–0.98)	0.45 (-0.07–0.85)
NT	0.93 (0.85–1.00)	0.89 (0.75–1.00)	0.91 (0.83–0.98)	0.86 (0.70–0.97)
Dent	0.96 (0.89–1.00)	0.91 (0.76–1.00)	0.93 (0.85–1.00)	0.87 (0.70–1.00)
OC	0.83 (0.70–0.93)	0.69 (0.47–0.88)	0.74 (0.61–0.87)	0.55 (0.30–0.76)
DP	0.85 (0.74–0.93)	0.14 (-0.14–0.55)	0.85 (0.74–0.93)	0.14 (-0.14–0.55)
Refer OHP	0.93 (0.87–1.00)	0.69 (0.18–1.00)	0.93 (0.87–1.00)	0.39 (0.22–1.00)

Dent, dentures; DP, dental pain; GandT, gums and tissues; NT, natural teeth; OC, oral cleanliness; Refer OHP, referral to an oral health professional (i.e. dentist, dental hygienist or denturist).
†95% confidence interval.

Intra-rater reliability based on videotapes

The ICC for intra-rater reliability for the total OHAT score, based on ratings of 10 videotapes, showed good reliability for both raters [rater 1: ICC = 0.95 (95% CI = 0.88–0.99); rater 2: ICC = 0.96 (95% CI = 0.89–0.99)]. Table 2.6 shows the per cent agreement for the individual categories and for referrals. As almost no variance was observed in the scoring, the Fleiss Kappa could not be calculated.

Table 2.6 Per cent agreement for individual categories and for referral to an oral health professional (n = 10)

Category	Rater 1	Rater 2
	Per cent agreement (95% CI†)	Per cent agreement (95% CI†)
Lips	0.93 (0.80–1.00)	1.00 (1.00–1.00)
Tongue	1.00 (1.00–1.00)	0.93 (0.80–1.00)
GandT	1.00 (1.00–1.00)	0.73 (0.53–0.93)
Saliva	1.00 (1.00–1.00)	1.00 (1.00–1.00)
NT	1.00 (1.00–1.00)	1.00 (1.00–1.00)
Dent	0.93 (0.80–1.00)	1.00 (1.00–1.00)
OC	1.00 (1.00–1.00)	1.00 (1.00–1.00)
DP	0.85 (0.74–0.93)	1.00 (1.00–1.00)
Refer OHP	1.00 (1.00–1.00)	1.00 (1.00–1.00)

Because of the lack of significant variance, Kappa was not calculated. Dent, dentures; DP, dental pain; GandT, gums and tissues; NT, natural teeth; OC, oral cleanliness; Refer OHP, referral to an oral health professional (i.e. dentist, dental hygienist or denturist).
†95% confidence interval.

2.4 DISCUSSION

Daily oral health care in long-term institutions is often viewed as a nursing task [27]. However, based on the scope of practice [92, 93] and the educational background [94] of SPs, they play a supplementary role in oral health promotion [27] and in the evaluation of oral health and dentition [95]. During a clinical swallowing examination, SPs inspect the oral cavity and dentition; however, a standardised assessment of oral health is typically not implemented. Therefore, the aim of this study was to evaluate the feasibility and reliability of the OHAT as used by SPs.

Feasibility

This study demonstrates that the OHAT is a feasible instrument and is quick and simple to administer (with no need for special equipment). At the end of the preparatory part of the study, a few uncertainties in completing the tool were reported by the SPs, which were subsequently addressed and resolved. The short training period between the end of the preparatory study and baseline was sufficient to increase the self-confidence among the SPs in completing the tool. During the actual study, the SPs reported that they were able to complete the assessment. Thus, the self-perceived ability to complete all OHAT categories did not change significantly during the actual study, although differences occurred in the need for relying on the visual training resources. However, the initial need of the SPs to consult the manual showed a progressive decrease after 14 weeks of practice with the OHAT. As a result of methodological differences, it was not possible to compare the questionnaire results from our study with those of the study by Chalmers et al. [90].

Reliability

The inter-rater reliability and test-retest reliability of the total OHAT score were adequate for this study. In evaluating the individual categories of the OHAT, a high level of inter-rater reliability was achieved with almost perfect agreement in seven of the eight individual categories and the need for referral, as well as with perfect agreement for 'dental pain'. Although the reliability was good for this study, only three raters were involved. Future studies should evaluate whether the reliability data are different when applying the OHAT to a larger sample of SPs. The inter-rater reliability data were higher in this study than in the previous study by Chalmers et al. [90], in which different types of nurses were involved as raters [90]. Differences in educational background, the availability of visual training resources that could easily be accessed and the manual could be possible explanations for the higher reliability between the raters in this study. Chalmers et al. [90] did not use the Fleiss Kappa to evaluate the inter-rater reliability for all subjects because the inter-rater agreement was only assessed for two raters. In this study, three raters were involved, which increases the likelihood that the measurements obtained at the same time by the raters represent the subjects' true oral health status. However, this study was limited by not examining the concurrent validity of the tool as determined by comparing the OHAT results obtained from SPs with a dental examination completed by a qualified dentist. To use the OHAT optimally among SPs and to

establish the high sensitivity (= the presence) and high specificity (= the absence of the target condition) [144] of the tool, concurrent validity warrants further assessment, which will be conducted in the next phase of this research. Regarding the test-retest reliability, excellent agreement was only reached for 'natural teeth' and 'dentures'. The test-retest results showed lower levels of agreement for 'gums and tissues', 'lips', 'dental pain', and the overall referral decision. The lower levels of agreement may be explained by possible changes in oral health [147, 148]. In particular, spontaneous changes in oral health within 2 weeks cannot be excluded [147, 148]. In this study, the participating subjects did not receive any recommendations within the 14-week period to improve their oral health care to minimise alterations in the test conditions. However, the subjects did not alter their oral care habits during the actual study, and standard oral care was performed if present. The study was limited by not examining the type of oral care practices the subjects received or applied themselves. The test-retest results from this study could not be compared with the study results of Chalmers et al. [90] because of the different methodologies. Additionally, the intra-rater reliability was assessed using videotapes. Despite a high ICC for both raters for the total OHAT score and a high percentage of agreement in the individual categories, caution is required in interpreting the results for the individual categories. In fact, it was not possible to calculate the Fleiss Kappa because of limited variance in the scores. Moreover, the use of very small samples (only 10 videotapes) may yield misleading results from a proportion-based Kappa-statistic [144]. However, videotapes offer the advantage of allowing the assessment of identical aspects of oral health status and reducing stress that influences clinical presentation [149]. Videos also provide dynamic images and have been applied as a medium for dental health education in previous reports [150, 151].

Scoring and interpretation of the OHAT

Regarding the distribution of the subjects' total OHAT scores, a score of 5 was most frequently obtained. This total score may indicate the severity of the oral health status. However, each item should be considered separately because referral could be determined based on a single aberrant category. Therefore, the clinical significance of the total score should be questioned, as this factor has not been evaluated in previous studies. Further investigation should focus on the correlation between the total score and differences in the severity of oral health status as well as the need for referral. The mean total OHAT score was higher in this study than previously reported mean total OHAT scores [90], possibly

because the three SPs judged oral health status more strictly than did other providers. In particular, the categories ‘dentures’ and ‘natural teeth’ were scored as ‘unhealthy’ for the majority of the subjects in this study, whereas a larger proportion of the subjects in the study by Chalmers et al. [90] scored ‘healthy’ for the same categories. This discrepancy in scoring necessitates further validation of the OHAT tool. Therefore, further research should focus on the accuracy of the OHAT tool when administered by SPs and nurses compared with a dental examination by a qualified dentist.

Type of residence and oral health

The score distribution for individual categories across the places of residence revealed that hospitalised patients had worse dental status (dentures and natural teeth) and more dry tissues compared with other subjects. Although there were no statistically significant differences between the residential care settings for the categories ‘dentures’ and ‘natural teeth’, the finding that dental status was worst in the hospitalised patients was corroborated by Pajukoski et al. [141]. This finding may be explained by differences in concomitant diseases and polymedication, rather than by the nature of the patient’s illness [141]. However, we did not perform an investigation of possible underlying conditions and etiological factors. Consequently, an irrefutable explanation is also lacking for the finding that a significant difference was found for ‘saliva’ between the hospitalised patients and the subjects residing in nursing homes. Oral cleanliness was worse in nursing homes, which could be attributed to the greater cognitive impairment of the inhabitants, resulting in difficulty performing oral hygiene. Behavioural difficulties associated with dementia, such as refusal to open the mouth, are seen as especially challenging tasks for oral care providers [97]. Additionally, greater accumulations of dental plaque and calculus have been found on natural teeth and dentures in patients with dementia [97]. Our study was limited by the lack of a standardised cognitive assessment battery in all subjects to evaluate their cognitive abilities. However, the presence of cognitive impairment was obvious in the hospitalised patients and the nursing home residents, as it was determined based on medical diagnoses. Caution is needed when interpreting the results of the subjects from the assisted living facilities, as the presence of cognitive impairment lacks a true medical diagnosis. In the assisted living facilities, where most of the ‘more independent’ subjects provide oral care themselves and had lower rates of cognitive impairment, oral cleanliness was better. However, approximately half of the elderly living in those assisted living facilities needed referral to an oral health

professional because of the condition of their natural teeth or dentures. The differences in the nature of the oral categories needing intervention may ultimately lead to establishing oral health care intervention programmes. However, replication of this study with equally balanced groups is recommended. Based on the quantitative and descriptive interpretations of this study, we suggest the following adjustments to the OHAT to improve the effectiveness of this assessment tool. A description of the number of pairs of teeth in the chewing position, as in the original BOHSE [98, 100], may provide additional information on mastication and food selection. Uncertainty in evaluating oral cleanliness could be resolved by incorporating additional illustrations in the publicly available visual training resources. However, the inter-rater reliability of this category was good. The category 'dental pain' assesses pain not only as a consequence of dental problems but also as a result of ulcers anywhere in the mouth. Therefore, assigning a different name to that category could be considered, further obviating the need for constant referral to the manual or available visual training resources of the OHAT. Future studies should evaluate whether the systematic implementation of an oral health assessment tool, such as the OHAT, during a swallowing examination would promote oral and dental care.

2.5 CONCLUSION

The findings of this study show that the OHAT is a feasible and reliable oral health assessment tool that can be used in clinical practice by SPs to screen oral health in a standardised manner in elderly dysphagic subjects. Future research is necessary to evaluate whether the implementation of an oral health assessment tool within a swallowing examination can promote oral health care awareness in daily clinical practice. We suggest exploring whether an oral health assessment tool used by SPs could have the potential to improve oral care management in dysphagic subjects.

Appendix 2.1 An oral health assessment tool for dental screening (developed by Chalmers et al. [90], and modified by the Halton Region's Health Department [103, 104]).

Resident:_____Completed by: _____ Date: _/_/_

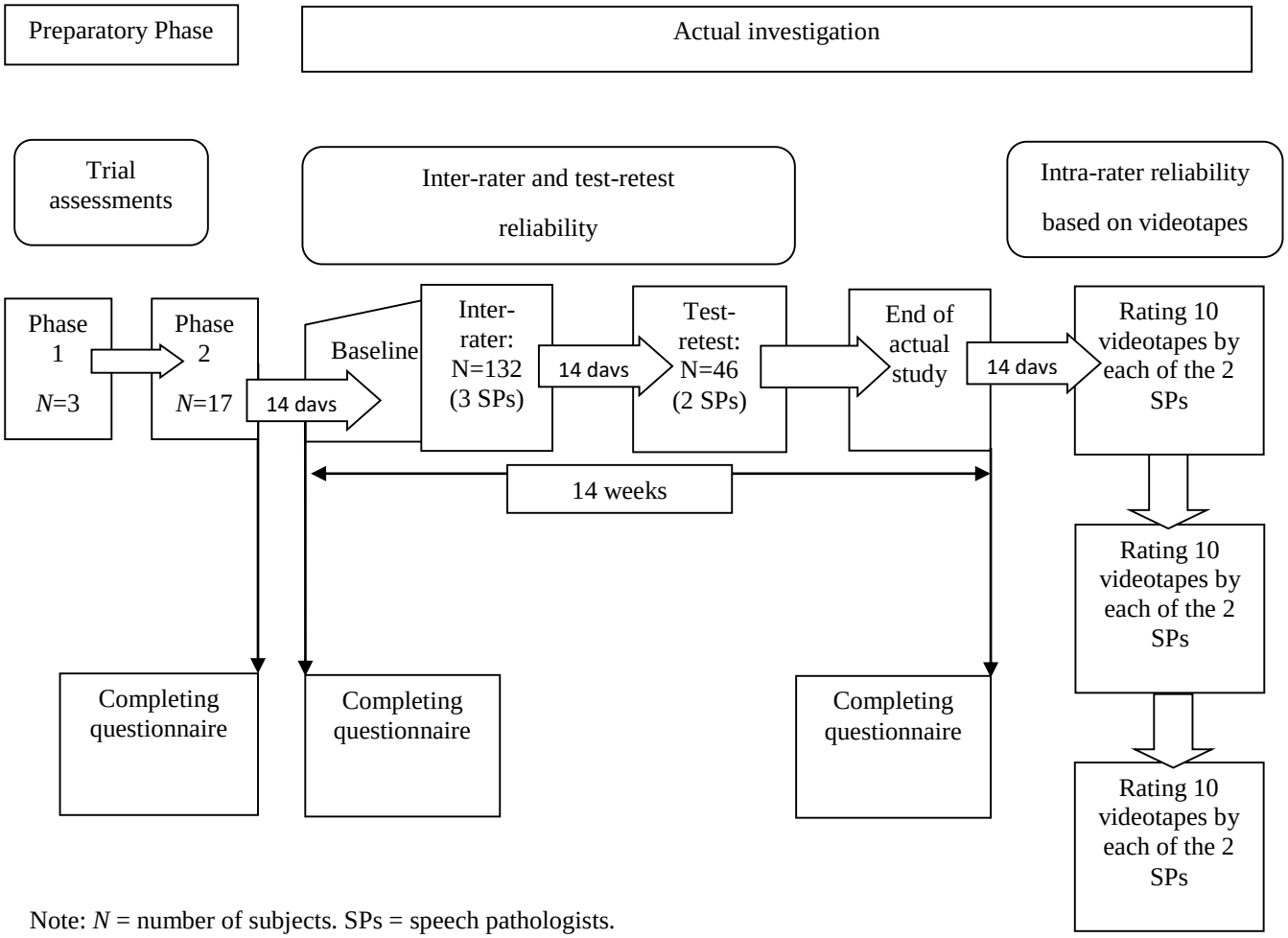
Category	0 = healthy	1 = changes	2 = unhealthy	Category scores
Lips	Smooth, pink, moist	Dry, chapped, or red at corners	Swelling or lump, white/red/ulcerated patch; bleeding/ulcerated at corners*	
Tongue	Normal, moist, roughness, pink	Patchy, fissured, red, coated	Patch that is red and/or white, ulcerated, swollen*	
Gums and tissues	Pink, moist, smooth, no bleeding	Dry, shiny, rough, red, swollen around 1–6 teeth, one ulcer/sore spot under dentures*	Swollen, bleeding around seven teeth or more, ulcers, white/red patches, generalized redness under dentures*	
Saliva	Moist tissues, watery and free-flowing saliva	Dry, sticky tissues, little saliva present, resident thinks they have dry mouth	Tissues parched and red, very little or no saliva present; saliva is thick, rope-like, resident complains of dry mouth*	
Natural teeth Yes/No	No decayed or broken teeth/roots	1–3 decayed or broken teeth/roots*	Four or more decayed or broken teeth/roots, or very worn down teeth, or less than four teeth with no dentures*	
Denture(s) Yes/No	No broken areas or teeth, dentures regularly worn, and named	1 broken area/tooth or dentures only worn for 1–2 hours daily, or dentures not named, or loose	More than 1 broken area/tooth, dentures missing or not worn due to poor fit, or worn only with denture adhesive*	
Oral cleanliness	Clean and no food particles or tartar in mouth or dentures	Food particles/tartar/plaque in 1 or 2 areas of the mouth or on small area of dentures or halitosis (bad breath)	Food particles/tartar/plaque in most areas of the mouth or on most of denture(s) or severe halitosis (bad breath)*	
Dental Pain	No behavioural, verbal or physical signs of pain	Verbal and/or behavioural signs of pain, such as pulling at face, chewing lips, not eating, aggression*	Physical signs of pain (swelling of cheek or gum, broken teeth, ulcers), as well as verbal and/or behavioural signs (pulling at face, not eating, aggression)*	

Referral to an oral health professional: □ yes □no

Total score: /16

An asterisk * and underline indicates referral to an oral health professional (i.e., dentist, dental hygienist, denturist) is required.

Appendix 2.2 Flow chart to define the different steps of the study



Note: *N* = number of subjects. SPs = speech pathologists.

Appendix 2.3 Questionnaire to evaluate the feasibility of the OHAT

Based on the original questionnaire, developed by Chalmers et al. [90]

1. Statements (Instruction: Please rate how strongly you agree or disagree with each of the formulated statements. Choose one of the following options: 'strongly disagree', 'disagree', 'agree' or 'strongly agree')
   a. I feel knowledgeable and prepared to use the OHAT.
   b. Using the OHAT improves my ability to detect dental pain and problems in residents' mouths.
   c. I had enough time to learn about the OHAT before it was implemented.
   d. I am able to complete the 'lips' category of the OHAT.
   e. I am able to complete the 'tongue' category of the OHAT.
   f. I am able to complete the 'gums and tissues' category of the OHAT.
   g. I am able to complete the 'saliva' category of the OHAT.
   h. I am able to complete the 'natural teeth' category of the OHAT.
   i. I am able to complete the 'dentures' category of the OHAT.
   j. I am able to complete the 'oral cleanliness' category of the OHAT.
   k. I am able to complete the 'dental pain' category of the OHAT.
2. Open-ended questions
   a. Do you experience difficulties when applying the Oral Health Assessment Tool?
      Yes/No, explanation:
      Do you think the intended section to refer to a dental professional is a necessary part of the oral health assessment tool?
      Yes/No, explanation:
   b. Do you have sufficient visual resources to perform the OHAT?
      Yes/No, explanation:
   c. Do you need the manual to perform the OHAT?
      Yes/No, explanation:

CHAPTER III

THE ROLE OF OROPHARYNGEAL DYSPHAGIA IN PREDICTING PLACE OF RESIDENCE AND FUNCTIONAL OUTCOMES AFTER STROKE: A ONE-YEAR PROSPECTIVE FOLLOW-UP STUDY

Study III is in review

ABSTRACT

Predicting outcomes after stroke is important for planning patient care and support and for gaining a better understanding of the disease course. Post-stroke oropharyngeal dysphagia (PSOD) is a common complication after stroke that may influence patient outcomes. This study determined whether the presence of PSOD during the initial swallowing examination post-stroke onset could predict place of residence post-discharge. The secondary aims were to evaluate whether the Mann Assessment of Swallowing Ability (MASA) and/or the Oral Health Assessment Tool (OHAT) could predict short- and long-term patient outcomes (i.e., pneumonia, feeding status, functional independence, residence and survival). We conducted a multi-centre prospective study with a one-year follow-up in 151 acute stroke patients. Dysphagia was present in 76.2% of patients at baseline. The odds ratio (OR) for not returning home after discharge was higher among PSOD patients (OR: 2.82, 95% confidence interval (CI) [1.02, 7.81]). The baseline MASA and OHAT scores were both significant predictors of pneumonia during hospitalization (OR MASA: 0.96, $p < 0.0001$; OR OHAT: 1.23, $p = 0.046$) and follow-up (OR MASA: 0.97, $p = 0.0002$; OR OHAT: 1.26, $p = 0.017$). The baseline MASA score was a significant predictor of feeding status, functional independence and place of residence during follow-up, although more information regarding feeding status and functional independence during the subsequent year could be obtained by repeating the MASA at different time points. Severe PSOD, as assessed by the MASA, was associated with increased mortality. This study confirmed that PSOD patients experience poor outcomes.

Keywords: Deglutition, Deglutition disorders, Oropharyngeal dysphagia, Outcome assessment, Stroke

3.1 INTRODUCTION

Stroke has been recognized worldwide as a leading cause of disability and death [1, 2]. Patients and families, as well as health care and insurance providers, often ask clinicians for a reasonable outcome prediction (e.g., if discharge to home is possible) very soon after stroke onset [11, 14]. Therefore, knowledge of the key factors that affect the outcomes is important for predicting stroke prognoses [11]. The factors that have improved outcomes after stroke include the increased number of dedicated stroke units that have been developed during previous decades [11, 12] and the current guidelines for stroke prevention [11, 13]. The factors that can negatively affect outcomes after stroke include increasing age, stroke characteristics (i.e., lesion location, clinical features and stroke severity) and related complications [11, 14, 15]. Post-stroke oropharyngeal dysphagia (PSOD), which is defined as the presence of abnormal swallowing physiology in the upper aerodigestive tract [16], is a common complication after stroke [1]. The presence of dysphagia is associated with poor outcomes, such as institutionalization [15], dehydration [19], malnutrition [2, 19], pneumonia [2, 20], decreased functional independence [15], increased mortality and morbidity [2, 10], increased length of hospital stay (LOS) [17], and increased health care costs [21]. Although studies have investigated the outcomes of PSOD during hospitalization and post- discharge [12, 14, 105-107], and during follow-up [1, 12, 15, 19, 108-110], the best practices for the management (including diagnosis and investigation) of PSOD remain undefined [2]. Furthermore, the reported incidence of PSOD varies greatly (i.e., ranges from 14% to 94% [17]), which can be explained by variations in the assessment method, the timing of the initial evaluation after stroke, and the lesion location [18]. Consequently, the reported short- and long-term outcomes associated with PSOD are often not consistent among studies [18].

Pneumonia is the most feared complication following stroke [20]. Post-stroke pneumonia is mostly caused by dysphagia, leading to aspiration of colonized oropharyngeal material (i.e., liquids, foods, and oral secretions) [2, 20, 22, 23]. Depending on the operational definition of pneumonia and the timing of the pneumonia diagnosis, the reported incidence of pneumonia in acute stroke patients with dysphagia ranges from 16.2% to 33% [18]. Furthermore, these patients have a 3.17-fold higher risk of developing pneumonia than stroke patients without dysphagia [18]. Because post-stroke pneumonia is a multifactorial phenomenon [20], tube feeding and a poor oral health status are also considered contributors to the development of this disease [20, 17]. Tube-fed patients show a higher rate of pneumonia than patients who receive food or liquid by mouth [22]. Poor oral health status increases the

risk of pneumonia, particularly among patients who have dysphagia and those receiving enteral feeding [2]. Pneumonia prevention should start as soon as patients are admitted to the hospital, and preventive measures include the early identification of dysphagia and the screening of patients for oral health [2]. Many bedside screening tests for evaluating swallowing capability are available [111]. However, there is no consensus regarding the preferred screening protocol, and few of these tests have sufficient psychometric properties [87]. Furthermore, there is limited evidence regarding the benefits of using dysphagia screening tools on health outcomes [16]. A method that is different from these swallowing screens is the use of a standardized clinical swallowing examination (CSE), which is performed by a speech-language pathologist (SLP). This clinical bedside swallowing assessment enables a thorough examination of the oral, pharyngeal, and laryngeal anatomy and physiology, establishes a foundation for the construction of a treatment plan, and allows the determination of the appropriate diet for each patient [114, 115]. The Mann Assessment of Swallowing Ability (MASA) is a CSE that has been specifically validated for assessing PSOD [87-89]. The MASA has been compared with videofluoroscopy [88, 89] and Fiberoptic Endoscopic Evaluation of Swallowing (FEES) [113] and demonstrates good psychometric properties [88, 89, 113]. An oral health screening can complement the swallowing examination [117]. Speech-language pathologists pay particular attention to oral health and dentition [95], both of which are essential for swallowing and speech production [31]. The feasibility and reliability of the Oral Health Assessment Tool (OHAT) in screening oral health in a standardized manner by SLPs was recently established [117]. The Functional Oral Intake Scale (FOIS), which is a stroke-validated tool for documenting changes in functional eating abilities over time, can be used to outline the functional impact of PSOD on the oral intake of stroke patients [152].

In addition to the variations in the methods used to identify dysphagia, the manner in which long-term outcomes are assessed can also influence outcome predictions. Most studies have used retrospective methods [108-110], chart reviews [108-110] and/or interview methods [15, 19, 109] to document the outcomes of PSOD and/or feeding status. Furthermore, the interval between successive measurements was often too long [15, 116]. Consistency in assessing swallowing and evaluating functional independence at regular intervals during follow-up [15] enables better descriptions of and predictions regarding who will have a persistent swallowing disorder, who will require a modified oral intake, who will have decreased functional independence, and who is at risk of long-term institutionalization. Furthermore, these improvements will help patients and their families to better

understand the course of the disease [11]. Therefore, longitudinal datasets with comprehensive measurements performed at different time points are required to build models that allow accurate predictions of mortality and short- and long-term outcomes [116] to enable the planning of future care and support [15].

The primary purpose of this study was to determine whether the presence of PSOD at baseline (i.e., during the first swallowing examination after the onset of stroke) could predict place of residence after discharge. We hypothesized that the presence of PSOD at baseline was associated with institutionalization after discharge or death. A secondary aim of this study was to evaluate whether baseline MASA and OHAT scores could predict the occurrence of pneumonia during hospitalization and follow-up. A third purpose of this study was to examine whether the baseline MASA score could predict patient outcomes in terms of feeding status, functional independence and place of residence during a one-year follow-up (i.e., at 1 month, 3 months, 6 months and 1 year). Because the MASA was performed at different time points during the subsequent year, we also investigated whether the relationships among the MASA score and feeding status, functional independence and place of residence changed over time by studying the associations among these factors at each time point. A fifth aim of the study was to investigate survival as a function of baseline MASA and/or OHAT scores while correcting for the other variables.

3.2 MATERIALS AND METHODS

3.2.1 Participants

We conducted a multi-centre prospective longitudinal study with a one-year follow-up in a cohort of 151 acute stroke patients who were referred to a SLP for a swallowing examination. Patients were included based on a clinical diagnosis of acute ischaemic or haemorrhagic cerebral stroke, which was confirmed by a neurological examination and neuroimaging exams, such as computed tomography (CT) and/or magnetic resonance imaging (MRI). Patients with a transient ischaemic attack (TIA) and/or epidural, subdural or subarachnoid haemorrhage were excluded. The following patients were also excluded: (1) patients with aggressive behavioural disorders that led to the refusal of a swallowing assessment, (2) patients with an inconclusive imaging exam, and (3) patients with trauma, anatomical alterations in the head and neck region or premorbid head and neck cancer that could cause mechanical dysphagia. Patients with a history of stroke were not excluded. Ethical approval for the consent procedure and the

experimental protocol of the study was granted by two independent ethics boards, namely, the Committee for Medical Ethics of the University Hospital of Antwerp and the Ethics Committee of the AZ Delta Hospital Roeselare-Menen (B300201215080). The study was conducted in full accordance with the World Medical Association Declaration of Helsinki. Verbal and written consent was obtained from all subjects or a legal representative if the patient could not provide consent.

3.2.2 Procedures

The study encompassed a screening of oral health, a thorough CSE, and an evaluation of the feeding status and functional independence at the following time points: (1) baseline, (2) 1 month post-stroke, (3) 3 months post-stroke, (4) 6 months post-stroke, and (5) 1 year post-stroke. The flow chart presented in Appendix 3.1 illustrates the different stages of the study.

Assessment during the acute phase and during hospitalization

Prior to the CSE, the oral health of the patients was evaluated by a SLP using the OHAT, which consists of eight categories ('lips', 'tongue', 'gums and tissues', 'saliva', 'natural teeth', 'dentures', 'oral cleanliness', and 'dental pain') with three possible scores ('0: healthy', '1: some changes present' and '2: unhealthy condition') [90, 117]. The total score, which encompasses the sum of all category scores, was used for the data analysis. Following the OHAT, a SLP performed the MASA on each patient, and all participating SLPs had extensive experience using both assessments. The proposed criteria of the MASA involved the classification of patients according to the presence and severity of dysphagia (i.e., ≥ 178 = absence of dysphagia, 168-177 = mild dysphagia, 139-167 = moderate dysphagia, and ≤ 138 = severe dysphasia) [88]. Boluses of liquid, jelly and, if possible, biscuits were used to evaluate the swallowing function. The presence of dysphagia was confirmed by the FEES [153]. Based on both the clinical and instrumental swallowing investigations, the type of diet was determined for each patient and scored according to the FOIS [152]. In addition to groups created based on the MASA classification (i.e., the presence and severity of dysphagia), additional subgroups were also created according to the type of diet at baseline, similar to other studies [14, 106]. Patients with an FOIS score of 6 or 7 were placed in the regular diet group, those with an FOIS score of 4 or 5 received a dysphagia diet, and those with an FOIS

score of 1 to 3 constituted the tube-fed group. Notably, not all patients with severe dysphagia obtained an FOIS score ≤ 3. In cases in which the attending doctor or the patient decided to not rely on tube feeding for the nutrition supply, the patient received a dysphagia diet and was assigned an FOIS score of 4 or 5. The Barthel index was administered by a nurse/occupational therapist to assess the functional independence of each patient [154]. The total Barthel scores ranged from 0 to 20, with lower scores indicating increased disability or greater functional dependence [155]. Demographic data, including data regarding premorbid conditions (e.g., living conditions, pre-stroke functional independence and pre-stroke functional oral intake), stroke characteristics and stroke severity, were recorded either through a review of patient medical records or by interviewing the patient or his/her legal representative. During their hospital stays, all patients received oral care for the prevention of pneumonia, and all dysphagic patients were supervised for swallowing by a SLP until they were discharged or their dysphagia resolved. The LOS and place of residence after discharge (i.e., home, other hospital (e.g., acute or rehabilitation hospital), or nursing home) were recorded. Pneumonia was diagnosed by the attending medical doctor and defined according to the Mann criteria [156]. For practical reasons, we did not repeat the evaluation of the swallowing function and the type of diet at the time of discharge unless the patient was discharged 1 month post-stroke.

Assessment during follow-up

During the follow-up, the FOIS score, Barthel index, place of residence, and occurrence of pneumonia and death were assessed, as were the MASA and OHAT scores. The FEES was repeated only in patients for whom oral feeding was reconsidered. The occurrence of pneumonia or recurrent stroke, as diagnosed by the attending medical doctors, was documented during the one-year follow-up.

3.2.3 Data analysis

The sample size was calculated using PASS 11 (Hintze J, NCSS, LLC, Kaysville, Utah, USA, https://www.ncss.com). All statistical analyses were performed using SPSS 20.0 (IBM, SPSS Inc., Chicago, IL, USA) and SAS 9.4 (SAS Institute Inc., Cary, NC, USA), and $p < 0.05$ was considered significant. The figures were created using R 3.3.2 (R Core Team, R: A language and environment for

statistical computing. R Foundation for statistical Computing, Vienna, Austria, https://www.R-project.org). The primary outcome was the place of residence after discharge (defined as 'home' versus 'not home'). The sample size calculation was based on a reported prevalence estimate for PSOD among stroke victims of 40%, with 60% of post-stroke patients not experiencing dysphagia [87, 157-159]. A logistic regression of the binary response variable (i.e., place of residence after discharge, if different from home) versus the binary independent variable (i.e., the presence of dysphagia) achieved 80% power (to detect a change in the probability of 'not returning home' from 66% to 88.6% by comparing the group without PSOD to the group with PSOD) with a sample size of 225 observations at a significance level of 0.05 [14]. The abovementioned change corresponds to an odds ratio (OR) of 4.00 [12, 15]. Adjustments were performed for the R-squared of 0.500 for other possible independent variables (i.e., age and pneumonia during hospitalization) in the logistic regression. The demographic data, stroke characteristics (including stroke severity) and clinical assessment results were tabulated according to the presence and severity of dysphagia and the type of diet at baseline, and the latter data were used only in the evaluations of the discharge destination (i.e., home, nursing home, other hospital, and death) and LOS. Normality was assessed using the Shapiro-Wilk test and Q-Q plots. For continuous variables, the differences between the groups were assessed using the Kruskal-Wallis test, and two-by-two post hoc tests with a multiple testing correction were used to evaluate the groups that differed significantly. ANOVA with Tukey post hoc tests was used to analyse the OHAT data. For categorical variables, chi-square tests or Fisher's exact test was used.

To study the effect of dysphagia at baseline on the discharge destination, the relationship between the baseline diet type and the place of residence after discharge, and LOS, we studied only those patients without stroke expansion or recurrent stroke during hospitalization (N = 140). According to the results, there were significant differences in the premorbid place of residence between patients with and without PSOD at baseline and between the 'baseline type of diet' groups. Therefore, patients residing in a nursing home before the onset of stroke were excluded from these analyses, and 112 patients were included in these analyses. The Mann-Whitney U test was used to compare LOS between patients with and without PSOD, and the Kruskal-Wallis test was used to compare LOS between groups based on the type of diet at baseline. For the total stroke population in our study (N = 151), Spearman's rank correlation coefficient (Spearman's rho) was used to investigate the correlations.

Logistic regression was performed to predict the occurrence of pneumonia during hospitalization as a function of the MASA score, OHAT score, and age at baseline, and patients with expanding/recurrent stroke were excluded from this analysis (N = 140). Receiver operating characteristic (ROC) curves and area under the curve (AUC) calculations enabled the evaluation of the diagnostic accuracy of baseline MASA and/or OHAT scores in predicting pneumonia during hospitalization. We also examined the usefulness of baseline MASA and OHAT scores as predictors of pneumonia during the one-year follow-up. Therefore, we used a generalized linear mixed model that included the baseline MASA score, baseline OHAT score, time and all two-way interactions between these terms as fixed effects and the subjects as a random effect. Time was considered a continuous variable since the small sample size did not allow time to be considered a categorical variable.

The value of the baseline MASA score as a predictor of outcomes over time (i.e., feeding status, functional independence and place of residence categorized as 'home versus institution' or more specifically, 'home as residence') was modelled using a linear mixed-effects model for continuous outcomes or a generalized linear mixed-effects model for dichotomized outcomes. For all models, we considered the MASA score, time and the interaction between MASA score and time to be fixed effects, while the subjects were considered a random effect. Time was considered categorical (i.e., 1 month, 3 months, 6 months, and 1 year) in all analyses. Data related to patients with stroke expansion/recurrence during the one-year follow-up were considered missing from the moment of expansion/recurrence onwards in all analyses in which the baseline MASA score was considered a predictor variable. The combined predictive value of the MASA score at baseline and at each follow-up time point for feeding status and functional independence at the corresponding time point was studied using linear regression. Analogously, logistic regression was fitted for place of residence at each time point. Patients who lived in a nursing home prior to the onset of stroke were withdrawn from all analyses in which place of residence was an outcome variable.

Survival after stroke among the different dysphagia severity categories was studied using Kaplan-Meier curves, and the time to death was modelled by a Cox proportional hazard (PH) model. All patients who experienced stroke expansion/recurrence during the one-year follow-up were withdrawn from the Kaplan-Meier analysis, and the group composition was based on the presence and severity of dysphagia at baseline. The Kaplan-Meier analysis was performed using data from 135 subjects. We explored different Cox-regression models using the following variables: the MASA score, the OHAT

score, age, the occurrence of pneumonia within one year post-stroke and the presence of stroke expansion/stroke recurrence within one year post-stroke. Unlike the Kaplan-Meier analysis, the Cox-regression analysis included all patients since a correction for stroke expansion/stroke recurrence during the one-year follow-up was applied.

3.3 RESULTS

3.3.1 Patient Characteristics

Overall, 118 (78.1%) patients suffered from ischaemic stroke, 15 (9.9%) patients suffered from haemorrhagic stroke, and 18 (11.9%) patients experienced both. Dysphagia was diagnosed in 115 patients at baseline (76.2%). Thirty-six (23.8%) patients did not exhibit PSOD. The mild dysphagia group consisted of 18 patients (11.9%). Moderate PSOD was present in 46 (30.5%) patients, and severe dysphagia affected 51 (33.8%) patients. The group fed a regular diet at baseline consisted of 40 (26.5%) patients, the group fed a dysphagia diet included 56 (37.1%) patients, and the group that received tube feeding comprised 55 (36.4%) patients. Table 3.1 presents the demographic data, stroke characteristics, stroke severity, and clinical assessment scores according to the presence and severity of dysphagia and the type of diet at baseline. Overall, pneumonia occurred in 30.5% of patients (N = 151) during hospitalization, and the incidence of pneumonia between the onset of stroke and 1 year post-stroke was 31.8%. Of the 151 patients, 16 (10.6%) patients showed expansion of their stroke during hospitalization or experienced stroke recurrence during follow-up.

3.3.2 Oropharyngeal dysphagia and type of diet at baseline and discharge destination after hospitalization

A moderate negative correlation ($r_s = -0.49$, $p < 0.001$) was observed between the MASA and the OHAT scores at baseline, and a very strong correlation ($r_s = 0.93$, $p < 0.001$) was observed between the MASA and the FOIS scores at baseline. A chi-square test with a continuity correction did not identify significant differences ($p = 0.08$) in the premorbid place of residence between patients with and without PSOD. Of the 151 patients, 11 (7.3%) patients suffered from stroke expansion/recurrence during hospitalization. Regarding the remaining 140 patients, a chi-square test with a continuity correction did identify

significant differences (p = 0.041) in premorbid place of residence between patients with and without PSOD. Similarly, significant differences (p = 0.013) in premorbid place of residence were found between the groups based on the type of diet at baseline. Of the 112 patients who did not reside in a nursing home before the onset of stroke, dysphagia was diagnosed in 72.3% of the cases. In the PSOD subgroup, 71.6% of patients could not return home. Of those dysphagic patients who did not return home, 24.1% died during hospitalization. The patients without PSOD could return home in 74.2% of cases. No individuals in this group died during hospitalization. A binary logistic regression model showed that the presence of dysphagia (p = 0.042) and the presence of pneumonia (p < 0.001) were significant predictors of place of residence after discharge, while age was not a significant predictor of place of residence after discharge (p = 0.154). Using this model, we were able to correctly predict the discharge destination in 69.6% of cases. The OR for not returning home in patients with PSOD was 2.82 (95% confidence interval (CI) [1.02, 7.81]). The mean LOS was significantly different between the two groups (p < 0.001). The subgroup without PSOD had a mean LOS of 15.74 days (SD = 9.57; min, max: 6, 47), while the subgroup with PSOD had a mean LOS of 29.12 days (SD = 15.67; min, max: 6, 101). Figure 3.1 shows the place of residence after discharge according to the type of diet at baseline. Significant differences in place of residence were found among these three groups (p < 0.001). Death was most prevalent in the group receiving tube feeding (i.e., 25.0% compared to 12.2% in the dysphagia diet group). More patients were transferred to another hospital or a nursing home in the groups receiving the dysphagia diet (34.1% were transferred to another hospital and 9.8% were transferred to a nursing home) or tube feeding (47.2% and 19.4%, respectively) than in those fed a regular diet (22.9% and 5.7%, respectively). The mean LOS was 17.17 days (SD: 10.05; min, max: 6, 47) in the regular diet group (N = 35), 24.00 days (SD: 10.85; min, max: 6, 62) in the dysphagia diet group (N = 41), and 35.06 days (SD: 18.85; min, max: 3, 101) in the tube feeding group (N = 36). The Kruskal-Wallis (p < 0.001) and post hoc tests revealed that LOS differed significantly among all the groups (‘regular diet versus dysphagia diet’ (p = 0.029), ‘regular diet versus tube feeding’ (p < 0.001), and ‘dysphagia diet versus tube feeding’ (p = 0.027)).

Table 3.1 Demographic data, stroke characteristics, stroke severity and clinical assessment results in the two groups

Characteristics	No dysphagia (N = 36)	Mild dysphagia (N = 18)	Moderate dysphagia (N = 46)	Severe dysphagia (N = 51)	*p*-value	Regular diet (N = 40)	Dysphagia diet (N = 56)	Tube feeding (N = 55)	*p*-value
Background characteristics									
Age (years)									
Median (IQR)[a] (Min, Max)	75.00 (21.00) (44, 88)	81.50 (17.50) (47, 90)	82.00 (12.25) (60, 95)	83.00 (12.00) (52, 94)	0.005[b]	74.00 (21.75) (44, 88)	82.50 (12.75) (58, 95)	83.00 (11.00) (52, 94)	<0.001[c]
Gender, N (%)					0.008[d]				0.072[d]
Male	21 (58.30)	10 (55.60)	11 (23.90)	24 (47.10)		23 (57.50)	19 (33.90)	24 (43.60)	
Female	15 (41.70)	8 (44.40)	35 (76.10)	27 (52.90)		17 (42.50)	37 (66.10)	31 (56.40)	
Barthel index premorbid					0.001[e]				0.001[f]
Median (IQR)	20.00 (1.00)	20.00 (1.25)	19.00 (7.25)	18.00 (7.00)		20.00 (1.00)	19.00 (3.75)	18.00 (7.00)	
(Min, Max)	(3, 20)	(13, 20)	(3, 20)	(0, 20)		(3, 20)	(3, 20)	(0, 20)	
FOIS premorbid					0.195[g]				0.659[g]
Median (IQR)	7.00 (0.00)	7.00 (0.25)	7.00 (0.00)	7.00 (0.00)		7.00 (0.00)	7.00 (0.00)	7.00 (0.00)	
(Min, Max)	(6, 7)	(6, 7)	(5, 7)	(5, 7)		(6, 7)	(6, 7)	(5, 7)	
Place of residence premorbid					0.075[d]				0.017[d]
Home, N (%)	33 (91.70)	16 (88.90)	36 (78.30)	36 (70.60)		37 (92.50)	46 (82.10)	38 (69.10)	
Nursing home, N (%)	3 (8.30)	2 (11.10)	10 (21.70)	15 (29.40)		3 (7.50)	10 (17.90)	17 (30.90)	
Stroke characteristics									
First stroke, N (%)	31 (86.10)	16 (88.90)	40 (87.00)	39 (76.50)	0.427[d]	34 (85.00)	50 (89.30)	42 (76.40)	0.178[d]
Recurrent stroke, N (%)	5 (13.90)	2 (11.10)	6 (13.00)	12 (23.50)		6 (15.00)	6 (10.70)	13 (23.60)	
TIA premorbid, N (%)	2 (5.60)	1 (5.60)	5 (10.90)	2 (3.90)	0.566[h]	3 (7.50)	4 (7.10)	3 (5.50)	0.922[h]
Stroke subtype					0.059[h]				0.009[h]
Ischaemic, N (%)	32 (88.90)	16 (88.90)	33 (71.70)	37 (72.50)		36 (90.00)	44 (78.60)	38 (69.10)	
Haemorrhagic, N (%)	2 (5.60)	2 (11.10)	8 (17.40)	3 (5.90)		2 (5.00)	9 (16.10)	4 (7.30)	
Both, N (%)	2 (5.60)		5 (10.90)	11 (21.60)		2 (5.00)	3 (5.40)	13 (23.60)	
Location of lesions					0.312[h]				0.914[h]

Table 3.1 continued

Characteristics	No dysphagia (N = 36)	Mild dysphagia (N = 18)	Moderate dysphagia (N = 46)	Severe dysphagia (N = 51)	*p*-value	Regular diet (N = 40)	Dysphagia diet (N = 56)	Tube feeding (N = 55)	*p*-value
Supratentorial, N (%)	32 (88.90)	12 (66.70)	39 (84.80)	43 (84.30)		34 (85.00)	45 (80.40)	47 (85.50)	
Infratentorial, N (%)	3 (8.30)	4 (22.20)	5 (10.90)	3 (5.90)		4 (10.00)	7 (12.50)	4 (7.30)	
Both, N (%)	1 (2.80)	2 (11.10)	2 (4.30)	5 (9.80)		2 (5.00)	4 (7.10)	4 (7.30)	
Lesion side					0.391[d]				0.113[d]
Left, N (%)	20 (55.60)	8 (44.40)	20 (43.50)	23 (45.10)		21 (52.50)	29 (51.80)	21 (38.20)	
Right, N (%)	9 (25.00)	8 (44.40)	20 (43.50)	15 (29.40)		11 (27.50)	22 (39.30)	19 (34.50)	
Both, N (%)	7 (19.40)	2 (11.10)	6 (13.00)	13 (25.50)		8 (20.00)	5 (8.90)	15 (27.30)	
Stroke severity									
NIHSS, Median (IQR) (Min, Max)	4.00 (3.50)[i] (2, 9)	4.00 (4.50)[j] (2, 8)	12.00 (7.50)[k] (4, 19)	22.00 (8.50)[l] (11, 36)	<0.001[m]	4.00 (3.50)[n] (2, 9)	11.50 (9.50)[o] (2, 35)	21.00 (9.00)[p] (7, 36)	<0.001[q]
Clinical assessments									
Period of initial CSE from onset, Median days, (IQR) (Min, Max)	2.50 (2.00) (0, 7)	2.00 (3.00) (1, 5)	2.50 (3.00) (0, 18)	2.00 (6.00) (0, 34)	0.750[g]	3.00 (2.00) (0, 7)	2.00 (2.75) (0, 18)	2.00 (6.00) (1, 34)	0.211[g]
MASA, Median (IQR) (Min, Max)	190.00 (12.00) (178, 200)	170.50 (5.00) (168, 177)	151.00 (11.25) (139, 166)	106.00 (40.00) (2, 138)	<0.001[r]	188.50 (13.75) (166, 200)	154.00 (24.25) (90, 177)	115.00 (49.00) (2, 161)	<0.001[s]
FOIS, Median (IQR)	7.00 (1.00)	5.00 (0.00)	4.00 (1.25)	1.00 (1.00)	<0.001[r]	7.00 (1.00)	4.00 (1.00)	1.00 (1.00)	<0.001[s]
(Min, Max)	(6, 7)	(4, 6)	(1, 6)	(1, 4)		(6, 7)	(4, 5)	(1, 3)	
OHAT, Mean (SD)	2.94 (2.08)	4.50 (2.46)	5.65 (1.92)	6.27 (2.63)	<0.001[t]	3.05 (2.09)	5.16 (2.15)	6.49 (2.45)	<0.001[u]
(Min, Max)	(0, 7)	(0, 9)	(1, 10)	(0, 11)		(0, 7)	(0, 10)	(0, 11)	
Barthel index, Median (IQR) (Min, Max)	13.50 (12.75) (1, 20)	10.50 (6.25) (2, 20)	2.00 (8.25) (0, 18)	0.00 (0.00) (0, 6)	<0.001[v]	12.00 (12.00) (1, 20)	4.00 (9.00) (0, 20)	0.00 (1.00) (0, 8)	<0.001[s]

TIA, Transient Ischaemic Attack; NIHSS, National Institutes of Health Stroke Scale; CSE, clinical swallowing examination; MASA, Mann Assessment of Swallowing Ability; FOIS, Functional Oral Intake Scale; OHAT, Oral Health Assessment Tool

[a] IQR = Interquartile range

Table 3.1 continued

[b] Post hoc Kruskal-Wallis: $p = 0.024$ for 'no dysphagia-moderate dysphagia', $p = 0.004$ for 'no dysphagia-severe dysphagia'

[c] Post hoc Kruskal-Wallis: $p = 0.001$ for 'regular diet-dysphagia diet' and for 'no dysphagia-tube feeding'

[d] Chi-square

[e] Post hoc Kruskal-Wallis: $p = 0.003$ for 'no dysphagia-severe dysphagia'; $p = 0.029$ for 'severe-mild dysphagia'; $p = 0.034$ for 'no dysphagia-moderate dysphagia'

[f] Post hoc Kruskal-Wallis: $p = 0.001$ for 'regular diet-tube feeding'

[g] Kruskal-Wallis test

[h] Fisher's exact test

[i] N = 12; [j] N = 5; [k] N = 17; [l] N = 29

[m] Post hoc Kruskal-Wallis test: $p < 0.001$ for 'no dysphagia-severe dysphagia' and 'mild-severe dysphagia'; $p = 0.001$ for 'moderate-severe dysphagia'

[n] N = 12; [o] N = 22; [p] N = 29

[q] Post hoc Kruskal-Wallis test: $p < 0.001$ for 'regular diet-tube feeding'; $p = 0.028$ for 'regular diet-dysphagia diet'; $p = 0.002$ for 'dysphagia diet-tube feeding'

[r] Post hoc Kruskal-Wallis test: $p < 0.001$ for 'no dysphagia-severe dysphagia', 'mild-severe dysphagia', moderate-severe dysphagia' and 'no dysphagia-moderate dysphagia'

[s] Post hoc Kruskal-Wallis test: $p < 0.001$ for all pairwise comparisons

[t] Post hoc analysis of variance test (Tukey HSD test): $p < 0.001$ for 'no dysphagia-moderate dysphagia' and 'no dysphagia-severe dysphagia'; $p = 0.030$ for 'mild-severe dysphagia'

[u] Post hoc analysis of variance test (Tukey HSD test): $p < 0.001$ for all pairwise comparisons, except for 'enteral feeding-dysphagia diet': $p = 0.008$

[v] Post hoc Kruskal-Wallis test: $p < 0.001$ for 'mild - severe dysphagia', 'no dysphagia-moderate dysphagia' and 'no dysphagia-severe dysphagia'; $p = 0.001$ for 'moderate-severe dysphagia'; $p = 0.002$ for 'mild-moderate dysphagia'

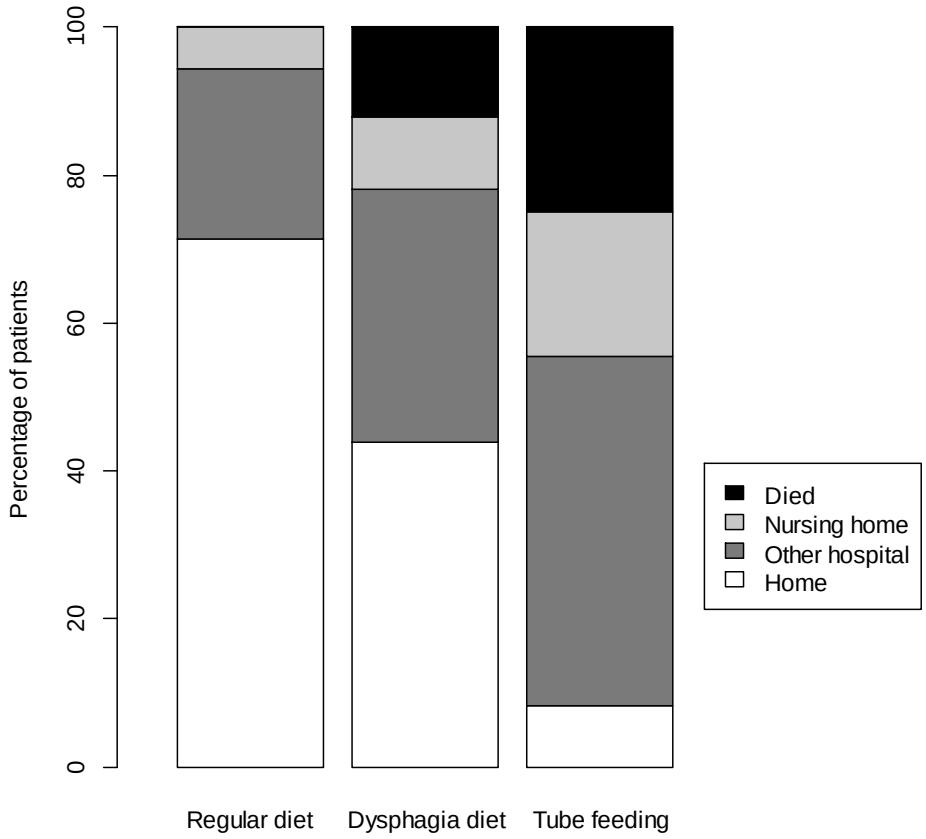


Figure 3.1 Place of residence after discharge from the hospital based on the type of diet at baseline (N = 112)

3.3.3 MASA and OHAT scores as predictors of pneumonia during hospitalization and follow-up

Age was omitted from the logistic regression model that was used to predict the occurrence of pneumonia during hospitalization as a function of the baseline MASA and/or OHAT scores because it was not a significant predictor. The MASA and OHAT scores were both significant predictors ($p < 0.0001$) when they were individually treated. After combining the MASA and OHAT scores into one model, we found that both the MASA ($p < 0.0001$) and OHAT ($p = 0.046$) scores remained significant predictors, while their interaction at baseline was not significant ($p = 0.487$). This model enabled correct predictions of the occurrence of pneumonia during hospitalization in 77.8% of cases. The OR for pneumonia during hospitalization was 0.96 (95% CI [0.94, 0.98] for the MASA score, indicating that for every 1-unit increase in the MASA score, the odds of pneumonia decreased by approximately 4%. A similar analysis of the OHAT scores showed that the OR for pneumonia during hospitalization was 1.23 (95% CI [0.997,

1.52]), indicating that the odds of pneumonia increased by 23% for every 1-unit increase in the OHAT score. The combined diagnostic accuracy of the MASA and OHAT scores for detecting pneumonia during hospitalization was represented by an AUC of 0.88 (Fig. 3.2), while the accuracy of the MASA score alone was represented by an AUC of 0.85.

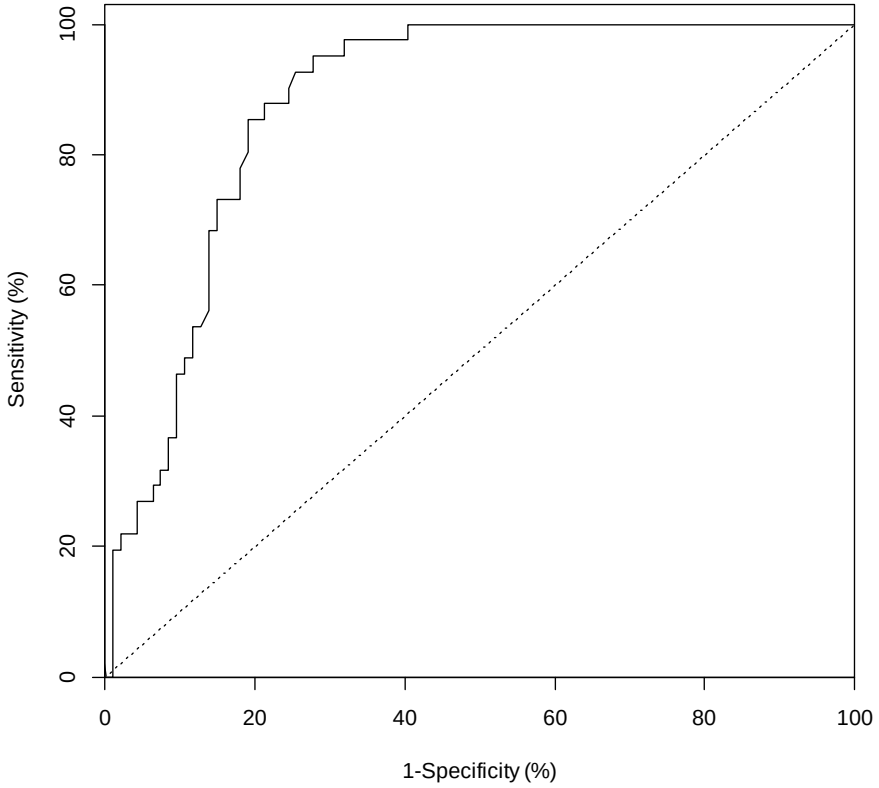


Figure 3.2 Receiver operating characteristic curve for the model for pneumonia with the baseline MASA score and the baseline OHAT score as predictors

Regarding the occurrence of pneumonia during the one-year follow-up, the MASA score at baseline, the OHAT score at baseline, and time were significant predictors (baseline MASA score: p = 0.0002; baseline OHAT score: p = 0.017; time: p = 0.002). There was no significant interaction between the baseline MASA score and time (p = 0.791), between the baseline OHAT score and time (p = 0.117), or between the MASA and OHAT scores at baseline (p = 0.091). The overall percentage correctly classified for pneumonia occurrence during follow-up in the model including the OHAT score, the MASA score and time was 92.7%. The OR for pneumonia during follow-up was 0.97 (95% CI [0.96, 0.99]) for the MASA score and 1.26 (95% CI [1.04, 1.52] for the OHAT score (Table 3.2).

Table 3.2 Effects of the baseline MASA score, baseline OHAT score and time as predictors of the occurrence of pneumonia during follow-up

Effect	OR[a]	Lower bound 95% CI[b,c]	Upper bound 95% CI[b,c]
Intercept	1.406	0.122	16.179
MASA score at baseline	0.974	0.961	0.987
OHAT score at baseline	1.259	1.043	1.520
Time	0.706	0.567	0.878

[a] odds ratio

[b] confidence interval

[c] Wald type

3.3.4 MASA score as a predictor of feeding status, functional independence and place of residence during follow-up

In the evaluation of the baseline MASA score as a predictor of outcomes over time, we noted a significant interaction between the baseline MASA score and time in the FOIS ($p < 0.0001$) and Barthel index ($p = 0.021$) outcomes. This finding indicates that the MASA score at baseline has a significant effect on feeding status and functional independence over time, and this effect is different depending on the time point considered. Figure 3.3 shows the FOIS estimates during the follow-up for different MASA values at baseline. The higher the MASA score at baseline, the higher the FOIS score over time. The FOIS score showed a greater improvement in the short term (from baseline to 3 months) than over the long term (from 6 months onwards). A more rapidly changing FOIS score was observed in patients with a lower MASA score at baseline (Fig. 3.3). Similar results were found for the Barthel index (Fig. 3.3). A higher MASA score at baseline resulted in higher Barthel index estimates over time. The Barthel index outcomes showed a stronger improvement in the short term (from baseline to 3 months) than in the long term but increased faster in patients with higher MASA scores than in patients with lower MASA scores from baseline to 1 month. The baseline MASA score and time were both significant predictors of the place of residence outcome of 'home versus institution' ($p < 0.001$); however, the interaction between the baseline MASA score and time was not significant ($p = 0.571$). The OR for returning home was 1.09 for the MASA (Table 3.3). The odds of returning home were smaller at 1 month than those at baseline (OR: 0.59); however, this ratio was not significantly different from 1 (95% CI [0.22, 1.62]). At 3 and 6 months and 1 year, the OR for returning home compared to baseline was significantly different from 1 and

increased over time (Table 3.3). This model using the baseline MASA score and time was able to correctly predict home as the residence in 78.1% of cases.

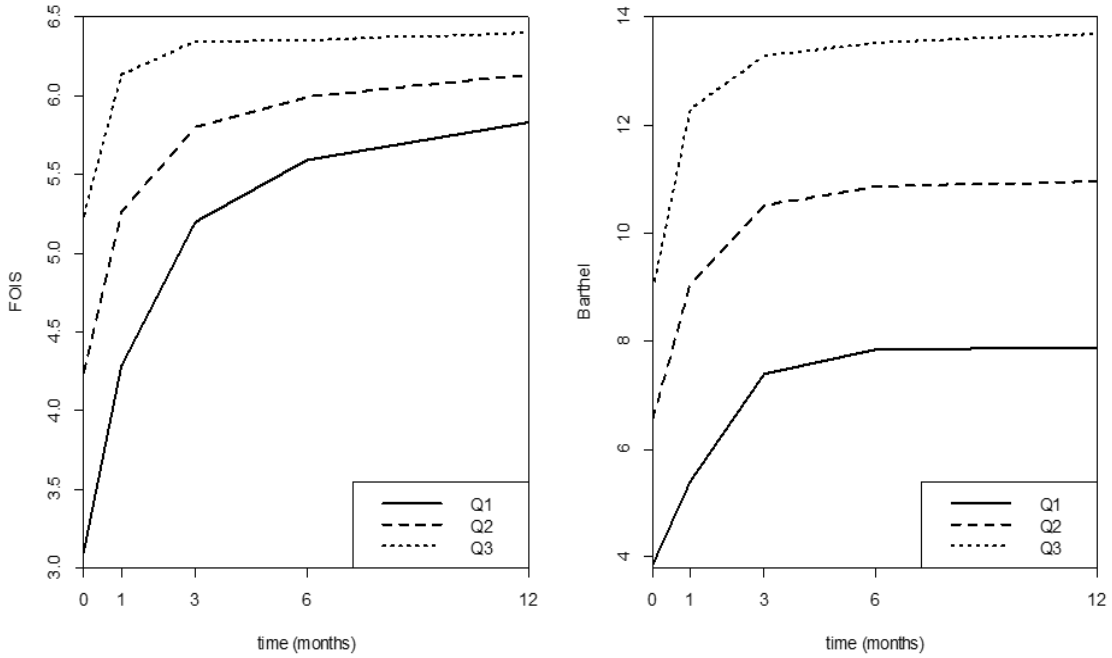


Figure 3.3 FOIS and Barthel index estimates over time based on the linear mixed model evaluated at the quartiles of the baseline MASA scores (Q1 = score 125, Q2 = score 152, and Q3 = score 176)

Table 3.3 Odds ratios and 95% CIs for the baseline MASA score and time as predictors of home as residence during follow-up

Effect	Time, months	OR[a]	Lower bound 95% CI[b,c]	Upper bound 95% CI[b,c]
Intercept		3.210^{E}-07	3.596^{E}-09	2.865^{E}-05
MASA score at baseline		1.094	1.066	1.124
Time	1	0.593	0.217	1.623
Time	3	5.157	1.699	15.652
Time	6	21.245	6.085	74.158
Time	12	44.155	11.188	174.269

[a] odds ratio

[b] confidence interval

[c] Wald type

By comparing the baseline MASA score and the MASA score at each particular time point as predictors of the FOIS and Barthel index outcomes at the different time points, we found that the effect of the baseline MASA score decreased during follow-up (Table 3.4). However, the effect of the MASA score, as assessed at a particular time point during the subsequent year, increased for the FOIS and Barthel index outcomes measured at that specific time point. The baseline MASA score and MASA score at each time point were both significant predictors of the FOIS and Barthel index outcomes during follow-up (Table 3.4). Table 3.5 shows the effect of the MASA score at baseline and the MASA score at particular time points on 'home as residence' at the different time points during the subsequent year. The effect of the baseline MASA score decreased during follow-up but remained significant at all time points. The MASA scores at the considered time points were not a significant predictor of 'home as residence'.

Table 3.4 Estimated effects of the MASA score at baseline and the MASA score at the considered time point as predictors of the FOIS and Barthel index outcomes at the different time points

Time point, months	Intercept	MASA baseline, 95% [CI][b], *p*-value	MASA at specific time point, 95% [CI][b], *p*-value	Time point, months	Intercept	MASA baseline, 95% [CI][b], *p*-value	MASA at specific time point, 95% [CI][b], *p*-value
FOIS 0[a]	-2.130	0.042, [0.037, 0.046], <0.001		Barthel 0[a]	-8.685	0.100, [0.080, 0.120], <0.001	
FOIS 1	-2.587	0.022, [0.016, 0.028], <0.001	0.029*MASA1, [0.024, 0.034], <0.001	Barthel 1	-19.795	0.135, [0.095, 0.174], <0.001	0.051*MASA1, [0.018, 0.085], 0.003
FOIS 3	-3.201	0.008, [0.001, 0.014], 0.026	0.045*MASA3, [0.035, 0.054], <0.001	Barthel 3	-24.211	0.099, [0.051, 0.147], <0.001	0.114*MASA3, [0.046, 0.182], 0.001
FOIS 6	-5.336	-0.006, [-0.012, -0.001], 0.031	0.068*MASA6, [0.059, 0.078], <0.001	Barthel 6	-30.349	0.079, [0.034, 0.124], 0.001	0.165*MASA6, [0.091, 0.239], <0.001
FOIS 12	-4.910	-0.006, [-0.012, -0.001], 0.025	0.066*MASA12, [0.056, 0.076], <0.001	Barthel 12	-38.240	0.081, [0.034, 0.127], 0.001	0.206*MASA12, [0.118, 0.293], <0.001

[a] at baseline

[b] confidence interval

Table 3.5 Estimated effects of the MASA score at baseline and the MASA score at the considered time point as predictors of home as the place of residence at the different time points during the subsequent year

Time point	OR[b] MASA baseline, 95% [CI][c,d], *p*-value[e]	OR[b] MASA at specific time point, 95% [CI][c,d], *p*-value[e]
Home 0[a]	1.065, [1.038, 1.093], <0.001	
Home 1	1.061, [1.013, 1.110], 0.006	1.030, [0.972, 1.091], 0.272
Home 3	1.051, [1.017, 1.087], 0.001	1.040, [0.986, 1.095], 0.127
Home 6	1.050, [1.020, 1.081], <0.001	1.002, [0.962, 1.044], 0.914
Home 12	1.032, [1.003, 1.061], 0.023	1.046, [0.996, 1.098], 0.063

[a] at baseline

[b] odds ratio

[c] confidence interval

[d] Wald type

[e] likelihood test

3.3.5 MASA and OHAT scores as predictors of survival

The Kaplan-Meier curve showed that the cumulative survival proportion was highest in the group with mild dysphagia and lowest in the group with severe dysphagia (Fig. 3.4). The median survival time of individuals with severe dysphagia was 105 days (95% CI [52.75, 157.26]), or less than 4 months. The log-rank test indicated that there was a significant difference in overall survival among the different groups ($p < 0.001$). A significant effect on the hazard rate was found for the baseline MASA score ($p < 0.001$) and age ($p < 0.001$) after correcting for the occurrence of pneumonia over one year and the presence of stroke expansion/recurrence. The hazard ratio for the MASA score was 0.98 (95% CI [0.97, 0.99]), and that for age was 1.08 (95% CI [1.04, 1.12]). Therefore, if the MASA score increased by 1 unit, the hazard of death decreased by approximately 2%, and if age increased by 1 year, the hazard of death increased by approximately 8%. After including the OHAT score at baseline in the Cox-regression model, we found that the baseline MASA score and age remained significant (MASA score: $p = 0.007$; age: $p < 0.001$); however, the OHAT score was not significant ($p = 0.224$). After including the OHAT score instead of the MASA score in the Cox-regression model, we found that the OHAT score showed a trend toward significance ($p = 0.058$). The hazard ratio for the OHAT score was 1.15 (95% CI [0.995, 1.33]), indicating that if the OHAT score increased by 1 unit, the hazard of death increased by approximately 15%.

Of the total patient population, 50 (33.1%) patients died during follow-up. The aetiology of the mortality during the one-year study involved the presence of pneumonia (N = 13, 26.0%), recurrent stroke (N = 2, 4.0%), the combination of pneumonia and other factors (e.g., recurrent stroke combined with pneumonia, progression in stroke-related symptoms) (N = 9, 18.0%) and other factors (e.g., cardiac failure, dehydration and/or malnutrition) (N = 24, 48.0%). In 2 patients (4.0%), the cause of death remained unknown.

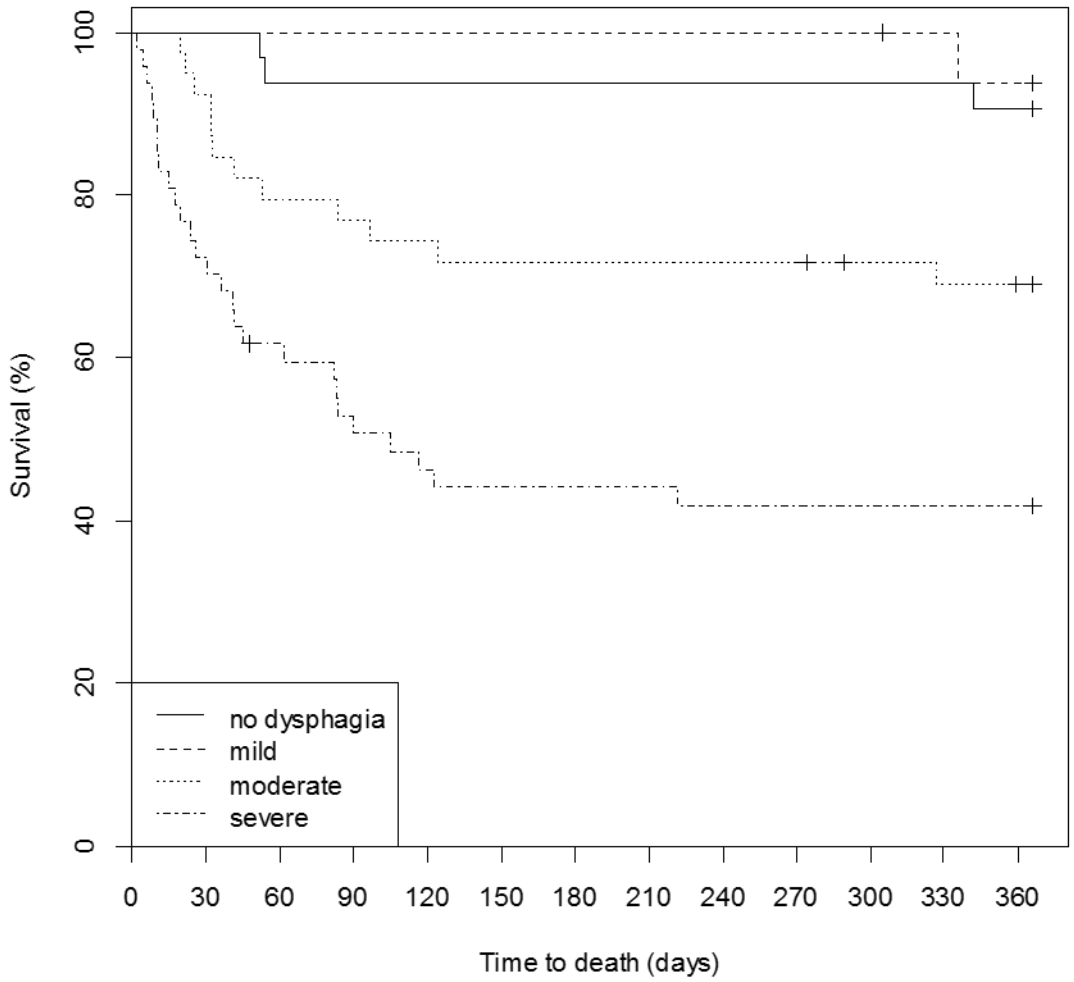


Figure 3.4 Kaplan-Meier survival estimates based on the presence and severity of dysphagia (i.e., MASA score) at baseline (N = 135)

3.4 DISCUSSION

3.4.1 Oropharyngeal dysphagia and type of diet at baseline and discharge destination after hospitalization

Oropharyngeal dysphagia affected three of four patients in our cohort of acute stroke patients, corroborating the incidence rates reported in other studies [160, 161]. Our study results confirmed the hypothesis that PSOD was associated with increased odds (OR: 2.82) of not returning home after discharge, which implies that there is a need to increase health care for these patients. This finding corroborates the results of previous studies [12, 14, 162, 163]. Using the predicted probabilities from the

model including the predictors dysphagia, the presence of pneumonia during hospitalization and age, we were able to predict the discharge destination correctly in nearly 70% of the cases. The mean LOS was significantly higher in dysphagic patients and was highest in tube-fed patients at baseline, confirming the results of other studies [14, 163]. Patients who received a regular diet at baseline were more likely to return home than those fed a dysphagia diet or those who were tube fed. Similarly, tube feeding showed the strongest association with mortality during hospitalization (i.e., 25.0% of tube-fed patients died compared to 12.2% of patients in the dysphagia diet group) and institutionalization after discharge (i.e., 66.6% of tube-fed patients were transferred to another institution, whereas 28.6% of the group fed a regular diet and 43.9% of the group fed a dysphagia diet were institutionalized). The association between the type of diet and the place of residence after discharge has already been reported in the literature [14, 106]. Maeshima et al. [14] found results similar to ours, as most (70%) of their patients who were fed a regular diet were sent home after discharge. Nakajima et al. [106] observed that half of their patients who could eat orally were transferred to home. These studies examined only the association between the type of diet at the time of discharge and the discharge destination but did not investigate the relationship between the type of diet at baseline and the discharge destination, which was investigated in this study. Although building a predictive model using the type of diet at baseline was outside the scope of this study, it warrants further investigation. The earlier and more accurately the prognosis can be predicted, the more effectively home health care services can be organized [14, 164].

3.4.2 MASA and OHAT scores as predictors of pneumonia during hospitalization and follow-up

The baseline MASA and OHAT scores, whether used separately and together, were predictive of pneumonia during hospitalization. We were able to correctly predict the occurrence of pneumonia during hospitalization in nearly 78% of cases using the model with the MASA score, OHAT score and age. The diagnostic accuracy of the MASA and OHAT scores was very high and further validates and confirms the clinical value of these assessment tools. The MASA has advantages over instrumental investigations in that it is inexpensive, non-invasive and easily performed by a SLP. The OHAT, although validated for use in elderly patients but not in specific diseases [90], can be used as a part of the CSE to evaluate the patient's oral health condition [117]. During follow-up, we found similar results regarding the prediction of the occurrence of pneumonia during hospitalization: the combination of the baseline MASA and

baseline OHAT scores could predict pneumonia over time. We were able to correctly predict pneumonia during follow-up in nearly 93% of cases based on the model with the MASA score, OHAT score and time. We found that pneumonia occurred most commonly in the short term after the stroke onset, a result consistent with those reported by Mann et al. [156]. Regarding the MASA, the OR for the prediction of pneumonia during hospitalization (OR: 0.96) was similar to that reported during follow-up (OR: 0.97); the OR for pneumonia for the OHAT was slightly higher during follow-up (OR: 1.26) than during the hospitalization period (OR: 1.23). The interaction between the OHAT score and MASA score at baseline was not significant during hospitalization or follow-up, indicating that the baseline MASA score was not dependent on the baseline OHAT score in predicting pneumonia occurrence and vice versa. The relationship between oropharyngeal dysphagia and pneumonia appears obvious, as dysphagia can lead to aspiration [23, 165], but prior studies [22, 166, 167] did not always find evidence to support this relationship. However, a systematic review of the literature [18] revealed that the risk of pneumonia was 3.17 times higher for dysphagic patients than for patients without dysphagia. Oral health status has also not always been found to be an independent predictor of pneumonia [168]. Because pneumonia is caused by the aspiration of colonized oropharyngeal material into the larynx and lower respiratory airways [22, 23], oral/dental status has been considered a contributing factor to pneumonia in post-stroke patients [22]. Notably, pneumonia is a multi-factorial phenomenon [20]; therefore, repeating binary logistic regression analyses while including more clinical predictors that may contribute to pneumonia occurrence is warranted. Nevertheless, the focus of this study was to evaluate the value of the baseline MASA and OHAT scores for predicting pneumonia during hospitalization and follow-up.

3.4.3 MASA score as a predictor of feeding status, functional independence and place of residence during follow-up

The results showed that a higher baseline MASA score correlated with a better feeding status and functional independence at baseline. The premorbid Barthel index was significantly lower in patients with severe dysphagia, which may have influenced the results. The effect of the baseline MASA score on feeding status and functional independence was stronger in the short term than over the long term. As the year progressed, the MASA score at a particular time point (e.g., the MASA score from 1 month to 1 year for the FOIS and the MASA score from 3 months to 1 year for the Barthel index) became more

informative than the baseline MASA score for predicting the feeding status and functional independence at that time point. These results highlight the need for repeated assessments and a high frequency of post-stroke follow-ups. The baseline MASA score was able to predict home as the residence during follow-up, and the effect was similar throughout the year. Using the predicted probabilities from the model including the baseline MASA and time, we were able to predict the place of residence correctly in approximately 78% of cases. The OR for returning home increased from 3 months to 1 year, which implies improvement among the stroke patients included in this analysis. The evaluation of the effect of the baseline MASA score combined with the MASA scores assessed at specific time points revealed that the baseline MASA score was a significant predictor of home as the place of residence during follow-up, although its effect decreased during the year, meaning that the baseline MASA score became less informative of the place of residence during the follow-up period. Repeated MASA assessments did not appear to be significant predictors of home as the place of residence during follow-up. Future studies should investigate other factors that influence home as the place of residence in the long term.

3.4.4 MASA and OHAT scores as predictors of survival

The MASA score and age were shown to be significant predictors of survival after correcting for the occurrence of pneumonia and the presence of stroke expansion/recurrence during the one-year follow-up. The OHAT score nearly reached significance when it was included instead of the MASA in the model. Prior studies have shown that poor oral health is associated with shorter survival [169, 170]. Future studies should focus on the value of the OHAT in predicting survival after stroke. Only a slight difference in survival was observed between the mild dysphagia group and the group without dysphagia, but this finding may have been influenced by the low number of patients in the group with mild dysphagia (N = 18). Repeating this study with a larger sample may reveal a more prominent difference. We found that the presence of severe PSOD was associated with increased mortality because less than half of the patients with severe PSOD survived one year following their stroke. A similar one-year mortality rate was observed in a study by James et al. [109]. All subjects were supervised by licensed SLPs during their hospitalization; thus, future studies could investigate the extent to which the type and frequency of dysphagia treatment influences the survival rate among individuals with dysphagia of different severities, if ethically feasible.

3.4.5 Further considerations

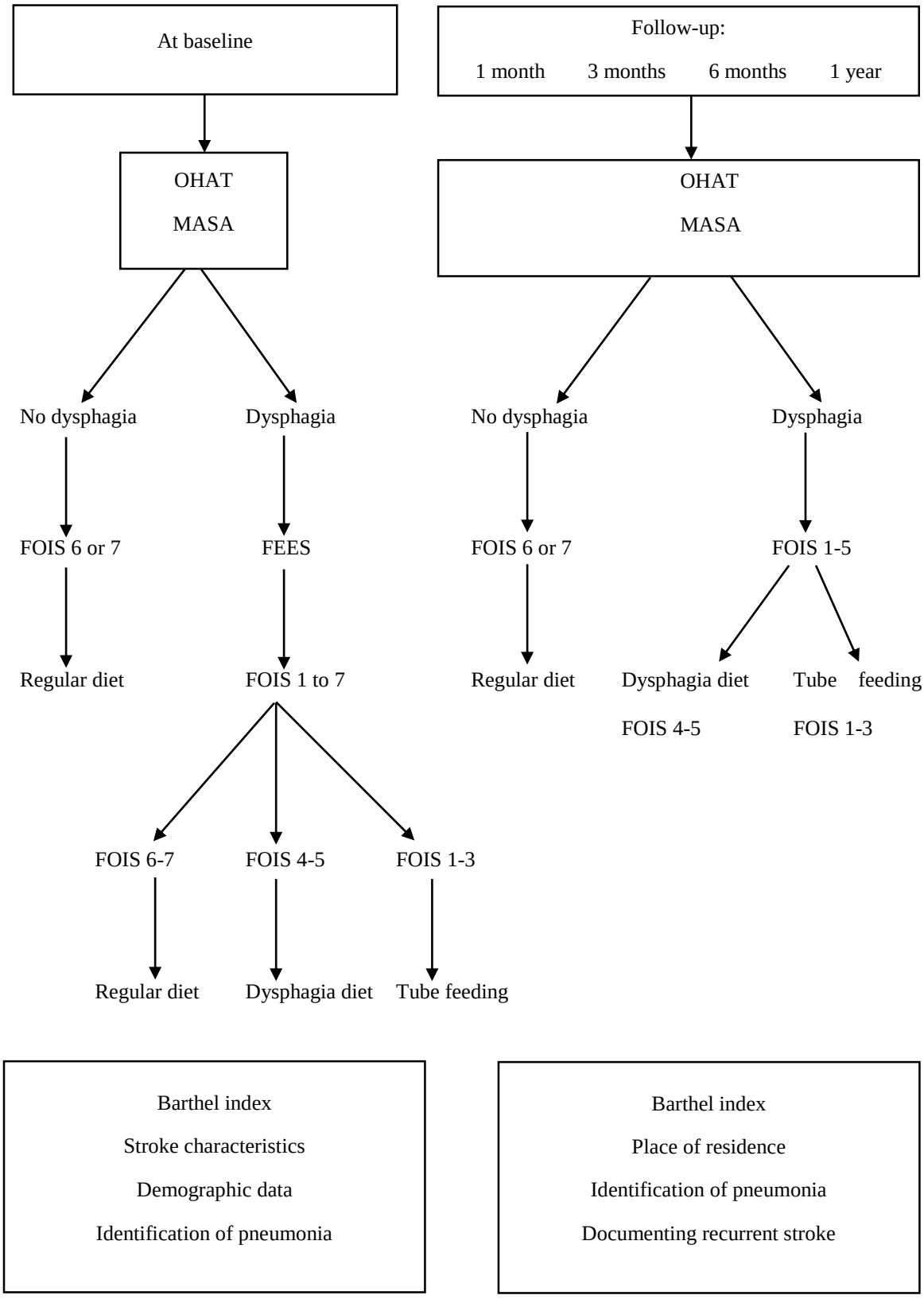
Stroke characteristics were not associated with the presence and severity of dysphagia, results consistent with those of previous studies [1, 66] that did not find an association between stroke subtype (i.e., ischaemic versus haemorrhagic) and the presence of dysphagia. Nevertheless, we did find that the stroke subtype was significantly associated with the type of diet at baseline (Table 3.1). Patients fed a regular diet at baseline were less likely to have a haemorrhagic stroke or a combination of both ischaemic and haemorrhagic stroke than patients fed another diet. Haemorrhagic stroke tends to be more severe than ischaemic stroke and is therefore associated with poorer outcomes than its counterpart [171, 172]. Similar to other studies [106, 89], in this study, we found that stroke severity was significantly associated with dysphagia occurrence and severity and the type of diet at baseline. In general, the number of patients with haemorrhagic stroke in our study was much lower than the number of patients with ischaemic stroke, which is not surprising, as haemorrhagic stroke is less common than ischaemic stroke [173]. However, it is clear that the relationships between stroke subtype or stroke severity and the presence and severity of dysphagia and the type of diet at baseline warrant further investigation. Nevertheless, this was not the focus of our study. This study was limited by the absence of a thorough dental examination to assess the oral health and dental status of the patients. The sensitivity and specificity of the OHAT when administered by SLPs has not been established [117]. However, our study supports further investigations on the potential value of the OHAT for predicting pneumonia and survival after stroke. We did not report the predictive ability of the OHAT for feeding status, functional independence or place of residence during follow-up. These results will be reported in a separate article. Various terms, such as chest infection, post-stroke pneumonia, aspiration pneumonia and stroke-associated pneumonia, are used to describe lower respiratory tract infections [174]. Although efforts have been made to propose standardized terminology and operational criteria for stroke-associated pneumonia, a consensus has not been currently validated [174]. In this study, we followed the Mann criteria [156] and used the term ‘pneumonia’ to refer to lower respiratory tract infections following stroke. Assessments of swallowing function and type of diet were not repeated at discharge for practical reasons. Furthermore, we were particularly interested in determining whether the presence of dysphagia at baseline was predictive of the discharge destination and in the relationship between the type of diet at baseline and the discharge destination. In future studies, it would be interesting to also assess the type of diet and the presence of dysphagia at discharge to determine the extent to which these parameters influence the discharge

destination. The sample size in this study was smaller than the estimated sample size. However, we reached statistical significance in many analyses. Furthermore, a post hoc power calculation, which was performed using PASS 11, revealed that logistic regression of the binary response variable 'residence after discharge different from home' on the binary independent variable 'presence of dysphagia' in a sample size of 112 observations (of which 27.7% did not have PSOD, and 72.3% had PSOD) achieved 90% power at a 0.05 significance level to detect a change in the probability ('discharge destination different from home') from the baseline value of 0.258 to 0.716. This change corresponds to an OR of 7.25. An adjustment was made because multiple regression of the independent variable of interest on the other independent variables in the logistic regression resulted in an R-squared value of 0.500.

3.5 CONCLUSION

For patients with PSOD, this study showed that having PSOD at baseline could predict failure to return home after discharge. Using the MASA as the primary metric of dysphagia and the OHAT to assess oral health, we found that both the baseline MASA and OHAT scores were significant predictors of pneumonia during hospitalization and follow-up. The baseline MASA score has been shown to be an important predictor of feeding status and functional independence, particularly in the short term. As the year progressed, the MASA score at a particular time point (i.e., the MASA score from 1 month to 1 year for the FOIS and the MASA score from 3 months to 1 year for the Barthel index) became more informative than the baseline MASA score for predicting the feeding status and functional independence at that time point. The baseline MASA score was shown to be a significant predictor of the place of residence during follow-up when combined with time and the MASA scores at specific time points. The MASA repeated at specific time points was not a significant predictor of place of residence during follow-up. The baseline MASA score was a significant predictor of survival following stroke, and severe dysphagia was associated with poor survival. Future studies should investigate the potential value of the OHAT for predicting survival after stroke and should assess the effect of dysphagia treatment on stroke survival.

Appendix 3.1 Flow chart representing the different stages of the study



CHAPTER IV

FEASIBILITY AND PSYCHOMETRIC PROPERTIES OF THE ADJUSTED DSWAL-QOL QUESTIONNAIRE FOR DYSPHAGIC PATIENTS WITH ADDITIONAL LANGUAGE AND/OR COGNITIVE IMPAIRMENT: PART I

Study III has been published in:

Simpelaere IS, Vanderwegen J, Wouters K, De Bodt M, Van Nuffelen G. Feasibility and Psychometric Properties of the Adjusted DSWAL-QoL Questionnaire for Dysphagic Patients with Additional Language and/or Cognitive Impairment: Part I. Dysphagia. 2017;32(3):401-19.

ABSTRACT

The Swallowing Quality-of-Life questionnaire (SWAL-QoL) is considered the gold standard for assessing health-related quality of life in people with dysphagia. However, many dysphagic patients struggle to complete this questionnaire because of additional functional sequelae such as language impairment and cognitive disorders. In this study, we sought to develop an adjusted Dutch version of the SWAL-QoL (aDSWAL-QoL) and to evaluate its psychometric properties and feasibility compared with the original questionnaire. We developed the aDSWAL-QoL based on recommendations from previous literature. The feasibility, internal consistency, test-retest reliability, and criterion validity of the aDSWAL-QoL were evaluated in 78 dysphagic patients, among whom 43 had additional language and/or cognitive impairments (DysLC). Statistical analyses were performed using SPSS 20.0. The aDSWAL-QoL had a higher degree of feasibility for the DysLC group. We obtained high Cronbach's α coefficients for total scale and for almost all subscales. Total aDSWAL-QoL scores showed excellent test–retest agreement and good criterion validity with respect to the DSWAL-QoL. Almost all subscales showed significantly moderate to good test–retest agreement and criterion validity. However, the psychometric properties of the 'Food selection' subscale were inadequate. The aDSWAL-QoL is a feasible, reliable, and valid tool for use with DysLC patients. Conversion of the aDSWAL-QoL into an audio computer-assisted self-administered format should be investigated. The construct validity of the aDSWAL-QoL will be evaluated in a separate report.

Keywords: Deglutition, Deglutition disorders, SWAL-QoL, Feasibility, Psychometrics, Language disorders, Cognition disorders

4.1 INTRODUCTION

Patient-reported outcome (PRO) measures, intended to measure health-related quality of life (HRQoL) via a questionnaire or structured interview [48, 49, 51, 52], are defined by the FDA-USHHS as "any report of the status of a patient's health condition that comes directly from the patient, without interpretation of the patient's response by a clinician or anyone else" [53]. PROs are important because they allow patients to communicate about their disease, describe the impact of the disease on their HRQoL [45], and evaluate treatment outcomes [45, 54] from their own perspective [45, 54, 55]. Patient involvement is crucial in the development and validation of these self-reported scales [65].

Use of PROs may therefore be impeded by certain functional impairments, such as communication disorders, often faced by individuals with dysphagia [66, 67]. Dysphagic patients may present with language impairments (i.e., aphasia) and/or cognitive disorders resulting from underlying diseases such as stroke and dementia [68, 69]. For example, aphasia is present in up to 40% of stroke patients [175]. Studies in Belgium and the Netherlands have revealed that the prevalence of cognitive impairment is up to 70% in acute [176], and long-term stroke survivors [177]. Such patients with language or cognition difficulties often struggle with PRO scales and may be unable to complete them [54, 70, 71] owing to comprehension difficulties [72] related to aphasia [73] or poor vision [74]. Indeed, they may face difficulties providing a response [75] that truly reflects their thoughts. Consequently, this group is at times excluded from HRQoL measurement studies [54, 71, 76]. This likely affects the validation of such scales [65], as the results cannot be generalized to the entire population. This problem may be alleviated by the use of interviewer-administered versions of HRQoL scales, which can yield the following advantages: (1) the interviewer can ask additional questions if inconsistent responses are provided; (2) further assistance, such as explanations about the questions, can be given; and (3) the patient's comprehension of the questions can be assessed [71, 77]. However, self-administered questionnaires have an advantage in that they often reveal more delicate health information and can control for interviewer bias [75, 77]. Furthermore, caution is necessary when interpreting HRQoL responses given in a face-to-face interview, because interview responses may be affected by social desirability and/or acquiescent bias. Social desirability bias refers to the tendency of patients to report a better health status and HRQoL when an interviewer is present, i.e. to provide socially desirable responses instead of those that truly represent their HRQoL [75, 77]. Acquiescence bias (which often

takes the form of patients consistently saying ‘yes’ to interviewer questions) refers to the tendency to compliantly agree with others [75, 77].

The Swallowing Quality of life Questionnaire (SWAL-QoL) [61-63] has been translated into Dutch (DSWAL-QoL) and validated for a Flemish population [48]. This multidimensional dysphagia-specific scale has good psychometric properties; it has been considered the gold standard for assessing HRQoL in individuals with oropharyngeal dysphagia [64]. Best practices and extant literature revealed that up to 50% of dysphagic patients frequently needed assistance to complete this questionnaire [48, 118, 119]. Although not specified in these SWAL-QoL studies [48, 118, 119], the need for assistance could have been due to the presence of language disorders [48, 118, 119] or cognitive impairment [118, 119]: patients with neurological disorders were involved in the validation process. Those studies [48, 118, 119] provided limited details about the assistance required despite the fact that this information could allow improvement of the scale’s feasibility for a specific patient population group, namely dysphagic patients with language impairment and/or cognitive disorders (DysLC). To increase the feasibility of the SWAL-QoL, Lemmens et al. [51] developed an interview-administered version. Although this interview version can be considered an important step forward, the format deviates from a self-administered questionnaire, which is conventionally preferred [77]. Such an adjusted DSWAL-QoL-questionnaire (aDSWAL-QoL)—based on aphasia- and cognitive-friendly recommendations from the literature [55, 70, 71, 120-123]—could facilitate self-reporting in DysLC patients and, consequently, increase the reliability of the ratings.

The primary purpose of this study was to develop the aDSWAL-QoL and to examine its feasibility and psychometric properties to be used by DysLC patients. We hypothesized that without compromising the psychometric properties, the aDSWAL-QoL would enhance the feasibility for DysLC patients to self-report about dysphagia-related HRQoL issues. A second aim of this study was to predict the need for assistance to complete the aDSWAL-QoL based on cognitive impairment, language comprehension, age, group, and functional dependency. We hypothesized that the required assistance would increase according to more impaired cognitive or language functions, greater age, being included in the DysLC group, and higher functional dependency. Considering the psychometric properties, we report on internal consistency, test–retest reliability, and criterion validity in this study. The results from the construct validity study of the aDSWAL-QoL will be reported in a separate article (part II).

4.2 MATERIALS AND METHODS

4.2.1 Participants

We conducted a cross-sectional study using convenience sampling and included 78 patients based on the following inclusion criteria: (1) native Dutch speakers, (2) adults (age ≥ 18 years old), and (3) suffering from oropharyngeal dysphagia of mechanical or neurological origin for at least 1 month. The presence of dysphagia was assessed by the Mann Assessment of Swallowing Ability (MASA) [88], and/or objectified by Fiberoptic Endoscopic Evaluation of Swallowing (FEES) [153]. The proposed criteria of the MASA involved the classification of patients according to dysphagia severity [88]. The level of oral intake was determined by the functional oral intake scale (FOIS) [152]. Levels 1–3 refer to varying degrees of non-oral (tube) feeding, levels 4–6 to varying degrees of oral feeding with restrictions, and level 7 to a normal diet without restrictions [152]. For the test–retest assessment, we selected patients with an objectively confirmed and unchanged level of dysphagia during the 2-week follow-up period. The exclusion criteria were (1) severe problems understanding written and spoken Dutch, resulting in the inability to complete the questionnaires; (2) severe attention and/or concentration problems that would affect patient ability to maintain concentration during the assessment; (3) the presence of pure esophageal dysphagia; (4) anosognosia, i.e. unawareness of the existence of dysphagia despite objective confirmation; and (5) both severe visual and hearing impairments that would prevent investigators from successfully providing assistance when required. We assessed language comprehension using the auditory and visual comprehension subtests of the Akense Afasie Test (AAT) [178], and screened cognitive impairment using the Mini Mental State Examination (MMSE) [143]. Standardized cut-off scores for the AAT language comprehension test and for the MMSE were respectively 107 [178] and 27 [179, 180]. Based on the results of the language and cognition tests, patients were grouped according to whether they suffered from dysphagia (Dys group) or DysLC. The total score on the language comprehension test was used to differentiate between the two population groups because the assistance offered by the investigators involved both auditory and visual cues. We also included patients with severe dysarthria or poor intelligibility: if they did not show language or cognitive impairment, we placed them in the Dys group. The Barthel Index [154] was administered by a nurse or an occupational therapist to establish the functional independence of each patient.

The patients were recruited from different settings, including hospitals, nursing homes, rehabilitation centres, and private speech-language pathologist (SLP) practices, and were identified by

SLPs, appropriate staff in nursing homes, and medical doctors based on the inclusion criteria. Ethical approval for the consent procedure and the experimental protocol of the study was granted by the Committee for Medical Ethics of the Antwerp University Hospital and Antwerp University (B300201318058) and the study was conducted in full accordance with the Declaration of Helsinki. Verbal and written consent were obtained from the subjects prior to the start of the study.

4.2.2 Measures

The Original DSWAL-QoL

The DSWAL-QoL is a valid and reliable multidimensional scale that contains 44 items allocated into the following 11 subscales: General burden, Eating desire, Eating duration, Symptoms, Food selection, Communication, Fear of eating, Mental health, Social functioning, Sleep, and Fatigue [48]. Each item is scored on a 5-point Likert scale ranging from 1 (='severely impaired quality of life') to 5 (='no impairment') to reflect the impact of dysphagia on an individual's HRQoL.

4.2.3 Procedures

Our study had two components: (1) the development of the aDSWAL-QoL and (2) the evaluation of its feasibility, reliability, and validity. The flow chart illustrates the different stages of the study (Appendix 4.1).

Development of the aDSWAL-QoL

We developed the aDSWAL-QoL during a five-step investigation involving an DSWAL-QoL scale adjustment process, an expert panel, initial trial assessments, preparation of the actual investigation, and additional trial assessments.

Step 1: Adjustment Process of the DSWAL-QoL

Based on previously reported recommendations [55, 70, 71, 120-123], we made aphasia- and cognitive-friendly adaptations to the original DSWAL-QoL. The language, structural, and visual components of the scale were modified (Appendix 4.2). We used a vocabulary profile of the Dutch language [181] to estimate the reference level of the selected words, as recommended in the Common European Framework of References for Languages [182]. In our modifications, we attempted to only use words that did not exceed the B1 level, which refers to the level of language comprehension and ability required to communicate and understand simple or familiar information about common topics. If a valid alternative was not available, we used more difficult words, because it was considered important to maintain the original meaning. Additionally, we incorporated a simple response category (‘yes’/‘no’) into the aDSWAL-QoL, because this response format is more easily rated compared with the 5-point response category. One exception was the Social functioning subscale, in which the 2-point response category was replaced by a 3-point response category to include ‘I do not know’ as an answer. This was done to maintain similarity with the middle response option from the 5-point response category (‘strongly agree’, ‘agree’, ‘I do not know’, ‘disagree’, ‘strongly disagree’) in the original subscale, thus enabling comparisons with the DSWAL-QoL.

Step 2: Expert Panel

Five SLPs with extensive dysphagia experience were involved in the evaluation of the adjusted questionnaire. They compared the original and adjusted questionnaires to ensure face validity of the instrument. All items were studied, and any issues were discussed. The language of the aDSWAL-QoL was adjusted until agreement was reached.

Step 3: Initial Trial Assessments

To compare the feasibility of completing the two questionnaires, we asked four DysLC patients to complete both the DSWAL-QoL and the aDSWAL-QoL. We particularly wanted to evaluate whether the revised words and sentences in the aDSWAL-QoL were easier to understand, if the response options were clear, if the format was more feasible, and if patients still needed assistance despite the various

modifications. The four patients involved in the trial assessments could not complete the DSWAL-QoL owing to language and cognitive disorders. We found that they were able to complete the aDSWAL-QoL, although they required extensive assistance from the investigators. Thus, we modified some words and sentences in the questionnaire and proposed guidelines for documenting the types of assistance given in the actual investigation. We sought to evaluate the extent to which patients needed assistance in completing the aDSWAL-QoL despite the performed modifications, as this information would facilitate future optimization of the mode of administration.

Step 4: Refining the Instrument and Preparation of the Actual Investigation

With the agreement of the expert panel, we performed further language modifications of the aDSWAL-QoL. For example, some words were replaced by more simple synonyms, given in the Dutch vocabulary profile. Throughout this process, we tried to maintain content validity at all times. Further refinement of the instrument involved three main steps. First, patients were screened for hemianopia or unilateral neglect by means of a line bisection task [183] prior to the administration of the questionnaires. This enabled us to determine whether both questionnaires should be offered to each patient in another format with vertical response scales. Second, we decided that patients would be asked to complete the questionnaire in the presence of two investigators. One investigator would be assigned to provide assistance to the patient when required and the other would document the need and type of assistance for each item. Therefore, we developed an 'investigator's protocol' to facilitate reporting. For both questionnaires, investigator B noted the presence (+) or absence (−) of each type of assistance for each of the 44 items on the protocol. Third, we prepared standard key questions following each item to investigate whether the response matched the patient's intended meaning, and to evaluate the comprehension of the items and response options. This method was based on principles of cognitive interviewing [53].

Step 5: Additional Trial Assessments

The final aDSWAL-QoL and the protocol were examined in two additional patients, prior to the start of the actual investigation, revealing no further issues.

Actual Investigation

After assessing each patient for dysphagia, language impairment, and cognitive impairment, we administered both the DSWAL-QoL and aDSWAL-QoL to each patient. The two questionnaires were completed in random order to minimize recall effects; a minimum of 15 and maximum of 30 minutes elapsed between administrating the first and the second questionnaire. The patients were encouraged to complete the questionnaires as independently as possible. While investigator A provided assistance, investigator B noted the need for and type of assistance. Assistance was given on request or when the patient failed to provide a response. If the patient could not respond due to comprehension problems despite the provided assistance or due to an item that was not applicable to that patient (i.e. in the case of patients that were tube-fed), the response was treated as 'missing'. After the patient made a response for each item, investigator A asked standard key questions (cognitive interviewing procedure). This procedure was applied in a random selection of 32 DysLC patients; owing to the associated burden, not all patients underwent this process. The original response provided by the patient was not justified or corrected. Each response was judged to be 'reliable' or 'unreliable'. We randomly selected 30 patients to assess the test–retest reliability.

4.2.4 Data analysis

We assessed the aDSWAL-QoL in terms of feasibility, reliability (internal consistency and test–retest reliability), and validity (criterion validity) using accepted quality criteria [59]. All statistical analyses were performed using SPSS 20.0 (IBM, SPSS, Inc., Chicago, Il, USA).

Feasibility of the questionnaire

We evaluated the feasibility of the newly developed aDSWAL-QoL in the targeted DysLC group only. Feasibility was determined according to (1) the ability of the patients to independently complete the aDSWAL-QoL versus the original DSWAL-QoL; (2) the need for and type of assistance required; (3) the number of missing responses for both questionnaires; and (4) the response reliability, defined according to whether responses matched the intent. The type of assistance was further specified as the need to (1) read items or questions out loud; (2) repeat items; (3) explain items; (4) read out loud and/or repeat

response options; (5) assist the patient in completing the 5-point response category (i.e. explanation of the 5-point response category); (6) provide motor assistance (i.e. marking the responses or indicating where the response categories can be found on the page, both for the 5-point and simple response categories); and (7) provide a visual demonstration (i.e. some items were visually demonstrated to improve the comprehension of the item. Additionally, we calculated the percentage of DysLC patients who required further assistance (i.e. an explanation) to complete the simple response category. We also assessed the need for and type of assistance required to complete each questionnaire at the retest in 15 DysLC patients.

Differences between the two questionnaires in terms of the need for and type of assistance required to complete the questionnaires, the number of missing responses, and the response reliability for the total scale and for the individual subscales of each questionnaire were pairwise assessed using the Wilcoxon signed-rank test (with $p < 0.05$ considered significant). We aimed to evaluate whether the 2-point response format (and the 3-point response format for the Social functioning subscale) of the aDSWAL-QoL produced similar responses as the 5-point response format of the aDSWAL-QoL. We therefore studied the correlations between the simple format and the 5-point response format using Spearman's rank correlation coefficient (Spearman's rho) with accepted cut-off values (<0.25 = poor, 0.25–0.50 = fair, 0.50–0.75 = moderate to good, >0.75 = good to excellent) [144] and the intraclass correlation coefficient (ICC) with a two-way random effects model with measures of absolute agreement. The correlation between the two response categories was considered good if the ICC reached at least ≥0.70 [59].

Reliability

The reliability of the aDSWAL-QoL was assessed by evaluating the internal consistency and test-retest reliability.

Internal Consistency

Internal consistency is defined as the extent to which the items of a scale are well-correlated and measuring the same concept [59]. Internal consistency is assessed using Cronbach's α; a high Cronbach's α ranges between 0.70 and 0.95.

Test–Retest Reliability

We chose a time period of 2 weeks between the test and retest moment, because we did not expect patients to recall their previous responses after this duration. We constructed one patient group encompassing an equal number of Dys and DysLC patients. We used the results of the 5-point and simple response categories from both Dys and DysLC patients to evaluate the test–retest reliability of the adjusted scale. We used similar criteria as in the feasibility analysis to assess test–retest reliability.

Validity

We assessed the validity of the aDSWAL-QoL by evaluating the criterion validity using both the 5-point and the simple response categories. The content validity of the SWAL-QoL has been well established [63]. Thus, we assumed that the aDSWAL-QoL would have similar content validity because this test measures the same concepts as the DSWAL-QoL [48] and SWAL-QoL [63].

Criterion Validity

Criterion validity is defined as "the extent to which scores on a particular questionnaire relate to a gold standard" [59]. In this study, the DSWAL-QoL is the gold standard and the aDSWAL-QoL is the particular questionnaire. To reduce the impact of varying degrees of language comprehension among the patients, we only assessed criterion validity in the Dys group. To this end, we used similar quality criteria [59] as those used in the feasibility and test–retest reliability analyses.

Associations Between Functional Measures and the Need for Assistance

We investigated the correlations between the level of cognitive impairment (measured by the MMSE) and the need for assistance in completing the aDSWAL-QoL in both patient groups using Spearman's rank correlation coefficient. We examined similar correlations for the level of language comprehension (measured by the AAT), and the level of functional dependency (measured by the Barthel Index). We also assessed the correlation between the amount of motor assistance and the Barthel Index. We used a linear regression model to predict the need for assistance in completing the aDSWAL-QoL as a function of the

MMSE, AAT, Barthel Index, age, and group status. Multicollinearity was assessed by means of the variance inflation factor and standard residuals were checked for normality using QQ-plots and the Shapiro–Wilk test.

4.3 RESULTS

4.3.1 Patient Characteristics

Of the 119 patients recruited during the inclusion phase, 78 patients met all inclusion criteria. Some patients were considered too severely affected to participate (based on MMSE < 5 and AAT < 37). Other reasons for exclusion from the data analysis were as follows: (1) not completing both questionnaires, (2) subjective dysphagia without objective identification, or (3) insufficiently determined aetiology of dysphagia. Of the 78 patients, 35 patients comprised the Dys group; the DysLC group consisted of 43 patients, of whom only four displayed no cognitive impairment (MMSE ≥ 27). Three patients (1 Dys and 2 DysLC) showed unilateral neglect or hemianopia; therefore, a vertical response format was offered. Table 4.1 represents the demographic characteristics of the patients. As indicated by the FOIS, this study included a small proportion of tube-fed patients (20% in the Dys group, 7% in the DysLC group). Age, as assessed by Shapiro–Wilk and QQ-plots, showed a normal distribution in each group although statistical differences for age—as assessed by the independent *t* test—were found between both groups. We found no statistical differences based on the level of dysphagia or gender in both groups, as assessed by the Mann–Whitney U-test and the Chi square test, respectively.

Table 4.1 Demographic characteristics of the patients (*n* = 78)

Characteristic	Dys (*N* = 35)		DysLC (*N* = 43)	
Age (years)				
Mean (SD) (min, max)	62	(13.01) (35–89)	77	(11.04) (52–94)
Gender, N, %				
Male	23	65.7	22	51.2
Female	12	34.3	21	48.8
Aetiology, N, %				
Stroke	14	40.0	23	53.5
Head and neck cancer	17	48.6	2	4.7
Parkinson's disease	1	2.9	6	14.0
Amyotrophic lateral sclerosis	1	2.9		
Multiple sclerosis	1	2.9		
Corticobasal degeneration	1	2.9		
Presbyphagia			8	18.6
Dementia			3	7.0
Cerebral palsy			1	2.3
Dysphagia (MASA)				
Mean (SD)	157.31	(16.68)	160.93	(10.27)
Dysphagia, N, %				
Severe (≤138)	7	20.0	1	2.3
Moderate (≤139–167)	16	45.7	27	62.8
Mild (≤168–177)	12	34.3	15	34.9
Normal swallowing (≤178–200)	0	0	0	0
Aspiration, N, %				
Severe (≤140)	8	22.9	3	7.0
Moderate (≤148)	3	8.6	3	7.0
Mild (≤149–169)	15	42.9	31	72.1
No aspiration (≤170–200)	9	25.7	6	14.0
Functional oral intake (FOIS)				
Mean (SD) (min, max)	4.40	(1.59) (1-6)	4.88	(0.76) (2-6)
Level, N, %				
1–3	7	20	3	7
4–5	20	57.2	35	81.4
6	8	22.9	5	11.6
7	0	0	0	0
MMSE				
Mean (SD) (min, max)	28.37	(1.11) (27–30)	20.35	(4.85) (5–28)
AAT				
Visual comprehension, mean (SD)	56.09	(2.54)	39.84	(10.24)
Auditory comprehension, mean (SD)	55.43	(2.81)	36.37	(9.54)
Total AAT score, mean (SD) (min, max)	111.54	(4.461) (107–120)	76.21	(17.69) (37–106)
Barthel index				
Mean (SD) (min, max)	16.79	(5.43) (3–20)	6.91	(6.21) (0–20)
Highest completed level of education, N, %				

Table 4.1 continued

Characteristic	Dys (*N* = 35)		DysLC (*N* = 43)	
Primary school	14	40.0	30	69.8
High school	12	34.3	11	25.6
University College	7	20.0	2	4.7
University	2	5.7		
Current Residence, N, %				
Home	27	77.1	6	14.0
Nursing home	4	11.4	30	69.8
Hospital	1	2.9	5	11.6
Rehabilation centre	3	8.6	1	2.3
Assisted living facility			1	2.3

4.3.2 Score distribution of the aDSWAL-QoL

We found positive coefficients of skewness for the subscales General burden, Eating duration, and Communication, indicating a greater number of low scores, whereas the other subscales were negatively skewed, indicating a lower number of low scores. The score distribution ranged from 0 to 100 for all subscales (except for the Symptoms and Mental health subscales) with mean scores ranging from 41.99 to 72.68 (Table 4.2). Table 4.2 shows the floor and ceiling effects for all subscales as well as for the total scale. Ceiling effects were only present for the Sleep and Food selection subscales. Floor effects were not observed for any subscale or for the total scale.

Table 4.2 Score distribution and floor and ceiling effects for the subscales of the aDSWAL-QoL and for the total scale

	N	No. items	Min	Max	Mean	SD	Coefficient of skewness	Floor effects (%)	Ceiling effects (%)
General burden	78	2	0	100	46.00	31.02	0.23	11.5	9.0
Eating duration	73	2	0	100	41.99	32.06	0.55	11.5	11.5
Eating desire	78	3	0	100	59.63	26.67	−0.55	3.8	7.7
Symptoms	77	14	18	97	64.55	17.32	−0.35	1.3[a]	1.3[b]
Food selection	75	2	0	100	72.68	24.97	−0.56	1.3	26.9[c]
Communication	77	2	0	100	53.26	26.95	0.07	3.8	11.5
Fear of eating	75	4	0	100	53.96	19.25	−0.26	1.3	2.6
Mental health	77	5	5	100	64.40	24.86	−0.46	2.6[a]	10.3
Social functioning	77	5	0	100	61.82	28.04	−0.43	2.6	12.8
Sleep	77	2	0	100	56.17	33.88	−0.39	12.8	15.4[c]
Fatigue	77	3	0	100	60.05	26.21	−0.32	2.6	11.5
Total scale	78	44	14	94	57.41	16.35	−0.06	1.3	1.3

[a] Floor score ≠ 0

[b] Ceiling score ≠ 100

[c] Effect present

4.3.3 Feasibility

Two DysLC patients were able to complete the aDSWAL-QoL entirely independently, while only 1 DysLC patient independently completed the DSWAL-QoL.

The DysLC group required significantly ($p < 0.05$) less assistance when completing the aDSWAL-QoL items compared with the DSWAL-QoL items. Regarding the types and degree of assistance required, we found significant differences between the two questionnaires (Table 4.3). The simple response format (i.e. 2- or 3-point response category) was completed by 44% of the DysLC group without further explanation, while the others—except for two patients—needed assistance to rate the simple response format in the first two items of the aDSWAL-QoL. We observed significantly fewer missing responses ($p < 0.05$) in the aDSWAL-QoL compared with the DSWAL-QoL questionnaire (38 and 150, respectively) for the DysLC group. Pairwise comparison of the response reliability for both questionnaires showed that the responses provided by the DysLC patients were more reliable for the aDSWAL-QoL than those for the DSWAL-QoL (Fig. 4.1); only one patient had more than 10 unreliable responses on the aDSWAL-QoL. The response reliability for the individual subscales was significantly ($p < 0.05$) lower for the DSWAL-QoL, with the exception of the Communication and Sleep subscales. We found a significantly high correlation between the aDSWAL-QoL scores resulting from the 5-point and simple response categories for the total scale (ICC = 0.87 [95% confidence interval = 0.75–0.93]; $r_s = 0.91$, $p < 0.01$) and for all subscales (ICC ≥ 0.68; $r_s \geq 0.75$, $p < 0.01$).

Table 4.3 Patients in the DysLC group ($N = 43$) required more assistance when completing the DSWAL-QoL versus the aDSWAL-QoL

Type of assistance	DSWAL-QoL %	aDSWAL-QoL %	Sig.[a]
Read out loud items			
Independent	21	61	0.000
Assistance	79	39	
Repeat items			
Independent	12	19	0.000
Assistance	88	81	
Explain items			
Independent	5	9	0.000
Assistance	95	91	
Read out loud response options			
Independent	5	16	0.000
Assistance	95	84	

Table 4.3 continued

Type of assistance	DSWAL-QoL %	aDSWAL-QoL %	Sig.[a]
Assistance to complete the 5-point response category			
Independent	2	14	0.000
Assistance	98	86	
Motor assistance			
Independent	23	47	0.001
Assistance	77	53	
Visual demonstration			
Independent	35	49	0.003
Assistance	65	51	

[a] Significant at 0.01 level.

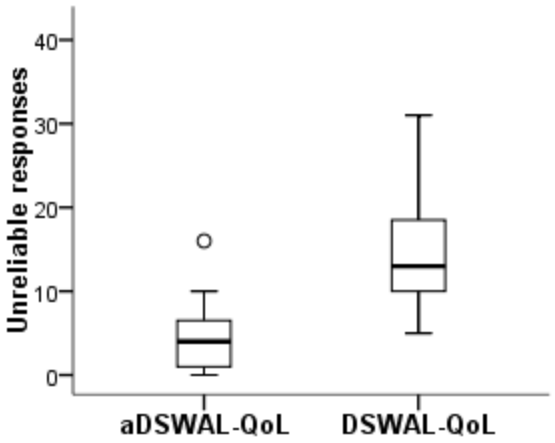


Figure 4.1 Pairwise comparison of the response reliability for both questionnaires for the DysLC group ($N = 32$)

Individual comparisons of the DSWAL-QoL and aDSWAL-QoL for both the test and retest moments revealed that for the aDSWAL-QoL, 15 DysLC patients required significantly less assistance ($p < 0.05$) when the same questionnaire was administered a second time. In particular, the 'read out loud response options' for the aDSWAL-QoL were requested significantly less frequently at the retest moment, and the 'repeat items' were requested significantly more frequently for the DSWAL-QoL at the second test moment (Table 4.4). When comparing both questionnaires at the first test moment, each type of assistance was required less frequently during the aDSWAL-QoL, except for the need for 'motor assistance'. At the retest moment, each type of assistance was required significantly less frequently when completing the aDSWAL-QoL (Table 4.4).

4.3.4 Reliability

Internal Consistency

Table 4.5 shows the internal consistency estimates for the 11 subscales and the total scale for the aDSWAL-QoL. Cronbach's α coefficients were high for total scale and for most subscales, good for Communication, but insufficient for Food selection and Fear of eating. Removing 'Do I know when I'm going to choke?' from the analysis improved the internal consistency for the subscale 'Fear of eating' (α raised from 0.41 to 0.75). Given that the Food selection subscale consists of only two items, no improvement is possible by removing the least consistent item. The internal consistency of the total aDSWAL-QoL scale did not improve by removing the Food selection subscale.

Table 4.4 Percentage of DysLC patients (*N* = 15) who required various types of assistance to complete the DSWAL-QoL versus aDSWAL-QoL at two different test times

Type of assistance	Ind.[a]/ Assist.[b]	Test			Retest			Test-Retest	
		DSWAL-QoL %	aDSWAL-QoL %	DSWAL-QoL/ aDSWAL-QoL[c] Sig.	DSWAL-QoL %	aDSWAL-QoL %	DSWAL-QoL/ aDSWAL-QoL[d] Sig.	DSWAL-QoL[e] Sig.	aDSWAL-QoL[f] Sig.
Read out loud items	Ind.	20	73	0.002[g]	13	93	0.001[g]	0.463	0.068
	Assist.	80	27		87	7			
Repeat items	Ind.	13	20	0.002[g]	0	20	0.005[g]	0.009[g]	0.583
	Assist.	87	80		100	80			
Explain items	Ind.	0	0	0.001[g]	7	13	0.001[g]	0.083	0.447
	Assist.	100	100		93	87			
Read out loud response options	Ind.	0	13	0.008[g]	0	33	0.002[g]	0.752	0.004[g]
	Assist.	100	87		100	67			
Assistance 5-point response category	Ind.	0	13	0.001[g]	7	13	0.002[g]	0.759	0.533
	Assist.	100	87		93	87			
Motor assistance	Ind.	40	40	0.317	20	40	0.043[h]	0.109	0.180
	Assist.	60	60		80	60			
Visual demonstration	Ind.	33	53	0.013[h]	27	53	0.016[h]	0.336	1.000
	Assist.	67	47		73	47			

[a] Independent

[b] Assistance

[c] DSWAL-QoL versus aDSWAL-QoL at the test moment

[d] DSWAL-QoL versus aDSWAL-QoL at the retest moment

[e] Test–retest comparison for DSWAL-QoL

[f] Test–retest comparison for aDSWAL-QoL

[g] Significant at 0.01 level

[h] Significant at 0.05 level

Table 4.5 Reliability estimates for the aDSWAL-QoL: internal consistency and test-retest reliability using both 5-point and simple response formats

aDSWAL-QoL	Internal Consistency		5-point response category Test-retest			Simple response category Test-retest		
	N	Cronbach's α	N	ICC (95% CI)	Spearman's rho	N	ICC (95% CI)	Spearman's rho
General burden	78	0.84	30	0.75 (0.51–0.88)	0.80[a]	30	0.75 (0.54–0.87)	0.75[a]
Eating duration	73	0.86	29	0.58 (0.28–0.78)	0.56[a]	27	0.66 (0.38–0.83)	0.67[a]
Eating desire	76	0.72	30	0.48 (0.15–0.72)	0.41[a]	30	0.66 (0.40–0.82)	0.60[a]
Symptoms	71	0.82	30	0.81 (0.64–0.91)	0.78[a]	30	0.74 (0.51–0.87)	0.72[a]
Food selection	75	0.33	29	0.45 (0.10–0.70)	0.31	27	0.59 (0.28–0.79)	0.42[b]
Communication	77	0.67	30	0.84 (0.69–0.92)	0.80[a]	29	0.66 (0.40–0.83)	0.64[a]
Fear of eating	69	0.41	30	0.79 (0.61–0.89)	0.80[a]	30	0.68 (0.43–0.83)	0.68[a]
Mental health	74	0.85	30	0.83 (0.68–0.92)	0.67[a]	30	0.81 (0.64–0.91)	0.81[a]
Social functioning	77	0.76	30	0.82 (0.65–0.91)	0.80[a]	30	0.89[c] (0.79–0.95)	0.88[a,c]
Sleep	77	0.81	30	0.78 (0.58–0.89)	0.77[a]	30	0.68 (0.42–0.83)	0.67[a]
Fatigue	77	0.85	30	0.88 (0.77–0.94)	0.86[a]	30	0.83 (0.67–0.92)	0.84[a]
Total scale	73	0.80	30	0.91 (0.83–0.96)	0.90[a]	30	0.89 (0.78–0.95)	0.87[a]

[a] Significant at 0.01 level

[b] Significant at 0.05 level

[c] Note that a 3-point response category was used instead of the 2-point response category

Test–Retest reliability

Total aDSWAL-QoL scores showed significantly excellent test-retest agreement (5-point response category: ICC = 0.91; r_s = 0.90; simple response category: ICC = 0.89; r_s = 0.87). All subscales showed significantly moderate to good test–retest agreement, except for the 'Food selection' subscale (Table 4.5).

4.3.5 Validity

Criterion Validity

Total aDSWAL-QoL scale scores showed significantly good agreement with the gold standard DSWAL-QoL for both the 5-point and simple response categories, indicating good criterion validity (Table 4.6). An evaluation of criterion validity for the individual subscales using the 5-point response category revealed that all subscales showed good agreement, except for the Eating desire and Food selection subscales. The General burden and Fear of eating subscales showed a good correlation (r_s = 0.75 and r_s =

0.73, respectively) with the corresponding DSWAL-QoL subscales; however, the agreement did not reach the ICC minimum standard of 0.70. For the simple response category, only the Eating duration, Communication, Mental health, Social functioning, and Sleep subscales showed good agreement (ICC ≥ 0.72) with the DSWAL-QoL subscales (Table 4.6). The missing data in this sample were due to the oral intake status, because not all items were suitable for tube-fed patients.

Table 4.6 Evaluation of criterion validity: correlation between aDSWAL-QoL and DSWAL-QoL (gold standard)

	N	Correlation between aDSWAL-QoL and DSWAL-QoL based on 5-point response format		N	Correlation between aDSWAL-QoL and DSWAL-QoL using the simple response format	
		ICC (95% CI)	Spearman's rho (r_s)[a]		ICC (95% CI)	Spearman's rho (r_s)[a]
General burden	35	0.68 (0.45–0.83)	0.75	35	0.41 (0.10–0.65)	0.54
Eating duration	32	0.81 (0.64–0.90)	0.70	32	0.72 (0.50–0.85)	0.57
Eating desire	35	0.58 (0.31–0.77)	0.57	35	0.64 (0.40–0.80)	0.63
Symptoms	35	0.74 (0.55–0.86)	0.76	35	0.65 (0.38–0.82)	0.76
Food selection	32	0.41 (0.09–0.66)	0.47	32	0.37 (0.03–0.63)	0.44[b]
Communication	35	0.80 (0.61–0.90)	0.84	34	0.72 (0.51–0.85)	0.74
Fear of eating	33	0.67 (0.40–0.83)	0.73	33	0.53 (0.08–0.77)	0.65
Mental health	35	0.79 (0.62–0.89)	0.81	35	0.79 (0.62–0.89)	0.79
Social functioning	35	0.80 (0.63–0.90)	0.81	35	0.80 (0.63–0.89)[c]	0.79
Sleep	35	0.83 (0.69–0.91)	0.81	35	0.72 (0.51–0.85)	0.78
Fatigue	35	0.75 (0.56–0.87)	0.77	35	0.56 (0.26–0.80)	0.66
Total scale	35	0.85 (0.72–0.92)	0.87	35	0.84 (0.70–0.91)	0.85

[a] Significant at 0.01 level

[b] Significant at 0.05 level

[c] Note that a 3-point response format was used instead of the 2-point response format

4.3.6 Associations Between Functional Measures and the Need for Assistance

We found significantly moderate negative correlations ($r_s = -0.51$, $p < 0.01$) between the MMSE and the total percentage of items on the aDSWAL-QoL that required assistance in the DysLC group, with no correlation ($r_s = -0.03$) in the Dys group (Fig. 4.2). Additionally, the Dys group showed no correlation ($r_s = 0.01$) between the AAT score and the total percentage of items on the aDSWAL-QoL for which assistance was needed, while this correlation was moderately negative ($r_s = -0.54$, $p < 0.01$) in the DysLC group. In both groups, we found significantly ($p < 0.01$) moderate negative correlations ($r_s = -0.56$ for Dys; $r_s = -0.45$ for DysLC) between the need for assistance and the Barthel Index (Fig. 4.2). We

also found significantly moderate negative correlations between motor assistance and the Barthel Index in both groups (Dys: $r_s = -0.49$; DysLC: $r_s = -0.44$; $p < 0.01$), indicating that patients who needed assistance in marking their responses were also more functionally dependent. Since AAT and MMSE were highly correlated ($r_s = 0.90$), the AAT was not included in the linear regression model to prevent multicollinearity. As the MMSE is widely used in clinical practice [184], we decided to keep the MMSE in the linear regression model. After correction for MMSE, age, and group, we found significant correlations between the need for assistance and the Barthel Index ($p = 0.001$), MMSE ($p = 0.021$), and group ($p = 0.036$).

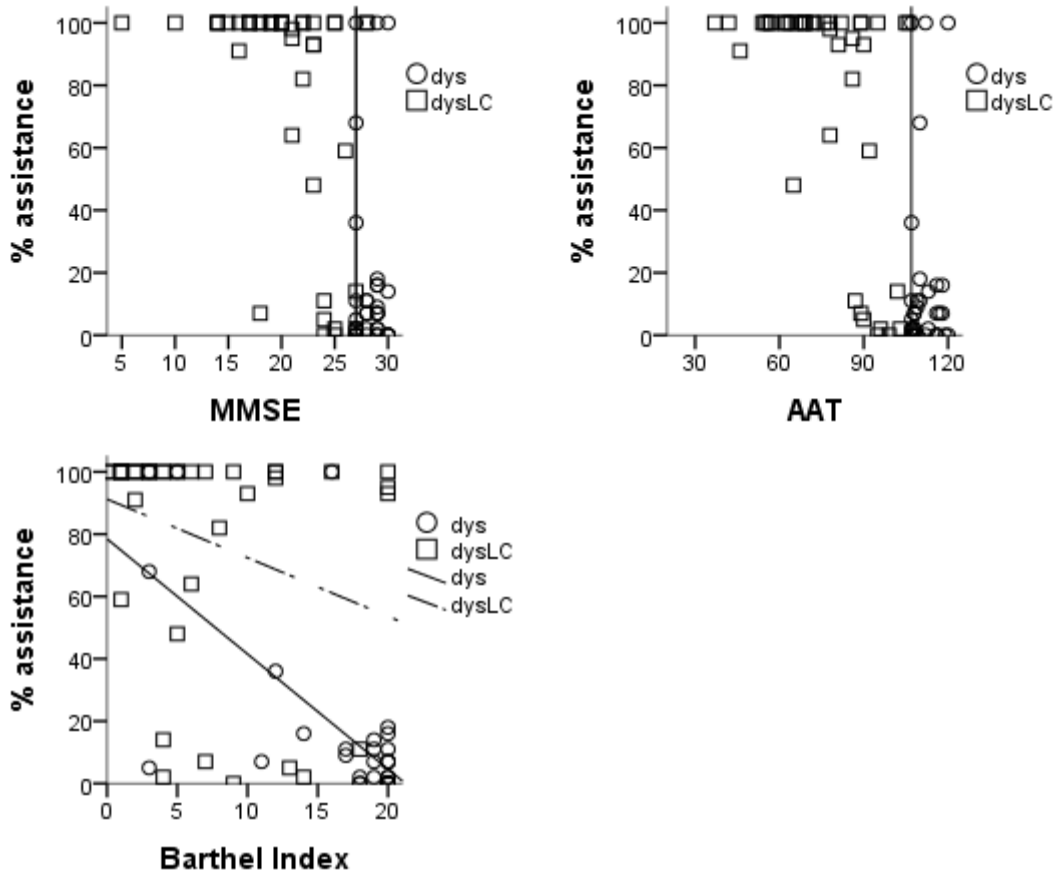


Figure 4.2 Scatterplots for MMSE, AAT, and Barthel Index data versus need for assistance (%). The *vertical line* indicates a cut-off score of ≥27 for the MMSE and ≥107 for the AAT

4.4 DISCUSSION

Our study results indicate that the aDSWAL-QoL is a feasible instrument for use with DysLC patients; it can be used in a population covering a broad range of language and cognitive abilities. We confirmed the internal consistency, test–retest reliability, and criterion validity for the total aDSWAL-QoL and for almost all of its subscales. However, the psychometric properties for the Food selection subscale were insufficient; to a lesser extent, this was also true for the Eating desire and Fear of eating subscales. The level of the MMSE, the type of group, and the Barthel Index could predict the need for assistance in completing the aDSWAL-QoL.

The aDSWAL-QoL was more feasible than the DSWAL-QoL for the DysLC group: they required less assistance, missed fewer responses, and gave more reliable responses. Even with the retest, the DysLC patients showed greater ease in completing the aDSWAL-QoL. We obtained high internal consistency and moderate to excellent test-retest agreement and criterion validity for the total aDSWAL-QoL scale and for almost all the subscales. However, we found that three subscales (Food selection, Fear of eating, and Eating desire) were least able to meet all psychometric requirements. We encountered the greatest difficulty with the Food selection subscale: we were unable to establish internal consistency, test–retest reliability, or criterion validity. The weak internal consistency for this subscale could be explained by the fact that it consists of only two items, which can inherently lead to a lower Cronbach's α [59]. The internal consistency was also too low for the Fear of eating subscale. Similar to the findings of Vanderwegen et al. [48], the internal consistency improved when we removed the item 'Do I know when I'm going to choke?' from the aDSWAL-QoL. For the Eating desire subscale, the criterion validity and test–retest reliability did not attain the recommended quality criteria [59]. A possible explanation for the unsatisfactory validity for both the Food selection and Eating desire subscales is that the language adjustments to the items in the aDSWAL-QoL were so profound that they no longer carried the same meaning. We were also unable to establish the reliability for repeated assessments with these two subscales, indicating weak stability over time. In contrast to our study, all the subscales reported by Vanderwegen et al. [48]—except for the Symptom subscale—attained the recommended test–retest reliability level. However, Vanderwegen et al. used a different population sample: they excluded patients suspected of having cognitive disorders. Lemmens et al. [51] showed that the interview-based version of the DSWAL-QoL also achieved good test–retest agreement (ICC $\geq$ 0.71) for all subscales in a group of patients with communicative and cognitive disorders. However, owing to methodological differences, it is

not possible to compare our study results with those of Lemmens et al. For example, communication problems were defined differently: patients with dysarthria but with preserved language comprehension and cognitive functioning were included in their population sample. Lemmens et al. used the DSWAL-QoL questionnaire validated for a Dutch population [185] instead of the DSWAL-QoL validated for a Flemish population [48]. Furthermore, only 30 patients in our study underwent the test–retest reliability investigation instead of the recommended sample size of 50 people [59]. In particular, many DysLC patients refused to repeat the complete investigation because of the associated burden. It should be noted that this group consisted of especially frail elderly people: the mean age was ≥77 years, and the majority of the DysLC patients were nursing home residents and stroke patients. To attain the final version of the aDSWAL-QoL, we decided not to remove any items or subscales until we had evaluated the construct validity by item analysis. Furthermore, repeating this study with a larger sample could have modified the psychometric properties for the subscales. We did not report the internal consistency or test–retest reliability of the DSWAL-QoL: the response reliability of the DSWAL-QoL was significantly lower than that of the aDSWAL-QoL in the DysLC group. However, this could be a source of bias. The psychometric properties of the aDSWAL-QoL were the prime area of interest in this study.

We prefer to use the 5-point response format in the aDSWAL-QoL; this is because more information can be obtained in this way than with the simple response category, i.e. 2- and 3-point response format [186]. In general, the reliability and validity for the total aDSWAL-QoL scale and most of its subscales were better when we used the 5-point response category. This finding was also observed by Preston and Colman [186]. If the patient is unable to respond to the 5-point response format, relevant information about dysphagia-related HRQoL can still be obtained using the simple response category. Overall, we observed strong correlation between the scores from both the 5-point and simple response categories for the total aDSWAL-QoL scale and all subscales. Furthermore, the psychometric properties, assessed by the simple response format, were sufficient for the total aDSWAL-QoL scale and for most of its subscales. In patients with severe language and cognitive impairment, self-reporting can be impossible. In such situations, proxy-reported outcome measures may be used to assess the patient's HRQoL: proxies can provide relevant information about how they think the patient would report on their HRQoL. Defined as health-care providers, spouses, parents, or relatives closely involved with the patient [84], proxies may be an alternative source of information about a patient's HRQoL that could otherwise be lost [47]. Patient-proxy agreement should be examined to evaluate the reliability of the proxy's ratings [47].

With respect to our second hypothesis, we found moderate negative correlations between the need for assistance in the DysLC group and cognitive impairment, language comprehension, and functional independency. This highlights the increasing need for assistance as functional levels decrease. For the Dys group, we found no such correlations for cognitive impairment or language comprehension. This may be explained by the almost complete lack of variance in the MMSE and AAT scores, i.e. the MMSE score ranged from 27 to 30. We found that in our sample population, an MMSE ≤20, DysLC group membership, and a Barthel Index <7 could predict the need for assistance in completing the aDSWAL-QoL. Future research should establish specific cut-off values for the MMSE and the Barthel Index that indicate which of the two scales (DSWAL-QoL or aDSWAL-QoL) should be administered.

The data for this study were obtained from a convenience sample. Thus, it may be difficult to generalize our results to the total population of patients with oropharyngeal dysphagia. However, it should be noted that the full spectrum of oropharyngeal dysphagia severity—as represented by the different MASA scores—was present in our study population. Only with Symptoms and Mental health did the HRQoL scores not have a range of 0–100. This indicates that all patients experienced dysphagia-related symptoms. All of them were free from the profound emotional impact of dysphagia on HRQoL. However, General burden, Eating duration, and Communication were positively skewed, which means that our study population was aware of the negative impact of dysphagia on their HRQoL, the longer mealtime duration when comparing themselves with other people, and the presence of communication difficulties. Ceiling effects were present for the Sleep and Food selection subscales, despite the high percentage of older patients in the DysLC group. A possible explanation for the ceiling effect for Sleep is the high frequency and chronic use of hypnosedative medication in Belgium [187]. The ceiling effect for Food selection was much higher, which can be explained by the fact that a high percentage of the population, in particular the DysLC population, was recruited from residential care facilities. In these care facilities, patients are dependent on staff for food choices, and thus often receive foods with a stable consistency across meals, such as texture-modified foods or thickened liquids. The aDSWAL-QoL was validated regardless of the underlying aetiology of dysphagia. Although stroke was the most common underlying aetiology among our DysLC patients, other aetiologies, such as Parkinson's disease, were also present. It would be useful to repeat this study with a balanced population. We included patients who had dysphagia for at least 1 month and not those in the immediate post-stroke phase; that was because the DSWAL-QoL has been validated for patients with long-term dysphagia. Additionally, patients are often

physically or mentally unable to self-report at that stage [188]. If the patient's condition permits, it would be interesting to assess dysphagia-related HRQoL in the acute phase following stroke and to evaluate whether the aDSWAL-QoL can be applied at this stage.

We decided to develop one tool that could be used by patients with language impairment and/or cognitive disorders for several reasons: (1) we expected that both patient groups would have difficulties in self-administering a PRO; (2) the combination of both language and cognitive impairment is common in dysphagic patients [66, 67]; and (3) the aphasia- and cognitive-friendly suggestions from the literature show considerable overlap, i.e. the recommended language and visual and structural modifications are both aphasia- and cognitive-friendly. It would be interesting for future research to investigate whether patients could be served more effectively by being dividing into further subgroups, i.e. one group having exclusively language impairment, and modifying the tool according to their needs. However, as our results indicate, the aDSWAL-QoL proved beneficial for patients with language impairment and/or cognitive disorders. Hence, the instrument may be used by a large population.

Both questionnaires were completed in the presence of two investigators. Although the investigators resisted interfering to prevent bias, as in the interview-administered modes, we wanted to evaluate the feasibility of the scale, specifically in terms of the need for assistance and the response reliability. Future research should investigate whether conversion of the pen-and-paper self-administered aDSWAL-QoL into an audio computer-assisted self-administered format could decrease patient reliance on the investigators for assistance. An audio computer-assisted self-administered format is advantageous in that it makes it possible to provide audible instructions while simultaneously providing instructions on the computer screen [75]. For patients with deafblindness, using assistive technology with Braille devices could meet their individual communication needs.

We used the MMSE and did not employ a standardized cognitive assessment battery to evaluate the cognitive abilities: that would have increased the patient burden for participation. We are aware of the impact that language disorders exert on MMSE results; however, this tool has been considered sensitive in screening for moderate to severe cognitive deficits in 1-month post-stroke patients and is the most widely used in hospitalized and outpatient settings [184].

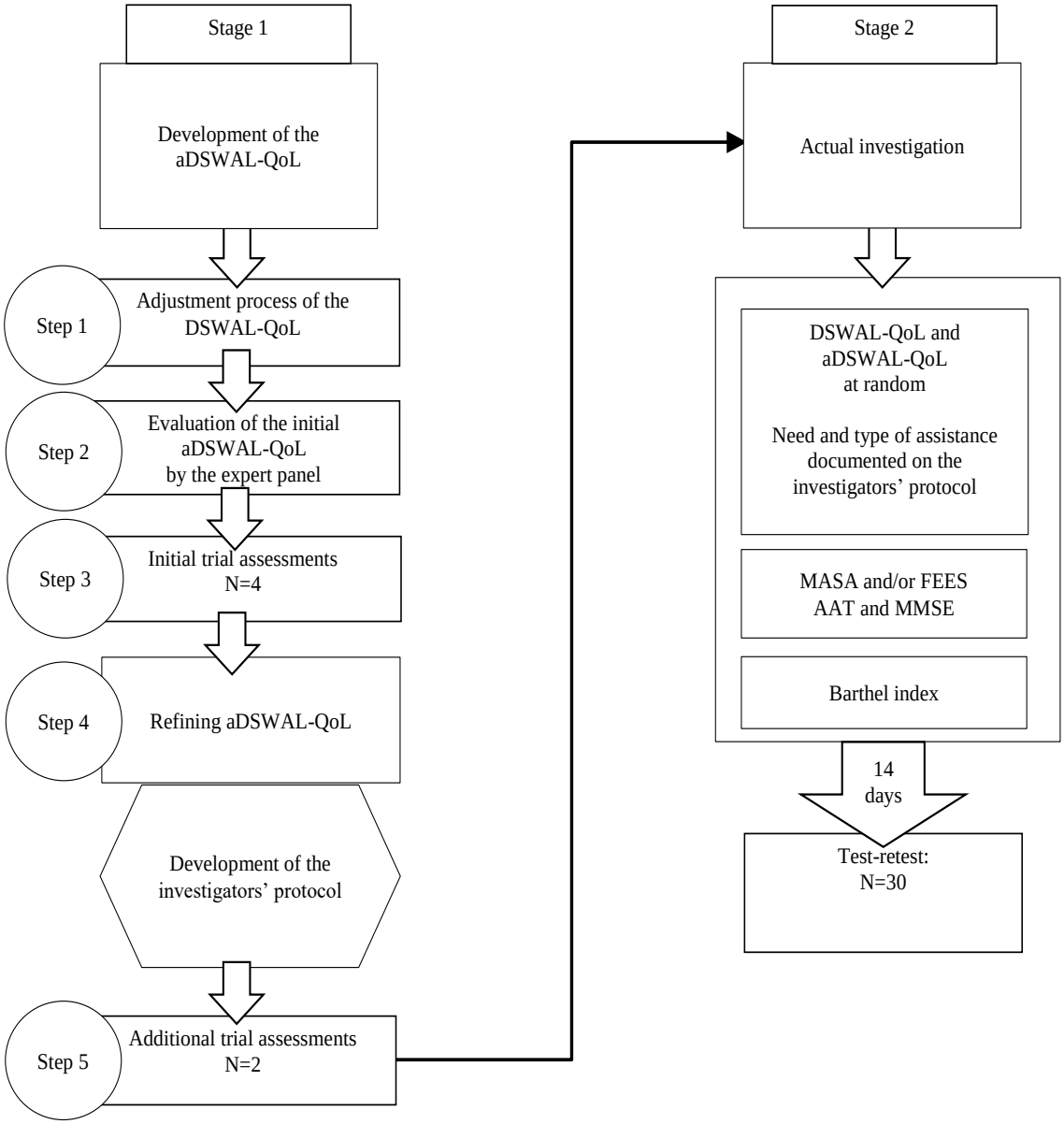
We observed that not all items of the DSWAL-QoL and aDSWAL-QoL were suitable for patients receiving enteral feeding. This is in agreement with previous SWAL-QoL validation studies [48,

51, 57, 63, 119]. We are currently investigating the added value of a newly developed PRO for patients receiving enteral feeding as a consequence of oropharyngeal dysphagia.

4.5 CONCLUSION

The main purpose of this study was to develop an adjusted version of the DSWAL-QoL to be used by DysLC patients, and to evaluate the feasibility and psychometric properties of such a scale. Our results demonstrate the aDSWAL-QoL to be a feasible instrument for use with DysLC patients, which may facilitate self-reporting about the impact of dysphagia on HRQoL. However, further research should focus on converting the pen-and-paper method into a computer-assisted self-administered format and then evaluate the accessibility of such a format for self-reporting HRQoL. We evaluated the psychometric properties, in particular the reliability and criterion validity, of the aDSWAL-QoL. However, caution is required when interpreting the results from the Food selection subscale, as the psychometric properties for this subscale are inadequate in the aDSWAL-QoL. Future research should evaluate the reliability and criterion validity of this subscale in a larger sample to determine the extent to which this subscale is different from the subscale used in the DSWAL-QoL. We were able to predict the need for assistance in completing the aDSWAL-QoL according to the level of the MMSE, the type of group, and the Barthel Index. However, specific cut-off values for specifying which questionnaire to administer for a particular patient are yet to be determined.

Appendix 4.1 Flow chart showing the different stages of the study



Appendix 4.2 Development of the aDSWAL-QoL: adjustment process of the DSWAL-QoL

Type of adjustments	Description		
Language modifications		Example	
	Reduction of information	DSWAL-QoL item	aDSWAL-QoL item
	Any redundant information was removed	"My swallowing problem makes it hard to participate in social activities"	"Are social activities difficult because of my swallowing problem?"
	Statements were transformed into questions	"I don't enjoy eating anymore"	"Do I enjoy eating?"
	Short and simple questions were used	"It's been difficult for me to speak clearly"	"Can I speak clearly?"
	Simplifying questions via syntax modifications		
	Pronouns were simplified so that only the first person was used.	"Choking when you eat food"	"Do I choke while eating?"
	Passive verbs were avoided	"My role with family and friends has changed because of my swallowing problem"	"Does my swallowing problem change my role with family and friends?"
	Simplifying questions via semantic modifications		
	Simple and straightforward language was used: high-frequency, concrete, meaningful words with the goal of preservation of the original information	"I get impatient dealing with my swallowing problem"	"Am I impatient because of my swallowing problem?"
	Complex words were replaced by simple synonyms, if possible	"I don't eat outdoors because of my swallowing problem"	"Do I want to eat alone because of my swallowing problem?"
	Explicit questions, which are easy to read, were used	"Figuring out what I can and can't eat is a problem for me"	"Do I know what I can and cannot eat?"
	Negative phrasing was avoided	"I never know when I am going to choke"	"Do I know when I'm going to choke?"
	Modification of the recall time from last month to last week	"In the last month"	"In the last week"

Appendix 4.2 continued

Type of adjustments	Description		
Structural modifications		Further explanation	
	Changing lay-out		
	Subscales were reorganized in the 5-point response category to follow a consistent format. The subscales with the same response format were placed together. The items within the individual subscales were presented in the same order as in the DSWAL-QoL.	*5-point response category and order of the subscales[a] in the DSWAL-QoL:*	*5-point response category and order of the subscales[a] in the aDSWAL-QoL:*
		6 different response formats	*only 2 different response formats, to reduce confusion*
		GB: “yes, strongly agree”, “quite a bit”, “somewhat”, “a little”, “no, strongly disagree”	GB: “yes”, “quite a bit”, “somewhat”, “a little”, “no”
		EATDUR: “strongly agree”, “quite a bit”, “somewhat”, “a little”, “no, strongly disagree”	EATDUR: “yes”, “quite a bit”, “somewhat”, “a little”, “no”
		EATDES: “strongly agree”, “quite a bit”, “somewhat”, “a little”, “no, strongly disagree”	EATDES: “yes”, “quite a bit”, “somewhat”, “a little”, “no”
		SYMPT: “almost always”, “often”, “sometimes”, “hardly ever”, “never”	FS: “yes”, “quite a bit”, “somewhat”, “a little”, “no”
		FS: “very much true”, “quite a bit”, “somewhat”, “a little”, “not at all”	SF: “yes”, “quite a bit”, “don't know”, “a little”, “no”
		COM: “yes, always”, “most of the time”, “sometimes”, “hardly ever”, “never”	SYMPT: “almost always”, “often”, “sometimes”, “hardly ever”, “never”

Appendix 4.2 continued

Types of adjustments	Description		
		FEAR: “yes, almost always”, “often”, “sometimes”, “hardly ever”, “no, never”	FAT: “almost always”, “often”, “sometimes”, “hardly ever”, “never”
		MH: “yes, always”, “often”, “sometimes”, “hardly ever”, “no, never”	SL: “almost always”, “often”, “sometimes”, “hardly ever”, “never”
		SF: “yes, strongly agree”, “agree”, “don't know”, “disagree”, “no, strongly disagree”	COM: “almost always”, “often”, “sometimes”, “hardly ever”, “never”
		FAT: “yes, always”, “most of the time”, “sometimes”, “hardly ever”, “no, never”	FEAR: “almost always”, “often”, “sometimes”, “hardly ever”, “never”
		SL: “yes, always”, “most of the time”, “sometimes”, “hardly ever”, “no, never”	MH: “almost always”, “often”, “sometimes”, “hardly ever”, “never”
	A simple response category was added (3-point response category and 2-point response category).		3-point scale: “yes, don't know”, no” for SF; 2-point scale: “yes, no” for all other subscales
	Separate pages for each item were provided		
	Each item was represented by the question and the two response formats		i.e. 5-point response category and simple response category
	White space was used between the question and response choices		
	In the case of hemianopia or unilateral neglect, the response scales were presented in a vertical format and the items were presented in the preserved visual field		
	The pages had a white background and page numbering was provided as in the DSWAL-QoL		

Appendix 4.2 continued

Types of adjustments	Description	
	Themes were separated by a visual line drawing	The following themes were provided: "General thoughts" for GB; "Daily meal" for EATDUR and EATDES; "Diet and eating habits" for FS; "Social activities" for SF; "Bodily symptoms" for SYMPT, FAT, and SL; "Communication" for COM; "Feelings" for FEAR and MH.
Visual modifications	Changing colors and fonts	
	Sans-serif fonts (Verdana) were used to increase readability	
	Large print was used: 20 point size for titles and 16 point size for other text	
	Keywords were placed in bold	
	Titles of the different themes were encircled	
	Visual support was provided	
	Visual line drawings (pictures) were used to support the idea that the topic/theme was going to change	
	A visual demonstration was provided, i.e. drooling (demonstration of what happens at the mouth)	
	Comprehension of the response scales was enhanced by providing symbols, colours, and/or line drawings for each response option	

[a] GB = General burden; EATDUR = Eating duration; EATDES = Eating desire; SYMPT = Symptoms; FS = Food selection; COM = Communication; FEAR = Fear of eating; MH = Mental health; SF = Social functioning; FAT = Fatigue; SL = Sleep

CHAPTER V

VALIDATION OF THE DUTCH VERSION OF THE SWALLOWING QUALITY-OF-LIFE QUESTIONNAIRE (DSWAL-QOL) AND THE ADJUSTED DSWAL-QOL (ADSWAL-QOL) USING ITEM ANALYSIS WITH THE RASCH MODEL: A PILOT STUDY

Study V has been published in:

Simpelaere IS, Van Nuffelen G, De Bodt M, Vanderwegen J, Hansen T. Validation of the Dutch version of the Swallowing Quality-of-Life Questionnaire (DSWAL-QoL) and the adjusted DSWAL-QoL (aDSWAL-QoL) using item analysis with the Rasch model: a pilot study. Health and Quality of Life Outcomes. 2017;15(1):66. doi:10.1186/s12955-017-0639-3.

ABSTRACT

The Swallowing Quality-of-Life Questionnaire (SWAL-QoL) is considered the gold standard for assessing health-related QoL in oropharyngeal dysphagia. The Dutch translation (DSWAL-QoL) and its adjusted version (aDSWAL-QoL) have been validated using classical test theory (CTT). However, these scales have not been tested against the Rasch measurement model, which is required to establish the structural validity and objectivity of the total scale and subscale scores. Thus, the purpose of this study was to examine the psychometric properties of these scales using item analysis according to the Rasch model. Item analysis with the Rasch model was performed using RUMM2030 software with previously collected data from a validation study of 108 patients. The assessment included evaluations of overall model fit, reliability, unidimensionality, threshold ordering, individual item and person fits, differential item functioning (DIF), local item dependency (LID) and targeting. The analysis could not establish the psychometric properties of either of the scales or their subscales because they did not fit the Rasch model, and multidimensionality, disordered thresholds, DIF, and/or LID were found. The reliability and power of fit were high for the total scales (PSI = 0.93) but low for most of the subscales (PSI < 0.70). The targeting of persons and items was suboptimal. The main source of misfit was disordered thresholds for both the total scales and subscales. Based on the results of the analysis, adjustments to improve the scales were implemented as follows: disordered thresholds were rescaled, misfit items were removed, and items were split for DIF. However, the multidimensionality and LID could not be resolved. The reliability and power of fit remained low for most of the subscales. This study represents the first analyses of the DSWAL-QoL and aDSWAL-QoL with the Rasch model. Relying on the DSWAL-QoL and aDSWAL-QoL total and subscale scores to make conclusions regarding dysphagia-related HRQoL should be treated with caution before the structural validity and objectivity of both scales have been established. A larger and well-targeted sample is recommended to derive definitive conclusions about the items and scales. Solutions for the psychometric weaknesses suggested by the model and practical implications are discussed.

Keywords: Deglutition disorders, Health-related quality of life, Oropharyngeal dysphagia, Outcome assessment, Psychometrics, Rasch model, SWAL-QoL

5.1 BACKGROUND

Health-related quality of life (HRQoL) refers to a complex, multidimensional construct and is based on the individuals' subjective perceptions of functioning and wellbeing among the physical, psychological and social domains of health [47, 52, 189]. The construct HRQoL is not directly measurable or is unobservable or latent [58] and should preferably be measured with patient-reported outcome (PRO) measures using multiple items, each assessing a different aspect of the underlying construct [190]. Many PROs have been developed to measure HRQoL in patients with oropharyngeal dysphagia [52]. These PROs use self-reported questionnaires, assess the presence and severity of dysphagia symptoms, and measure the influence of dysphagia on a person's HRQoL [48, 52, 56, 57]. These PROs add useful information to the clinical swallowing examination and instrumental investigations [49] and can be used as outcome measures of therapeutic interventions [48, 52, 56, 57]. The applicability and appropriateness of a PRO in a specific population depend on the target population (i.e., persons with oropharyngeal dysphagia), its feasibility, and the quality of its psychometric properties (i.e., reliability and validity) [58, 83].

The Swallowing Quality-of-Life questionnaire (SWAL-QoL) is a 44-item disease-specific scale that is distributed into 10 subscales and the Symptom scale [63]. The SWAL-QoL is considered the gold standard for assessing HRQoL in oropharyngeal dysphagia [64]. The psychometric properties of the SWAL-QoL as well as the Dutch translation of the SWAL-QoL (DSWAL-QoL) and its adjusted version (aDSWAL-QoL) have been demonstrated to be sufficient according to classical test theory (CTT) [48, 63, 124]. The CTT-psychometric assessment included an examination of internal consistency based on Cronbach's alpha, test-retest reliability via the intraclass correlation coefficient (ICC), and/or construct validity based on principal component analysis (PCA) techniques [48, 63, 124]. Some drawbacks related to CTT methods are recognized [125, 126], such as test and sample dependence [125-127] and the assumption of equal weight for all of the items even if there is a difference in the level of difficulty [127]. The scale's total sum score is based on ordinal values and the standard error of measurement is assumed to be constant [83, 126], in contrast to the Rasch methodology.

The Rasch model within modern item response theory (IRT) has been considered the gold standard against which scales summarizing item responses must be tested [80]. Item analysis using the Rasch model involves formal testing of a scale against a mathematic measurement model that specifies what should be expected in the item responses to provide interval-based measures instead of ordinal

values [81, 82]. Interval measures are preferable to ordinal scales because they provide meaningful information about the relative differences and equivalences within the categories of the scale and enable the use of parametric statistics, which provide more powerful and precise results [125, 144]. If the observed data fit the model, the following can be concluded: interval data have been generated, the measurement scale demonstrates structural validity and objectivity, and the total score is statistically sufficient [80]. Structural validity is an aspect of construct validity and evaluates the extent to which the scores of a HRQoL-PRO are an adequate reflection of the dimensionality of the construct being measured [58]. Objectivity implies invariance, which indicates that the comparison between two persons should be independent of which particular items have been used and vice versa [79]; therefore, the instrument should work in the same manner across all persons and items. In contrast to CTT, Rasch measurements allow for the provision of scale-independent person estimates and sample-independent item estimates [190]. The total score of a scale is statistically sufficient if the assessment of the latent variable (i.e., HRQoL) is only a function of that total score and does not depend on the conditional distribution of the item responses underlying the total score [80]. Four assumptions should be satisfied for a measurement scale to meet the criteria of validity, objectivity and statistical sufficiency: 1) unidimensionality (all items in the scale measure the same single construct) [79, 80], 2) monotonicity (the scale items function hierarchically from easy to difficult, with increased item scores corresponding to increased levels of underlying ability) [80, 191], 3) local item independency (a person's score on one item does not depend on their score on another item), and 4) no differential item functioning (DIF, i.e., a particular item's score does not differ due to other factors, e.g., age, for persons with equal ability levels) [79, 80]. In addition to identifying measurement weaknesses, analysis with the Rasch model provides potential solutions for scale improvement. Such improvement has previously been demonstrated with the Taiwan Chinese version of the EORTC QLQ-PR25 questionnaire [128], St. George's Respiratory Questionnaire (SGRQ) [127], and the Patient-Rated Elbow Evaluation (PREE) questionnaire [129].

The purpose of this study was to assess the structural validity and objectivity of both the DSWAL-QoL and aDSWAL-QoL scales and subscales and the statistical sufficiency of the total score and subscale scores using item analysis with the Rasch model.

5.2 METHODS

5.2.1 Participants

A portion of the data was derived from a previous validation study of the aDSWAL-QoL, which has been extensively reported elsewhere [124]. Therefore, the design will be briefly described. A cross-sectional study using convenience sampling was conducted and included 108 persons, among whom 78 were involved in the previous study [124]. People were selected if they were (1) native Dutch speakers, (2) adults (age ≥18 years old) and (3) had oropharyngeal dysphagia of mechanical or neurological origin as assessed with the Mann Assessment of Swallowing Ability (MASA) [88] and/or the Fiberoptic Endoscopic Evaluation of Swallowing (FEES) [153]. Persons without oropharyngeal dysphagia but with a confirmed language and/or cognitive impairment as measured by the auditory and visual comprehension subtests of the Akense Afasie Test (AAT) [178] and the Mini Mental State Examination (MMSE) [143] were also included in this study. Persons were classified into three groups according to whether they suffered from dysphagia (Dys group), had dysphagia accompanied by a language impairment and/or cognitive disorder (DysLC group), or suffered from a language impairment and/or cognitive disorder without the presence of dysphagia (LC group). The proposed criteria for the MASA [88] (further specified in Table 5.1) and the standardized cut-off scores of 107 for the AAT [178] and 27 for the MMSE [179, 180] were used to compose the groups. The exclusion criteria were as follows: (1) severe problems understanding written and spoken Dutch resulting in the inability to complete the questionnaires; (2) severe attention and/or concentration problems that affected the person's ability to maintain concentration during the assessment; (3) the presence of purely esophageal dysphagia; (4) anosognosia, i.e., being unaware of the existence of dysphagia despite clinical confirmation; and (5) severe visual and hearing impairments that prevented the investigators from successfully providing assistance when required. The people were recruited from different settings that included hospitals, rehabilitation centers, nursing homes, and private speech-language pathologist (SLP) practices and were identified by SLPs, the appropriate staff in nursing homes and medical doctors based on the inclusion criteria. Verbal and written consent were obtained from the participants prior to the start of the study. Ethical approval for the consent procedure and the experimental protocol of the study was granted by the Committee for Medical Ethics of the Antwerp University Hospital and Antwerp University (B300201318058), and the study was conducted in full accordance with the Declaration of Helsinki.

Table 5.1 Demographic characteristics of the subjects (*N* = 108)

Characteristic	Dys (*N* = 35)		DysLC (*N* = 43)		LC (*N* = 30)	
Age (years)						
Mean (SD)	62	(13.01)	77	(11.04)	81	(14.65)
(Min, Max)		(35–89)		(52–94)		(20–95)
Gender, *N*, %						
Male	23	65.7	22	51.2	12	40.0
Female	12	34.3	21	48.8	18	60.0
Etiology, *N*, %						
Stroke	14	40.0	23	53.5	4	13.3
Head trauma					1	3.3
Head and neck cancer	17	48.6	2	4.7	0	0.0
Parkinson's disease	1	2.9	6	14.0	4	13.3
Amyotrophic lateral sclerosis	1	2.9				
Multiple sclerosis	1	2.9				
Corticobasal degeneration	1	2.9				
Presbyphagia			8	18.6		
Dementia			3	7.0	18	60.0
Depression					3	10.0
Cerebral palsy			1	2.3		
Dysphagia (MASA)						
Mean (SD)	157.31	(16.68)	160.93	(10.27)	188.67	(5.41)
Dysphagia, *N*, %						
Severe (≤138)	7	20.0	1	2.3		
Moderate (≤139-167)	16	45.7	27	62.8		
Mild (≤168-177)	12	34.3	15	34.9		
Normal swallowing (≤178-200)	0	0	0	0	30	100.0
Aspiration, *N*, %						
Severe (≤140)	8	22.9	3	7.0		
Moderate (≤148)	3	8.6	3	7.0		
Mild (≤149-169)	15	42.9	31	72.1		
No aspiration (≤170-200)	9	25.7	6	14.0	30	100.0
MMSE						
Mean (SD)	28.37	(1.11)	20.35	(4.85)	19.10	(4.77)
(Min, Max)		(27–30)		(5–28)		(4–26)
AAT						
Visual Comprehension,						
Mean (SD)	56.09	(2.54)	39.84	(10.24)	38.77	(9.09)
Auditory Comprehension,						
Mean (SD)	55.43	(2.81)	36.37	(9.54)	34.90	(8.64)
Total score AAT, Mean (SD)	111.54	(4.46)	76.21	(17.69)	73.67	(15.46)
(Min, Max)		(107–120)		(37–106)		(31–95)
Highest completed education, *N*, %						
Primary school	14	40.0	30	69.8	18	60.0
High school	12	34.3	11	25.6	11	36.7
University college	7	20.0	2	4.7	1	3.3
University	2	5.7				
Place of living, *N*, %						

Table 5.1 continued

Home	27	77.1	6	14.0	3	10.0
Nursing home	4	11.4	30	69.8	23	76.7
Hospital	1	2.9	5	11.6	1	3.3
Rehabilitation center	3	8.6	1	2.3	3	10.0
Assisted living facility			1	2.3		

Abbreviations: *Dys* patients suffering from dysphagia, *DysLC* patients with dysphagia accompanied by language impairment and/or cognitive disorders, *LC* patients suffering from language impairment and/or cognitive disorders without the presence of dysphagia, *SD* standard deviation, *Min* Minimum, *Max* Maximum, *N* Number of persons, *%* percentage of people, *MASA* Mann Assessment of Swallowing Ability, *MMSE* Mini Mental State Examination, *AAT* Akense Afasie Test. Note that a portion of the data was published in a previous validation study of the aDSWAL-QoL [124].

5.2.2 Measures

DSWAL-QoL and aDSWAL-QoL

The DSWAL-QoL is a condition-specific PRO scale that measures the effect of dysphagia on a person's HRQoL [48]. The DSWAL-QoL has been validated for a Flemish population [48] and consists of 44 items that are grouped into the following 11 subscales: General burden, Eating desire, Eating duration, Symptoms, Food selection, Communication, Fear of eating, Mental health, Social functioning, Sleep, and Fatigue. The DSWAL-QoL uses a 5-point Likert scale that ranges from 1='severely impaired quality of life' to 5='no impairment'. Based on Likert's method of summated ratings, the scores are transformed into subscale and scale scores that range from 0 = 'strong effect of dysphagia on HRQoL' to 100 = 'no effect on HRQoL' [62]. To increase the feasibility of the DSWAL-QoL for DysLC people, an adjusted version (aDSWAL-QoL) has been developed [124]. Both versions (DSWAL-QoL and aDSWAL-QoL) have been validated using CTT [48, 124]. The aDSWAL-QoL has similar content as the DSWAL-QoL (the abbreviated item contents of the DSWAL-QoL and aDSWAL-QoL are presented in Additional file 5.1) and also uses 5-point response categories. In contrast to the DSWAL-QoL, the number of different response formats in the aDSWAL-QoL is reduced to three (instead of six) and the subscales following the same response format are placed together. Additionally, the response categories in the aDSWAL-QoL are supported by visual line drawings, symbols and colors (Additional file 5.2).

Procedures

People completed both the DSWAL-QoL and aDSWAL-QoL in a random order to minimize recall effects. A minimum of 15 and a maximum of 30 minutes elapsed between the administration of the first and the second questionnaires. The people were encouraged to complete the scales as independently as possible, while assistance was provided when required (i.e., on request or when the patient failed to provide a response) [124].

The Rasch model

The Rasch model is based on a probabilistic Guttman pattern [81, 83]. This indicates that the probability of affirming a certain response to an item is a logistic function of the difference between the level of the measured construct as expressed by the person and as represented by the item and only a function of that difference [79, 81]. Person and item parameter estimates are placed on the same linear logit scale by transforming the original ordinal raw data into equal interval level measures (logits or log-odd units). The logit scale represents the latent trait [81] (i.e., dysphagia-related HRQoL) and both parameters are centered around a mean item location of zero [79]. Positive values for the person and item parameters indicate high ability levels (i.e., better HRQoL) and difficult items, and negative values indicate low ability levels (i.e., worse HRQoL) and easy items [192].

Data analysis

The DSWAL-QoL and aDSWAL-QoL include polytomous variables, and significant likelihood ratio tests ($p < 0.001$) indicated that the unrestricted parameterization of the model (partial credit) should be used rather than the rating scale model [193]. The item analysis with the Rasch model was performed using RUMM2030 [194], which integrates a pairwise conditional maximum likelihood algorithm in the estimation of the item and person parameters [191]. The following properties were examined: overall fit to the model, internal consistency reliability, unidimensionality, threshold ordering, individual item and person fits and differential item functioning (DIF), local item dependency (LID) and targeting. Item analysis with the Rasch model also yielded an iterative process in which strategies such as rescaling the

response categories and item reductions were applied to improve the model fit and the construction of the scale.

Overall fit to the model

The overall fit to the model was assessed by evaluating three overall fit statistics, specifically two item-person interaction statistics and one item-trait interaction statistic [195]. The overall item and person fit were evaluated by inspecting the mean item and mean person standardized fit residuals (FRs) [195], which should be close to zero with a standard deviation (SD) of '<1.4' [192]. The item-trait interaction that assesses whether the relative difficulties of the items remained constant across the different ability groups of patients [192, 195] was measured using a chi-square statistic (χ^2). Specifically, the χ^2 summarizes the differences between the observed and the expected values and was considered to be non-significant ($p > 0.05$) to fit the model expectations.

Reliability

The internal consistency reliability was assessed with the Person Separation Index (PSI), which is an estimate similar to Cronbach's alpha (α) coefficient [195]. The PSI assesses how adequately the set of items can distinguish subjects on different levels of the scale [196], and a value ≥ 0.70 is required [81, 192]. The PSI is also an indicator of the power of the generated fit statistics [129].

Unidimensionality

The unidimensionality of the scale was measured by performing t-tests on the two most divergent subsets of items [192]. The items with the greatest positive and negative loadings on their first residual factor (resulting from PCA) were used to create the two subsets [192, 197]. The scale was considered unidimensional if < 5% of the person estimates exhibited a significant difference in the scores for the two subtests [192, 197] or if the lower bound of an exact binomial confidence interval is < 5% [82]. Performing the *t*-test requires at least 12 category thresholds in each of the two subsets [82].

Threshold ordering of the polytomous items

After the investigation of the overall fit statistics, the ordering of the response categories was examined using a threshold map and category probability curves [79, 81]. In the cases of both the DSWAL-QoL and aDSWAL-QoL items, there are 5 response categories, resulting in 4 thresholds (= transitional points) [125]. Figure 5.1 represents a category probability curve in which the thresholds of a certain item are well ordered and form distinctive regions. Monotonicity was expected, and in cases of disordered thresholds, the item was rescored by combining adjacent categories [81, 192].

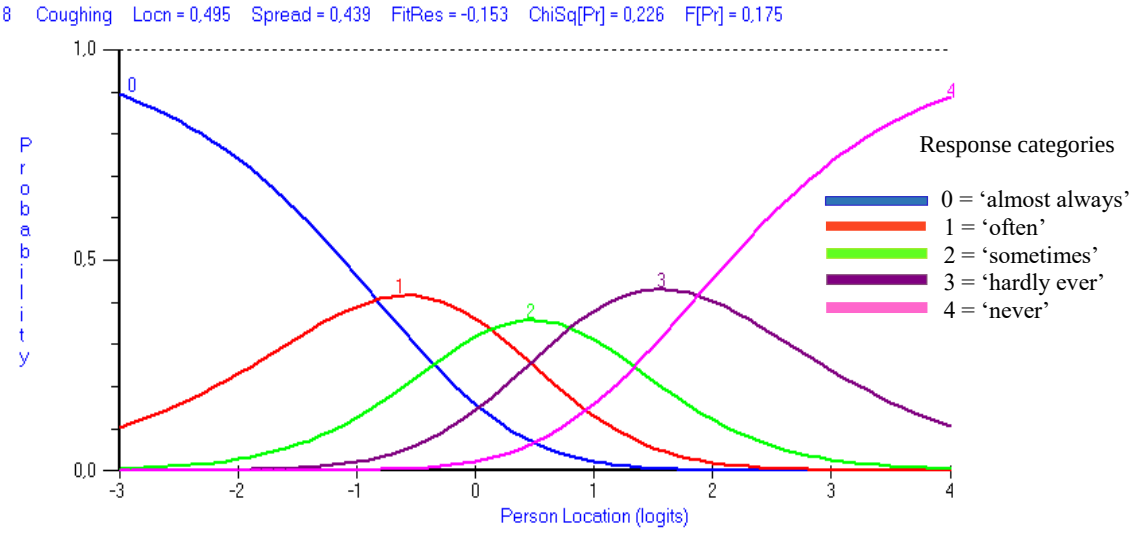


Figure 5.1 Category probability curve with ordered thresholds. Category probability curve displaying ordered thresholds for item 8 of the total DSWAL-QoL scale. This item has five response categories, resulting in 4 thresholds that increase in their location on the latent trait in a manner consistent with the increase in the underlying trait being measured. Note that the original response category structure of '1 to 5' was transformed into '0 to 4' by the Rumm software. Each response category (0, 1, 2, 3, 4) has a point (indicated by a peak in the curve) along the latent trait at the point of the most probable response.

Individual item and person fit

Individual item and person fit were assessed using the FR and χ^2. A person and item FR ± 2.5 and a χ^2 statistic above a Bonferroni-adjusted α-value of 0.05 [192, 198] indicated a fit to the model. Misfitting items or persons were removed to improve the overall fit of the model.

Differential item functioning

Items were also checked for DIF to ensure that the items of the scale were not biased by the person factors (i.e., language and cognitive impairments and dysphagia) and that the different class intervals followed the expected values of the characteristics of the items themselves. Thus, it was possible to investigate whether the different groups of the sample responded differently to an individual item despite the equal location on the latent trait [79]. The detection of DIF (i.e., uniform [81] and non-uniform DIF [79, 81, 197]) was made possible via the application of analysis of variance (ANOVA) to the fit residuals [79]. Uniform DIF was adjusted by splitting the item into group-specific items [79]. Items with non-uniform DIFs were considered to misfit the model and were removed [195].

Local item dependency

Local item dependency, which might be caused by response dependency (i.e., when a person's response to an item depends on the response to another item) or by trait dependency (i.e., multidimensionality), was investigated using the residual correlation matrix [191, 199]. Local item dependence was considered to be present if the item residual correlations were > 0.3 above the average of all of the item residual correlations [191, 198]. By grouping the items into one "super-item", called a testlet, the LID can be adjusted [197]. For all analyses, the Bonferroni correction was applied to adjust for multiple testing and was calculated based on the number of items [79].

Targeting of persons and items (person-item threshold distributions)

Targeting was examined after fitting the best solutions for the DSWAL-QoL and aDSWAL-QoL total scales and subscales. Targeting was analyzed by comparing the person and item threshold distributions.

To be acceptable, the mean person locations were expected to approximate the mean item threshold location (i.e., 0.0 logits) and the item locations were expected to cover approximately the same range of the logit scale as the person locations [79, 195].

Sample size

A sample size of 108 persons was suggested to provide 95% confidence that the item calibration or the estimated item difficulty will be within ± 0.5 logits [200].

5.3 RESULTS

5.3.1 Participants

In total, 108 persons were included. The Dys group consisted of 35 persons, 43 persons comprised the DysLC group, and 30 persons were in the LC group. The mean age of the total sample was 73.50 years (SD: 14.79). Table 5.1 presents the demographic characteristics of the persons. Comparison of the three groups revealed that head and neck cancer (48.6%) were most common in the Dys group, stroke (53.5%) was most common in the DysLC group, and dementia (60.0%) was most common in the LC group.

5.3.2 Evaluation of the measurement properties of the DSWAL-QoL and aDSWAL-QoL

Tables 5.2 and 5.4 display the results for the overall fit statistics before and after the implementation of the solutions suggested by the Rasch model for the DSWAL-QoL and aDSWAL-QoL scales, respectively. Table 5.3 provides an overview of the item level fit statistics of both scales.

Evaluation of the measurement properties of the DSWAL-QoL

The analysis revealed that the reliability was good, with a PSI of 0.93 and an excellent power of fit without extreme scores. However, the total DSWAL-QoL scale was found to misfit to the Rasch model as indicated by the item FR SD and person FR SD > 1.4, and by the presence of a significant item-trait interaction (Table 5.2). Multidimensionality was present as confirmed by the 16.09% statistically

significant different person estimates based on the two subsets of items. Disordered thresholds were found in 38 items, which indicated that the categorization of these items did not work as intended (Table 5.3). For example, the category probability curve for item 22 revealed that the estimates of the thresholds defining categories 2 and 3 did not form distinctive regions on the latent trait; therefore, these scores (i.e., 'somewhat' and 'a little') were at no time the most probable responses (Fig. 5.2). Seven items did not fit (items 5, 6, 7, 32, 34, 40 and 41; Table 5.3), and the individual person fit revealed that 11 persons fell outside the FR range of ± 2.5.

Table 5.2 Overall fit statistics for the DSWAL-QoL scale

Analysis		Item-person interaction		Item-trait interaction		Reliability	Unidimensionality	% ext
Scale	Initial scale (# items) Rescaled scale (suggestions by the Rasch model to improve model fit) (# items after adjustments)	Item FR Mean (SD)	Person FR Mean (SD)	χ^2 (df)	p	PSI +ext/÷ext	t-test %, (95% CI)	
Total scale	IS Total scale (44)	0.43 **(1.80)**	0.18 **(1.62)**	181.03 (44)	**<0.001**	0.93/0.93	***16.09 (10.9-19.2)***	0
	RS Total scale (rescore 41 items, delete 6 misfit items) (38)	0.03 (1.11)	–0.28 **(1.87)**	55.47 (38)	**0.032**	0.93/0.93	***15.24 (11.1-19.4)***	0.9
General burden	IS General burden (2)	0.33 (0.15)	–0.49 (1.0)	3.37 (2)	0.185	0.33/–0.61	NA	39.8
	RS General burden (rescore 2 items) (2)	0.70 (0.17)	–0.88 **(1.66)**	0.49 (2)	0.782	0.17/–2.64	NA	65.0
Eating duration	IS Eating duration (2)	0.39 (0.29)	–0.32 (1.05)	4.38 (2)	0.112	0.18/–1.32	NA	45.1
	RS Eating duration (rescore 2 items) (2)	0.96 (0.29)	**–1.59 (3.00)**	3.82 (2)	0.148	–0.10/–3.47	NA	52.9
Eating desire	IS Eating desire (3)	1.08 (0.72)	–0.10 **(1.46)**	8.74 (3)	**0.033**	0.13/–0.51	NA	30.8
	RS Eating desire (rescore 3 items) (3)	1.11 (0.56)	–0.73 **(2.72)**	8.85 (3)	**0.031**	0.28/–0.69	NA	30.8
Symptoms	IS Symptoms (14)	0.04 (0.94)	–0.17 (1.16)	13.66 (14)	0.475	0.83/0.84	***9.3 (4.9-13.4)***	6.6
	RS Symptoms (rescore 11 items) (14)	0.01 (1.05)	–0.21 (1.07)	18.91 (14)	0.168	0.85/0.84	***8.2 (3.8-12.6)***	6.6
Food selection	IS Food selection (2)	0.21 (0.04)	–0.39 (0.66)	2.96 (2)	0.227	0.46/0.20	NA	40.6
	RS Food selection (rescore 2 items) (2)	0.47 (0.21)	–0.70 (1.28)	0.77 (2)	0.682	0.54/0.25	NA	40.6
Communication	IS Communication (2)	0.41 (0.06)	–0.70 (1.06)	0.19 (2)	0.908	0.77/0.48	NA	35.6
Fear of eating	IS Fear of eating (4)	0.32 (0.96)	–0.30 (1.14)	16.90 (4)	**0.002**	0.55/0.41	NA	25.5
	RS Fear of eating (rescore 4 items, DIF-split 1 item) (5)	0.17 (0.92)	–0.45 (1.22)	8.05 (5)	0.153	0.54/0.38	NA	25.5
Mental health	IS Mental health (5)	0.20 (0.76)	–0.41 (1.35)	4.45 (5)	0.487	0.75/0.75	NA	30.5
	RS Mental health (rescore 1 item) (5)	0.26 (0.79)	–0.37 (1.23)	2.33 (5)	0.801	0.74/0.74	NA	30.5
Social functioning	IS Social functioning (5)	0.17 (0.64)	–0.41 (1.17)	5.09 (5)	0.405	0.72/0.59	NA	31.1
	RS Social functioning (rescore 5 items) (5)	0.40 (0.92)	–0.39 **(1.44)**	8.19 (5)	0.146	0.72/0.44	NA	36.8
Fatigue	IS Fatigue (3)	0.50 (0.11)	–0.52 (1.10)	2.30 (3)	0.513	0.69/0.56	NA	20.8
	RS Fatigue (rescore 3 items) (3)	0.49 (0.23)	–0.26 (0.88)	1.27 (3)	0.737	0.61/0.38	NA	30.2
Sleep	IS Sleep (2)	0.86 (0.06)	–0.54 (1.22)	2.13 (2)	0.345	0.27/–0.88	NA	43.4

Table 5.2 continued

Analysis		Item-person interaction		Item-trait interaction		Reliability	Unidimensionality	% ext
Scale	Initial scale (# items) Rescaled scale (suggestions by the Rasch model to improve model fit) (# items after adjustments)	Item FR Mean (SD)	Person FR Mean (SD)	χ^2 (df)	p	PSI +ext/÷ext	t-test %, (95% CI)	
	RS Sleep (rescore 2 items) (2)	0.47 (0.05)	**–1.63 (1.90)**	0.63 (2)	0.729	0.17/–2.40	NA	43.4
Satisfactory fit		0.00 (<1.40)	0.00 (<1.40)		>0.05	≥0.70	<5% or LCI ≤5%	

Abbreviations and symbols: *DSWAL-QoL* Dutch version of the Swallowing Quality-of-Life, *IS* initial scale, *RS* rescaled scale based on the suggestions from the Rasch methodology, *DIF* differential item functioning, *FR* fit residual, *SD* standard deviation, *PSI* Person separation index with extremes (+) and without extremes (÷), χ^2 *(df)* chi-square (degrees of freedom), *LCI* lower confidence interval, *% ext* percentage of extreme scores, *NA* not applicable (T-tests were not performed when there were too few thresholds in each subset or when the items were split for DIF when using the RUMM software). Bold indicates misfit to the Rasch model. Bold italic indicates multidimensionality.

Table 5.3 Item analysis: Overview of the item-level fit statistics of the DSWAL-QoL and aDSWAL-QoL

Scale (# items)	Disordered thresholds		Misfitting items		DIF		LID	
	DSWAL-QoL Item No	aDSWAL-QoL Item No	DSWAL-QoL Item No	aDSWAL-QoL Item No	DSWAL-QoL Item No	aDSWAL-QoL Item No	DSWAL-QoL Item No	aDSWAL-QoL Item No
Total scale (44)	1–7,9,11–15,17, 19,22–44	1–7,10,11,13,14, 17–41,43	5: FR= 6.1[a] 6: FR= 6.2[a] 7: FR= 4.1[a] 32: FR= –2.3[a] 34: FR= –2.0[a] 40: FR= 3.4[a] 41: FR= 2.6	2: FR= –2.8[a] 5: FR= 2.7[a] 6: FR= 5.9[a] 25: FR= 3.2 29: FR= 5.4[a] 32: FR= –2.7[a] 40: FR= 3.5[a] 41: FR= 3.2	5[b]	1,2,6,9,32,43[b]	clusters within and across subscales	
General burden (2)	1,2							
Eating duration (2)	3,4	3				3[b]		
Eating desire (3)	5,6,7	5,6,7						
Symptoms (14)	9,11–19,20	9,11–15,17,19,20					8/9	9/10, 19/20
Food selection (2)	22,23	22,23						
Communication (2)		24						
Fear of eating (4)	26,27,28,29	26,27,28,29		29: FR=3.2[a]	26[b]	26[b]		26/28
Mental health (5)	30	30,33,34				31[c]		
Social functioning (5)	35,36,37,38,39	35,36,37,38,39						
Sleep (2)	40,41	40						
Fatigue (3)	42,43,44							

Abbreviations: *DSWAL-QoL* Dutch version of the Swallowing Quality-of-Life, *aDSWAL-QoL* adjusted DSWAL-QoL, *No* number, *DIF* differential item functioning, *LID* local item dependence, *FR* fit residuals

[a]Indicates significant chi-square at $p < 0.001$.

[b]Uniform DIF

[c]Non-uniform DIF

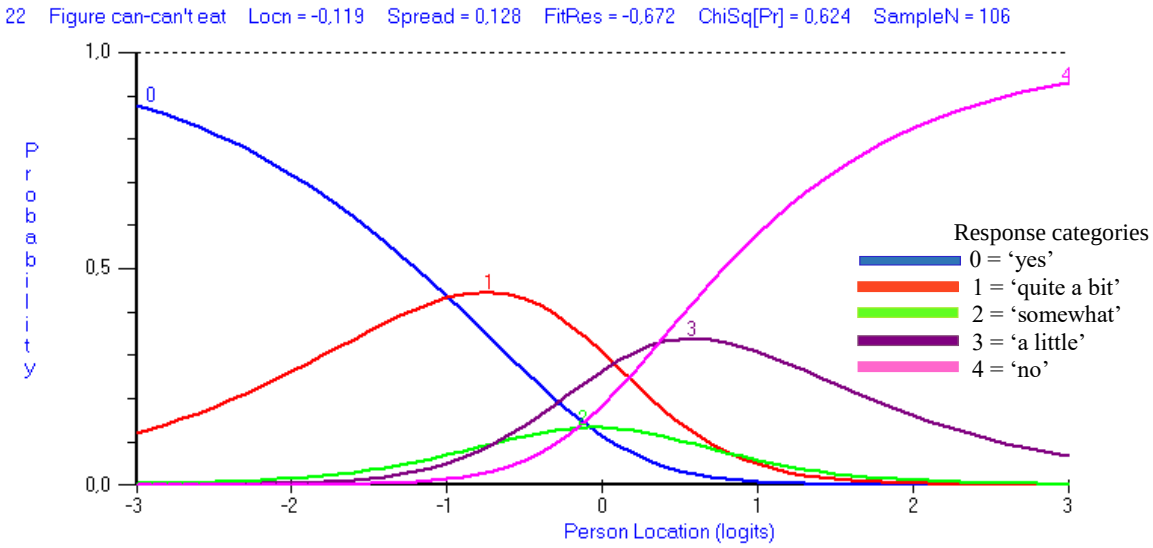


Figure 5.2 Category probability curve with disordered thresholds. Category probability curve graphically highlighting the disordered thresholds for item 22 of the total DSWAL-QoL scale. The point at which the lines for the adjacent response categories intersect in item 22 indicates that the transition between categories 2 and 3 is lower on the trait than the transition between categories 0 and 1. Response categories 2 and 3 never have a point on the continuum at which the most probable response is located.

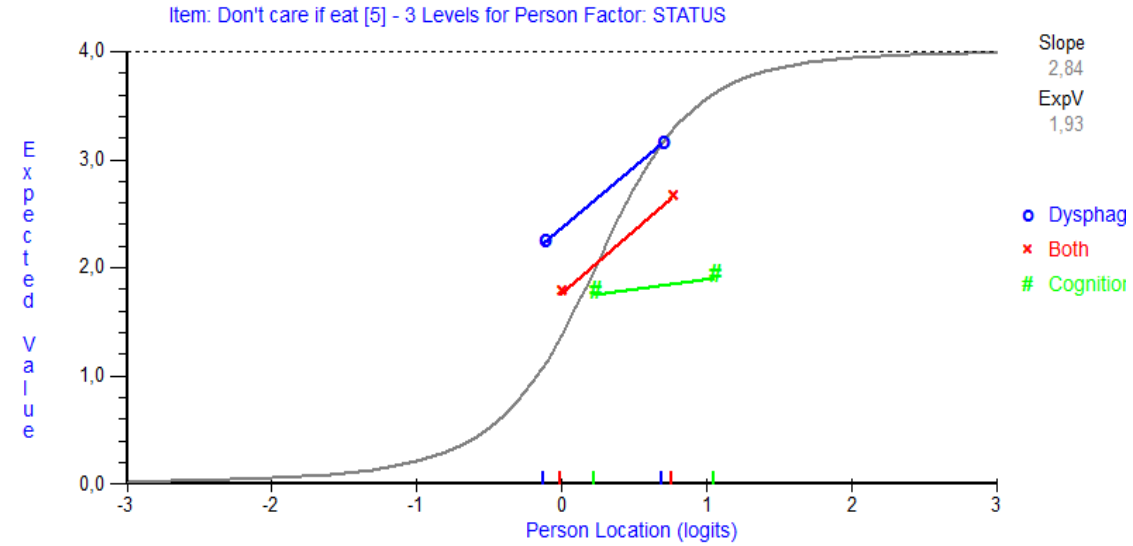


Figure 5.3 Item characteristic curve displaying a uniform DIF. Item characteristic curve displaying a uniform DIF for item 5. Despite the equal ability level, the three groups responded differently. The cognitive group obtained a prominently lower score than those of the two other groups.

As illustrated in Figure 5.3, item 5 exhibited a uniform DIF by group for all three groups, and the cognitive group obtained a prominently lower score compared with the those of persons in the other two groups given an equal ability level. Residual correlations > 0.3 were found for clusters of items within

and across subscales; thus, LID was present. To achieve a satisfactory overall model fit, it was necessary to rescore 38 items. For example, scores 2 and 3 for item 22 were collapsed into the score '1'; therefore, the Rasch-suggested scoring solution revealed a three-point response category, i.e., '0, 1 and 2', for this item (Additional file 5.1). During this process of rescoring items, three additional items exhibited disordered thresholds were also rescored. It was also necessary to delete six items (items 5, 6, 7, 34, 40 and 41; Table 5.2). The overall model fit improved for the items and no further DIF was present. However, the overall fit worsened for the person FR SD, the item-trait interaction remained significant, and 14 persons did not fit. Due to the limited sample size, the misfit persons were not removed. Multidimensionality remained despite the adjustments, and LID could not be resolved due to its unclear pattern.

For the Symptoms and Mental health subscales, the reliabilities were acceptable (PSI $\geq$ 0.75; Table 5.2). For the other subscales, the reliabilities were below the recommended level and the power of fit was low. A large number of extreme scores were present for all subscales with the exception of the Symptoms scale. There was no pattern of the extremes across the three person groups. All of the DSWAL-QoL subscales, with the exception of the Eating desire and Fear of eating subscales, exhibited satisfactory overall fit statistics. For the Eating desire subscale, the person FR mean (SD) of –0.10 (1.46) indicated some misfit of the persons. Inspection of the individual person fits revealed five misfit persons. The Symptom subscale exhibited a lack of unidimensionality, whereas the other subscales could not be subjected to t-tests due to insufficient numbers of items (i.e., thresholds). None of the items exhibited misfit, but disordered thresholds were found for a majority of the items within all subscales with the exception of the Communication subscale (Table 5.3). A uniform DIF was identified for item 26 from the Fear of eating subscale and was biased toward the cognitive group, which obtained higher scores. Local item dependency was demonstrated between items 8 and 9 from the Symptoms subscale. Adjustments of the subscales were performed, and after the items with disordered thresholds were rescored and item 26, which exhibited DIF, was split, all of the items exhibited ordered thresholds, the LID disappeared, and the item-trait interaction improved for the Fear of eating subscale (Table 5.2). However, the overall person FR mean values and/or SDs significantly increased for some of the subscales (i.e., General burden, Eating duration, Eating desire, Social functioning and Sleep), and the item-trait interaction remained significant for the Eating desire subscale. For the General burden, Eating duration, Eating desire, Social functioning and Sleep subscales, the numbers of misfit persons were N = 4, N = 14, N = 11, N = 7, and N = 33,

respectively. Unidimensionality could not be established for the Symptom subscale, and improvement for reliability also could not be identified for the subscales. After the solutions, the numbers of extreme scores increased for the General burden, Eating duration, Social functioning and Fatigue subscales (Table 5.2). The misfit persons were not removed because of the limited sample size.

Evaluation of the measurement properties for the aDSWAL-QoL

The analysis of the total aDSWAL-QoL scale revealed that the reliability was good (PSI = 0.93), the power of fit was excellent, and there were no extreme scores. Nonetheless, the total aDSWAL-QoL scale significantly deviated from the Rasch model (Table 5.4). Approximately 18% of the person estimates on the two most divergent subsets of items were significantly different, which indicated multidimensionality. Individual item analysis revealed 37 items with disordered thresholds (Table 5.3). Eight items (items 2, 5, 6, 25, 29, 32, 40 and 41) exhibited individual item misfit, and six items (1, 2, 6, 9, 32 and 43) exhibited uniform DIFs (Table 5.3). Most of the items with DIF were biased toward the LC group, which obtained higher scores, with the exception of item 43. Individual person misfits were found for 16 persons. Local item dependency was present between several item pairs within and across the subscales. After the rescoring of 37 items and six more items that exhibited disordered thresholds during the iterative process, the individual item fits improved for items 2 and 25 and the DIFs disappeared for items 6, 9 and 43. The other misfit items and the items that exhibited DIF remained. The iterative process revealed that to improve the overall fit, it was necessary to remove items 5, 6, 7, 29, 32, 40 and 41 and to split five items that displayed uniform DIFs (items 1, 2, 27, 43 and 44; Table 5.4). The person FR still indicated some misfit among the persons (SD > 1.4). The misfit persons (N = 19) were not removed because of the relatively limited sample size. Assessing the unidimensionality of the scale was no longer possible because of the item split. Since the LID showed an unclear pattern, it was not possible to create testlets.

For the Symptoms and Mental health subscales, the reliabilities were acceptable (PSI ≥ 0.72; Table 5.4). For the other subscales, the reliabilities were below the recommended level and the power of fit was low. Extreme persons were identified in all subscales, although the magnitudes were lowest for the Eating desire, Symptoms and Fear of eating subscales. Again, there was no pattern of extremes across the three population groups. The overall fits to the model were demonstrated for all of the aDSWAL-QoL subscales with the exception of the Communication, Fear of eating and Social functioning subscales

(Table 5.4). The overall person FR indicated a misfit for the Communication subscale (SD = 1.57), and further analysis revealed 22 misfit persons. Multidimensionality was present for the Symptom subscale; however, the other subscales could not be subjected to the test for multidimensionality because there were too few items. Disordered thresholds were found for 29 items across all subscales with the exception of the items in the General burden and Fatigue subscales (Table 5.3). At the individual item level, item 29 of the Fear of eating subscale did not fit the model. Differential item functioning was found in 3 items (items 3, 26, and 31), and LID was found between two item pairs from the Symptoms subscale (9–10, 19–20) and between item pairs 26–28 from the Fear of eating subscale. After rescoring all of the items with disordered thresholds (the DIFs for items 26 and 31 disappeared after rescaling), removing the misfit item 29 and splitting item 3 for DIF, ordered thresholds were found for all items and the item-trait interactions improved for the Fear of eating and the Social functioning subscales (Table 5.4). The item FR SD increased for the Communication subscale and the item-trait interaction became significant for the Eating desire subscale ($p < 0.05$). The person FR SDs increased for the Communication and Social functioning subscales. The numbers of misfit persons were $N = 21$ and $N = 9$ for the Communication and Social functioning subscales, respectively, and misfit persons were not removed because of the relatively limited sample size. The number of extreme scores remained unchanged or increased. Improvement for reliability could not be identified for the subscales. The LID disappeared between items 26 and 28 from the Fear of eating subscale. For the Symptom subscale, the lack of unidimensionality remained and the LID persisted between items 9 and 10, disappeared between items 19 and 20 and appeared between items 14 and 18. Adjusting the LID (i.e., creating two testlets: item pair 9–10 and item pair 14–18) did not improve the overall fit statistics for the Symptom subscale (item FR (SD) = –0.10 (0.64); person FR (SD) = –0.32 (1.16); item-trait interaction: χ^2 (df) = 12.16 (12); $p = 0.433$); however, the reliability increased (PSI = 0.98).

Targeting of persons and items

After the adjustments, targeting was suboptimal for both the DSWAL-QoL and aDSWAL-QoL total scales and subscales (Table 5.5). For both the total scales and subscales, the item locations did not cover the same ranges of the logit scale as the person locations. At the positive and negative ends of the trait, no item thresholds were found at the person locations, which indicated that these persons exhibited higher or

lower ability levels that could not be measured by the items of the scale. For the total aDSWAL-QoL scale and the aDSWAL-QoL Symptoms subscale, at the negative end of the trait, no persons were located at the item thresholds, which indicated that the average item difficulties of some of the items were too low.

Table 5.4 Overall fit statistics for the aDSWAL-QoL scale

Analysis		Item-person interaction		Item-trait interaction		Reliability	Unidimensionality	% ext
Scale	Initial scale (#items) Rescaled scale (suggestions by the Rasch model to improve model fit) (# items after adjustments)	Item FR Mean (SD)	Person FR Mean (SD)	χ^2 (df)	p	PSI +ext/÷ ext	t-test %, (95% CI)	
Total scale	IS Total scale (44)	0.19 **(1.95)**	–0.01 **(1.64)**	302.65 (44)	**<0.001**	0.93/0.93	***17.59 (13.5-21.7)***	0
	RS Total scale (rescore 43 items, delete 7 misfit items, DIF-split 5 items) (42)	–0.19 (0.95)	–0.46 **(1.86)**	50.96 (42)	0.137	0.92/0.92	NA	0
General burden	IS General burden (2)	0.17 (0.10)	–0.57 (0.86)	1.58 (2)	0.453	0.69/0.51	NA	38.9
Eating duration	IS Eating duration (2)	–0.12 (0.32)	–0.29 (0.62)	3.95 (2)	0.139	0.70/0.42	NA	24.5
	RS Eating duration (rescore 1 item + DIF-split 1 item) (3)	0.06 (0.16)	–0.29 (0.60)	2.70 (3)	0.440	0.68/0.37	NA	24.5
Eating desire	IS Eating desire (3)	0.32 (0.51)	–0.24 (0.80)	6.17 (3)	0.104	0.61/0.40	NA	2.8
	RS Eating desire (rescore 3 items) (3)	0.37 (0.35)	–0.22 (0.73)	10.75 (3)	**0.013**	0.57/0.32	NA	10.2
Symptoms	IS Symptoms (14)	–0.05 (0.88)	–0.29 (1.25)	10.23 (14)	0.741	0.87/0.86	***10.8 (6.6-15.0)***	2.8
	RS Symptoms (rescore 9 items) (14)	–0.08 (0.64)	–0.29 (1.17)	12.02 (14)	0.605	0.86/0.85	***11.8 (7.5-16.0)***	2.8
Food selection	IS Food selection (2)	0.74 (0.77)	–0.10 (0.67)	5.21 (2)	0.074	–0.44/–1.31	NA	29.9
	RS Food selection (rescore 2 items) (2)	0.82 (0.78)	–0.10 (0.55)	1.11 (2)	0.574	–0.39/–1.33	NA	29.9
Communication	IS Communication (2)	0.25 (0.24)	–1.09 **(1.57)**	2.44 (2)	0.295	0.46/0.18	NA	18.1
	RS Communication (rescore 1 item) (2)	0.39 **(1.47)**	–1.14 **(1.89)**	1.12 (2)	0.570	0.43/–0.03	NA	27.6
Fear of eating	IS Fear of eating (4)	0.82 **(1.65)**	–0.36 (1.37)	57.75 (4)	**<0.001**	0.26/0.23	NA	3.8
	RS Fear of eating (rescore 2 items + delete 1 misfit item) (3)	0.66 (0.58)	–0.19 (1.06)	7.01 (3)	0.072	0.15/–0.31	NA	38.6
Mental health	IS Mental health (5)	–0.02 (0.83)	–0.48 (1.20)	4.46 (5)	0.486	0.71/0.72	NA	22.6
	RS Mental health (rescore 4 items) (5)	0.08 (0.91)	–0.41 (1.02)	6.66 (5)	0.247	0.73/0.72	NA	24.5
Social functioning	IS Social functioning (5)	0.08 (0.70)	–0.24 (0.96)	11.93 (5)	**0.036**	0.50/0.39	NA	31.8
	RS Social functioning (rescore 5 items) (5)	0.35 (0.81)	–0.49 **(1.49)**	3.801(5)	0.578	0.62/0.46	NA	31.8
Fatigue	IS Fatigue (3)	0.12 (0.66)	–0.49 (1.03)	1.64 (3)	0.651	0.70/0.60	NA	14.3
Sleep	IS Sleep (2)	0.37 (0.10)	–0.38 (0.82)	3.15 (2)	0.207	0.43/0.03	NA	29.5
	RS Sleep (rescore 1 item) (2)	0.13 (1.40)	–0.29 (0.72)	2.50 (2)	0.287	0.39/–0.06	NA	35.2
Satisfactory fit		0.00 (<1.40)	0.00 (<1.40)		>0.05	≥0.70	<5% or LCI ≤5%	

Abbreviations and symbols: *aDSWAL-QoL* adjusted DSWAL-QoL, *IS* initial scale, *RS* rescaled scale based on the suggestions from the Rasch methodology, *DIF* differential item functioning, *FR* fit residual, *SD* standard deviation, *PSI* Person separation index reported with extremes (+) and without extremes (÷), *χ^2 (df)* chi-square (degrees of freedom), *LCI* lower confidence interval, *% ext* percentage of extreme scores, *NA* not applicable (T-tests were not performed when there were too few thresholds in each subset or when the items were split for DIF when using the RUMM software). Bold indicates misfit to the Rasch model. Bold italic indicates multidimensionality.

Table 5.5 Targeting of the DSWAL-QoL and aDSWAL-QoL total scales and subscales after fitting solutions

	DSWAL-QoL				aDSWAL-QoL			
	Item location		Person location		Item location		Person location	
	Mean (SD)	Range	Mean (SD)	Range	Mean (SD)	Range	Mean (SD)	Range
Total scale	0.0 (0.76)	–2.06; 1.20	0.89 (1.30)	–3.43; 5.36	0.0 (0.80)	–1.98; 2.05	1.05 (1.35)	–1.68; 5.98
General burden	0.0 (0.28)	–0.19; 0.19	–0.43 (1.52)	–1.92; 2.00	0.0 (0.28)	–0.15; 0.15	–0.43 (1.52)	–3.30; 3.01
Eating duration	0.0 (0.01)	–0.01; 0.01	–0.30 (1.20)	–1.65; 1.65	0.0 (0.72)	–0.61; 0.79	–0. 80 (2.16)	–4.97; 2.59
Eating desire	0.0 (1.15)	–0.13; 0.16	0.15 (1.16)	–1.89; 1.89	0.0 (1.25)	–1.09; 1.37	0.53 (1.54)	–3.52; 3.78
Symptoms	0.0 (1.63)	–1.19; 0.79	1.32 (1.48)	–1.98; 4.60	0.0 (0.70)	–1.52; 0.96	1.07 (1.25)	–1.15; 4.63
Food selection	0.0 (1.22)	–0.16; 0.16	1.35 (2.00)	–2.92; 3.48	0.0 (0.87)	–0.61; 0.61	1.14 (1.67)	–2.85; 2.58
Communication	0.0 (1.34)	–0.24; 2.40	0.26 (2.56)	–4.04; 3.81	0.0 (0.04)	–0.03; 0.03	0.36 (1.41)	–2.37; 2.32
Fear of eating	0.0 (0.80)	–1.30; 0.27	1.42 (1.72)	–3.12; 3.92	0.0 (0.48)	–0.56; 0.28	0.98 (1.20)	–2.51; 2.19
Mental health	0.0 (0.28)	–0.24; 0.44	0.97 (1.63)	–3.00; 2.94	0.0 (0.52)	–0.72; 0.47	1.58 (1.77)	–3.41; 3.52
Social functioning	0.0 (0.47)	–0.80; 0.42	0.08 (1.89)	–2.89; 2.77	0.0 (0.53)	–0.85; 0.58	1.08 (1.61)	–3.09; 3.01
Fatigue	0.0 (0.37)	–0.41; 0.30	0.54 (1.78)	–3.00; 2.78	0.0 (0.32)	–0.26; 0.35	0.59 (1.45)	–3.14; 3.04
Sleep	0.0 (0.08)	–0.06; 0.06	0.33 (1.46)	–2.17; 2.17	0.0 (0.62)	–0.44; 0.44	0.48 (1.38)	–1.80; 2.27

Abbreviations: *DSWAL-QoL* Dutch version of the Swallowing Quality-of-Life, *aDSWAL-QoL* adjusted DSWAL-QoL, *SD* standard deviation

5.4 DISCUSSION

The analysis did not support the structural validity or objectivity of either the DSWAL-QoL or the aDSWAL-QoL total scales and subscales or the statistical sufficiency of the total scores and subscale scores. Misfit to the Rasch model, multidimensionality and/or the presence of DIF were found. Comparing the subscales of both versions, the Eating desire subscale of the aDSWAL-QoL exhibited an overall fit to the model in contrast to its corresponding subscale in the DSWAL-QoL, while the Communication and Social functioning subscales in the DSWAL-QoL did fit the model. For all other subscales, the results for the overall fits were similar. A large number of extreme scores were present in both versions of the scale, and this phenomenon was even greater for the DSWAL-QoL subscales. These extreme scores influenced the reliability and the power of fit. The presence of low levels of PSI and the high percentage of extreme scores reflected suboptimal targeting for the subscales of both versions. The suboptimal targeting for the total scales and subscales resulted in decreased estimation precisions of the item and person parameters [79]. The misfit items were most present when all of the items were treated as one total scale and were quite similar in both versions. Local item dependence was present between items within and across the subscales for both the DSWAL-QoL and aDSWAL-QoL scales.

The main sources of misfit for both scale versions were disordered thresholds for the items in the total scale and the individual subscales with the exception of the items of the Communication subscale in the DSWAL-QoL and the General burden and Fatigue subscales in the aDSWAL-QoL. We expected fewer disordered thresholds in the aDSWAL-QoL because the aDSWAL-QoL has been proven to be more feasible for use in groups with additional language and/or cognitive impairments [124]. Note that the 5-point response category in the aDSWAL-QoL contains similar content as the DSWAL-QoL. Some patients might have interpreted the graphic support (i.e., the symbols that were intended to enhance the comprehension of the response categories) in a different manner than what was intended. Nonetheless, it was obvious that the original scoring structures for most of the items of both the total scales and subscales did not work as intended (Additional file 5.1). The latter may be because the people were not able to discriminate between the response categories. Either the different categories were not well defined or the difference in meaning was too subtle (i.e., what is the difference between ‘somewhat’ and ‘a little’?). Additionally, the incorrect assumption that the Likert scale is an interval scale is common, although the categories of a 5-point Likert format represent a qualitative variable that is actually only sequential and ordinal [201]. To obtain linear, equal-interval level results, testing of the ordering of the response

categories against the Rasch measurement model and subsequent rescaling of the items with disordered thresholds is required.

The presence of DIF by group was found in both versions of the scale but was most prominent for the aDSWAL-QoL. Most of the items that displayed DIF were biased toward the cognitive group, which tended to obtain higher scores for these items. This finding indicates that this group overestimated their HRQoL. The latter was expected because this group did not suffer from oropharyngeal dysphagia. The DSWAL-QoL and aDSWAL-QoL are disease-specific scales developed for people with oropharyngeal dysphagia. The LC group did not meet this condition; thus, the appropriateness of including this group in the study could be questioned. The objectivity of the scale can only be established by the Rasch methodology if one important requirement is satisfied, i.e., if the items and the sample are within the specific frame of reference for which the scale was developed [80]. Nonetheless, including this patient group was important because it enabled the evaluation of whether the scale and subscales were influenced by DIF. It was also expected that this LC group would exhibit extreme scores (i.e., all of the maximum scores for HRQoL) because of the absence of oropharyngeal dysphagia. However, there was no pattern in the extreme scores across the three groups. This issue leads to the question of the extent to which the scales are completed in a 'reliable' manner (i.e., whether they truly capture the patient's perspective). Compared to the other two groups, more people in the cognitive group had an underlying etiology of dementia. The language and cognitive impairments were also greater in the LC group; thus, this group likely had more problems understanding the questions of both the DSWAL-QoL and aDSWAL-QoL. Next to impaired language functions, dementia encompasses a large spectrum of behavioral and other cognitive impairments, such as changes in personality and behavior, impaired reasoning and handling of complex tasks, poor decision-making ability and poor judgment [69]. The finding that this LC group did not demonstrate extreme scores as expected indicates that caution should be exercised in the use of these scales with dysphagic people with dementia because the dementia-related factors might influence the responses.

The use of the 5-point response category for the Social functioning subscale of the aDSWAL-QoL should be questioned because the middle response category of this subscale included 'I don't know'. The literature indicates that respondents do not interpret this type of middle response category as expected from the integer scoring (i.e., monotonically). Consequently, disordered thresholds can occur because

these categories differ from other response options in their probabilities of being selected. With respect to the integer scoring [202], it would be appropriate to reformulate this response category.

After the adjustments, a potential scoring structure was suggested for all items of both the DSWAL-QoL and aDSWAL-QOL total scales and subscales (Additional file 5.1). The scoring structure was often different for some of the items when they were treated as one scale instead of being part of the subscales. In most of the items, the 5-point response category was rescaled to a 3- or 4-point response format. A disadvantage of using different response formats is that it might lead to confusion and cause erroneous responses [203]. The analysis suggested that, rather than including 44 items, the total DSWAL-QoL scale should be rescaled to a 38-item scale and the total aDSWAL-QoL should be rescaled to a 42-item scale in which five items are group specific. The proposed numbers of items for the subscales of both versions are displayed in Additional file 5.1 and in Tables 5.2 and 5.4. Note that a large-scale empirical study is needed to confirm the scoring structures of both the scales and subscales.

The overall fit improved for the total aDSWAL-QoL scale but not for the total DSWAL-QoL after the adjustments, and the person FR SDs remained high for both total scales. Misfits were also demonstrated for the Eating desire subscale of both scales after adjusting the items. The issue of fit is a relative matter in the Rasch methodology because it depends on the sample size [79]. The reliabilities were high for both total scales but low for most of the subscales of both questionnaires. When comparing studies that performed cross-cultural adaptations of the SWAL-QoL [48, 57, 124, 185], we observed differences in reliability. These studies used Cronbach's α to evaluate the internal consistency. However, relying on Cronbach's α is only justified if the data are normally distributed. Multiple ceiling or floor effects were observed in those studies [48, 57, 124, 185], raising questions about the accuracy of the internal consistencies of these scales. It is not possible to compare our study results (based on the PSI) with studies that have used Cronbach's α because α includes extreme scores, whereas the estimate of the PSI requires extrapolated values for extreme scores [204]. Specifically, the calculation of Cronbach's α assumes equal standard errors (SEs) in all of the scores. This assumption contrasts with the calculation of the PSI in which the SE increases as the scores become more extreme [204]. After adjusting for LID in the Symptom subscale of the aDSWAL-QoL, the reliability increased, which indicated multidimensionality [199]. Whether LID is response or trait dependent may be difficult to distinguish in polytomous analysis [199]. For both the DSWAL-QoL and aDSWAL-QoL total scales, LID was present between and across subscales. Thus, due to an unclear pattern of the LID, it was not possible to create

testlets that would resolve LID. Since LID might be caused by multidimensionality, it could be suggested to reconsider the dimensional structure of both scales using factor analytic approaches. Although Vanderwegen et al. [48] performed traditional (linear) PCA on the DSWAL-QoL, PCA only identifies the variables that show the strongest linear relationship with each other and tries to explain for as much of the total variance in the data [144]. Therefore, factor analysis for ordinal data [205] is required as it identifies the (number of) latent constructs and the possible underlying factor structure of a set of variables [144]. Furthermore, multidimensionality could not be resolved by the Rasch methodology for either of the total scales. This finding indicates that each item of both scales should be scored separately and should be considered as a single item [191]. The main strength of this study was that by using a modern test theory approach, both scales could be improved (e.g., ordered thresholds for the items). However, we could not establish the structural validity and objectivity of either the total scales or the subscales and the total score and subscale scores remained statistically insufficient.

Limitations and future research

One major limitation of this study was the relatively limited sample size. A sample size of at least 64 to 144 persons is required to achieve 95% confidence that the item calibration is within ± 0.5 logits [200]. Our study sample of 108 patients was within the recommended sample size (i.e., sufficient for the total scales). Subjects with extreme scores were excluded from the analysis because they did not contain information for the estimation of the item and person threshold parameters [196]. Thus, for the subscales, the effective sample size in this study was smaller than the original sample size. The low PSI and the low power of fit had to be taken into account when interpreting the results for the subscales. To derive definitive conclusions about the items and the scales, well-targeted and sample sizes ≥ 250 people are recommended [196, 200]. Therefore, this study must be interpreted as a pilot study. Nonetheless, clinicians and researchers cannot longer rely on both the DSWAL-QoL and aDSWAL-QoL total scores and subscale scores as an indicator of how a patient's HRQoL is affected by oropharyngeal dysphagia. Until the psychometric properties have been established in a larger sample, we suggest to use the proposed scoring structure (Additional file 5.1) for each individual item, taking into account to derive only qualitative information from that item. Items suggested to be removed from the total scales and subscales should be interpreted with caution. The psychometric weaknesses of these scales indicate to reconsider the cross-cultural validation process [206] and to evaluate if the translations and adaptations

meet accepted standards of cross-cultural validation [58, 191]. We recommend using IRT for further validation of the original SWAL-QoL [63] and all of its translations. Most of the subscales exhibited a lack of sufficient items to allow for the assessment of the unidimensionality of the scale. After all, multiple items enable the improvement of the reliability because random errors of measurement can be averaged out. Multiple items increase the scope of a scale and are less open to variable interpretation [207]. Scales that are too extensive do not function well in routine clinical practice [63] because the patient's burden in completing multiple items can be onerous and time consuming [208]. With 44 items, both of the SWAL-QoL versions are still long and extensive scales. Therefore, it would be beneficial to create and validate a shorter version. We analyzed the two versions of the SWAL-Qol separately. For future studies, it may be useful to merge the two datasets and perform DIF analysis using the version as a person factor. We used a residual correlation of $r > 0.30$ above the average of all the correlations for the detection of LID [198], although this criterion might be regarded as arbitrary [209]. If a more strict criterion of $r > 0.20$ was used, we might have found more LID.

5.5 CONCLUSIONS

This is the first study to examine the structural validity and objectivity of both the DSWAL-QoL and aDSWAL-QoL total scales and subscales and the statistical sufficiency of the total scores and subscale scores using item analysis with the Rasch model. However, the analysis could not establish these psychometric properties because a misfit to the model, multidimensionality, disordered thresholds, DIF and/or LID were found. This analysis with the Rasch model identified areas that require further investigation. Our study highlighted the fact that relying on the DSWAL-QoL and aDSWAL-QoL subscale scores and total scale scores to make conclusions about a person's dysphagia-related HRQoL should be undertaken with caution before the psychometric requirements have been established. The adjustments suggested by the Rasch model induced scale improvement, as the disordered thresholds were rescaled, the misfit items were removed and the DIF was resolved. Although we were not able to derive definitive conclusions about the items and the scales, this study illustrated the added value of the use of Rasch analysis in the detection of the psychometric strengths and weaknesses of these rating scales. Therefore, this study can be viewed as an essential step forward toward the further improvement of these scales.

Additional file 5.1. Scoring structure after the adjustments as suggested by the Rasch model

			DSWAL-QoL		aDSWAL-QoL	
	Item content	Subscale	Total scale	Subscales	Total scale	Subscales
1	Difficult dealing	Burden	0,1,1,1,2	0,0,1,1,2	0,1,1,1,2[a]	**0,1,2,3,4**
2	Major distraction	Burden	0,1,1,1,2	0,0,1,1,2	0,1,1,1,2[a]	**0,1,2,3,4**
3	Longer time to eat	Eating duration	0,1,1,1,2	0,1,1,1,2	0,1,1,1,2	0,1,2,2,3[a]
4	Takes forever to eat	Eating duration	0,1,1,1,2	0,0,1,1,2	0,1,1,1,2	**0,1,2,3,4**
5	Don't care if I eat or not	Eating desire	~~0,0,0,1,1~~	0,1,1,1,2	~~0,1,1,1,2~~	0,1,1,1,2
6	Don't enjoy	Eating desire	~~0,0,0,1,1~~	0,1,1,1,2	~~0,1,1,1,2~~	0,1,1,2,3
7	Rarely hungry	Eating desire	~~0,0,0,1,1~~	0,1,1,1,2	~~0,1,1,1,2~~	0,1,1,2,3
8	Coughing	Symptoms	**0,1,2,3,4**	**0,1,2,3,4**	**0,1,2,3,4**	**0,1,2,3,4**
9	Choking on food	Symptoms	0,1,1,1,2	0,1,1,1,2	0,0,1,1,2	0,1,1,2,3
10	Choking on liquids	Symptoms	**0,1,2,3,4**	**0,1,2,3,4**	0,0,1,1,2	**0,1,2,3,4**
11	Thick saliva, phlegm	Symptoms	0,1,1,1,2	0,1,2,2,3	0,1,1,1,2	0,0,1,2,2
12	Gagging	Symptoms	0,1,1,1,2	0,1,2,2,3	0,0,0,1,2	**0,1,2,3,4**
13	Excess saliva, phlegm	Symptoms	0,1,1,1,2	0,1,1,1,2	0,1,1,2,2	0,1,2,2,3
14	Clear throat	Symptoms	0,1,1,1,2	0,1,1,1,2	0,1,1,1,2	0,1,1,1,2
15	Drooling	Symptoms	0,1,1,1,2	0,1,1,1,2	0,1,1,1,2	**0,1,2,3,4**
16	Problem chewing	Symptoms	0,1,1,1,2	0,1,2,2,3	0,1,2,2,3	**0,1,2,3,4**
17	Food stick throat	Symptoms	0,1,1,1,2	0,1,2,2,3	0,1,1,1,2	0,1,2,2,3
18	Food stick mouth	Symptoms	**0,1,2,3,4**	0,1,1,1,2	0,1,1,1,2	0,1,1,2,3
19	Food/liquid dribble from mouth	Symptoms	0,1,1,1,2	0,1,1,1,2	0,1,1,1,2	0,1,1,1,2
20	Food/liquid dribble from nose	Symptoms	0,1,1,1,2	0,1,2,2,3,	0,0,1,1,2	0,1,2,2,3
21	Cough out of mouth when food stuck	Symptoms	0,1,1,1,2	**0,1,2,3,4**	0,1,1,1,2	0,0,1,1,2
22	Figure can - can't eat	Food selection	0,1,1,1,2	0,1,2,2,3	0,1,1,1,2	0,1,1,1,2
23	Difficult dealing	Food selection	0,1,1,1,2	0,1,2,2,3	0,1,1,1,2	0,1,1,1,2
24	Hard understand me	Communication	0,1,1,1,2	**0,1,2,3,4**	0,1,1,1,2	0,0,1,2,2
25	Hard speaking clear	Communication	0,1,1,1,2	**0,1,2,3,4**	0,1,1,1,2	**0,1,2,3,4**
26	Afraid choking foods	Fear	0,1,1,1,2	0,1,1,1,2[a]	0,1,1,1,2	0,1,1,2,2
27	Afraid pneumonia	Fear	0,1,1,1,2	0,1,1,1,2	0,1,1,1,2[a]	**0,1,2,3,4**
28	Afraid choking liquids	Fear	0,1,1,1,2	0,1,1,1,2	0,1,1,1,2	0,0,1,2,2
29	Never know when choke	Fear	0,1,1,1,2	0,1,1,1,2	~~0,1,1,1,2~~	~~0,0,1,2,2~~
30	Depressed	Mental health	0,1,1,1,2	0,0,1,1,2	0,1,1,1,2	0,1,1,2,2
31	Impatient dealing	Mental health	0,1,1,1,2	**0,1,2,3,4**	0,1,1,1,2	0,0,1,1,2
32	So careful annoy	Mental health	0,1,1,1,2	**0,1,2,3,4**	~~0,1,1,1,2~~	**0,1,2,3,4**
33	Frustrated	Mental health	0,1,1,1,2	**0,1,2,3,4**	0,1,1,1,2	0,1,1,2,3
34	Discouraged	Mental health	~~0,1,1,1,2~~	**0,1,2,3,4**	0,1,1,1,2	0,0,1,1,2
35	Do not go out	Social functioning	0,1,1,1,2	0,1,1,1,2	0,1,1,1,2	0,1,1,1,2
36	Hard social life	Social functioning	0,1,1,1,2	0,1,1,1,2	0,1,1,1,2	0,1,1,1,2
37	Change work activity	Social functioning	0,1,1,1,2	0,1,1,1,2	0,1,1,1,2	0,1,1,1,2
38	Dislike social gathering	Social functioning	0,1,1,1,2	0,1,1,1,2	0,1,1,1,2	0,1,1,1,2
39	Role change	Social functioning	0,1,1,1,2	0,1,1,1,2	0,1,1,1,2	0,1,1,1,2
40	Feel exhausted	Fatigue	~~0,1,1,1,2~~	0,0,1,2,2	~~0,1,1,1,2~~	**0,1,2,3,4**
41	Feel weak	Fatigue	~~0,1,1,1,2~~	0,0,1,2,2	~~0,1,1,1,2~~	**0,1,2,3,4**
42	Feel tired	Fatigue	0,1,1,1,2	0,0,1,2,2	0,1,1,1,2	**0,1,2,3,4**
43	Trouble falling asleep	Sleep	0,1,1,1,2	0,1,1,1,2	0,1,1,1,2[a]	0,0,1,1,2
44	Trouble staying asleep	Sleep	0,1,1,1,2	0,1,1,1,2	0,1,1,1,2[a]	**0,1,2,3,4**

Note: The score range from 1 to 5 is transformed into the range of 0 to 4 in the RUMM2030. Bold indicates that the original scoring structure worked as intended. Items crossed out indicate that these items were removed from the scale.

[a] items were split for DIF.

Additional file 5.2 Representation of the 5-point response categories for the different subscales of the aDSWAL-QoL. These pictures demonstrate how the 5-point response categories for the different subscales are presented in the aDSWAL-QoL with respect to the real size. Note that the translation of the response format into English can slightly deviate from the formulation in the Flemish language.

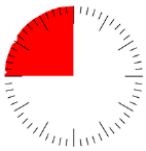
The General burden, Eating duration, Eating desire and Food selection subscales used the following response format:

√√	√	√X	X	X X
yes	quite a bit	somewhat	a little	no
				

The Social functioning subscale used the following response format:

√√	√	?	X	X X
yes	quite a bit	don’t know	a little	no
				

The Symptoms, Communication, Fear of eating, Mental health, Fatigue and Sleep subscales used the following response format:

 almost always	 often	 sometimes	 hardly ever	 never
				

CHAPTER VI

PATIENT-REPORTED AND PROXY-REPORTED OUTCOME MEASURES FOR THE ASSESSMENT OF HEALTH-RELATED QUALITY OF LIFE AMONG PATIENTS RECEIVING ENTERAL FEEDING: A SYSTEMATIC REVIEW PROTOCOL

Study VI has been published in:

Simpelaere I, White A, Bekkering GE, Geurden B, Van Nuffelen G, De Bodt M. Patient-reported and proxy-reported outcome measures for the assessment of health-related quality of life among patients receiving enteral feeding: a systematic review protocol. JBI Database of Systematic Reviews and Implementation Reports. 2016;14(7):45-75. doi:10.11124/JBISRIR-2016-002982.

ABSTRACT

Review question/objective: The objective is to systematically review the psychometric properties and the clinical utility of patient-reported outcome measures (PROMs) and proxy-reported outcome measures that assess health-related quality of life (HRQoL) among patients receiving enteral feeding to make recommendations for use in clinical practice and research. The purpose of this systematic review is to evaluate the psychometric properties and the clinical utility of:

- Measures to assess HRQoL among patients receiving enteral feeding, regardless the cause of receiving enteral feeding.
- Measures to assess HRQoL in the subgroup "patients receiving enteral feeding as a consequence of dysphagia".
- Measures (1 and 2) that are completed by patients' proxy to assess HRQoL among patients receiving enteral feeding.

The research question is: What are the psychometric properties and the clinical utility of these measures? We will summarize evidence on the following properties: validity (content validity, criterion-related validity, construct validity, floor and ceiling effects), reliability (reproducibility and internal consistency) and responsiveness and clinical utility (interpretability and feasibility to complete the PROM and the proxy-reported outcome measure).

Keywords: Enteral feeding, Patient-reported outcome measures, Quality of life, Validity

6.1 INTRODUCTION

Patient-reported outcome measures (PROMs) can be used to assess patients' health-related quality of life (HRQoL) from the patients' perspective [210]. The Food and Drug Administration of the US Department of Health and Human Services defines PROMs as "any report of the status of a patient's health condition that comes directly from the patient, without interpretation of the patient's response by a clinician or anyone else" [53 (p.2)]. Therefore, PROMs can be used to measure patients' own perceptions of their health status [210, 211], HRQoL [210, 211] and individual treatment effects [48]. However, self-reporting can be problematic or impossible for some patients [212]. In these situations, proxy-reported outcome measures may be used for the assessment of the patient's HRQoL as the proxies can provide relevant information about how they think the patient would report on his or her HRQoL. Patient-reported and proxy-reported outcome measures are also used in the context of enteral feeding [213, 214]. Enteral feeding is an alternative feeding strategy to be used in patients with insufficient oral food intake [215], such as patients with severe dysphagia and other conditions [216-218]. Although patients may benefit from this feeding strategy in terms of nutritional status [44, 216], the impact on patients' HRQoL is equally important [44]. Therefore, PROMs or proxy-reported (i.e. when self-reporting is complicated) outcome measures should be evaluated in addition to conventional clinical assessments and instrumental investigations [49]. As a means to select the most suitable PROM and proxy-reported outcome measure for assessing HRQoL among patients receiving enteral feeding, it is important to evaluate the psychometric properties and the clinical utility of these measurement instruments [58].

Quality of Life (QoL) is defined by the World Health Organization as follows: "individuals' perceptions of their position in life in the context of the culture and value systems in which they live and in relation to their goals, expectations, standards and concerns. It is a broad-ranging concept incorporating in a complex way the person's physical health, psychological state, level of independence, social relationships, personal beliefs and their relationships to salient features of the environment" [38 (p.3)]. There is still a lack of consensus among researchers about the definition of QoL that can explain the different constructs or concepts that are selected for their measurement instruments [39, 40]. Quality of life varies between populations, time periods and diseases [41-43]. Also, the concept is subjective as it is based on a perception by the individual around his or her wellbeing [44]. Individuals from the same environment seem to have a similar conception about QoL (e.g. "reaching happiness" and "satisfaction in life" are important constructs of QoL in individuals living in the Western countries) [39]. Health-related

quality of life is distinct from QoL as it is directly associated to a health state of an individual [39, 45]. Various definitions of HRQoL exist [42, 46]. However, agreement is reached in that HRQoL is a multidimensional concept (i.e. consists of physical, psychological and social domains of health), it is based on the individual's perception of his or her wellbeing, encompasses the current health situation and the future perspective and is not restricted to the disease [39, 42].

Assessment of HRQoL is useful for measuring the outcome of disease and treatment effectiveness [47]. The most common way to measure HRQoL issues is by patient-reported outcome (PRO) instruments, using a standardized questionnaire or an interview [48, 49, 50-52]. Patient-reported outcome measures can be generic instruments, disease-specific instruments or battery of scales. Generic PROMs can be used in diverse patient populations and are suitable for comparing HRQoL and the efficiency of different treatment strategies across these populations [50, 219, 220]. Generic instruments can examine a broad range of domains useful to various health conditions and diseases [42, 50]. In contrast, disease-specific PROMs assess one specific patient population in which a particular disease manifests, measure the HRQoL domains relevant to that specific patient population and the impact of specific treatments on that population group [52, 50, 220]. Battery of scales comprises multiple scales that measure one particular HRQoL domain (i.e. social functioning) [50]. Such unidimensional HRQoL instruments will assess a relevant domain in depth, but does not allow to benchmark populations. Moreover, the burden on patients to complete multiple scales can become onerous and time consuming, which may impede results [50, 221]. In this review, we are interested in multidimensional PROMs, such as the generic and disease-specific scales. These multidimensional scales are intended to measure the constructs of interest and consist of subscales that measure different aspects of health [219]. To select the most appropriate PROMs for measuring HRQoL among patients receiving enteral feeding, it is important that these PROMs measure the relevant constructs or concepts that are crucial for patients receiving enteral feeding [130]. Clear definitions about HRQoL in enteral feeding and the constructs being measured should have been provided in the studies that have established the psychometric properties of the PROMs of interest. We will collect these definitions (cf. data extraction).

Among patients with cognitive, speech and language disorders, or in some other populations (i.e. infants), self-reported methods may be inappropriate or impossible [47, 50]. Proxies, defined as healthcare providers, spouses, parents or relatives who are closely involved with the patient [84], can act as alternative sources to obtain valuable information about patients' HRQoL [47, 222]. The proxies are

asked to assess the patient's HRQoL in the manner that they believe the patient would report on this [212]. The use of proxy-reported outcome measures allows to reveal information about the patient's HRQoL that may otherwise be lost [85]. However, these proxy-reported outcome measures are not the same as PROMs as the PRO information is obtained by proxies [45, 53]. Furthermore, relying on proxy-provided information is only useful if proxies are able to report on the different HRQoL domains and if their ratings are sufficiently reliable [85]. The reliability of the proxy ratings can be evaluated by examining patient-proxy agreement. This requires that the PROM and the accompanying proxy-reported outcome instrument measure the same constructs with identical items and that these instruments and items are valid and reliable [77, 86]. As various terms are available for these proxy versions (i.e. proxy-reported outcome measures, parent proxy-report instruments) [53, 86], we will make use of the term 'proxy-reported outcome measures' in this systematic review. Considering the patient-proxy agreement, a review of Sneeuw et al. [47] revealed that proxy raters tend to report more HRQoL problems than patients do themselves. However, the subsequent underestimation of patient's HRQoL should be tempered as discrepancies in patient-proxy agreement were only found significant in a minority of cases [47].

There is a growing interest to take into account HRQoL measures in patients receiving enteral tube feeding in hospital and among outpatients as well [44, 223]. Enteral tube feeding refers to the delivery of complete nutrition into the stomach. The tube can be placed through the nose via the esophagus into the stomach (nasogastric feeding or NG-feeding) or directly into the stomach through the skin on the abdomen (percutaneous gastrostomy (PEG) feeding or percutaneous endoscopic gastrostomy-feeding) [218]. Enteral tube feeding is an alternative for patients with insufficient oral food intake [215] and is most commonly (64.1%) [216] used in patients with dysphagia [216, 218]. Dysphagia or swallowing disorder refers to any dysfunction in the swallowing process as a result of anatomical or physiological deficits in the mouth, pharynx, larynx and/or esophagus [224, 225]. Enteral feeding therapy can benefit patients with dysphagia, who cannot meet their nutritional needs through any other intervention [216]. Although most commonly used in patients with dysphagia, also other conditions can benefit from enteral tube feeding, such as patients with gastrointestinal dysfunction, patients with mental health disorders, patients with psychological problems and patients with increased nutritional requirements [218].

The provision of enteral feeding can yield nutritional benefits [218, 226] and allow weight maintenance [216] in all populations (infants [227], children [228], adults [229] and the elderly [216]).

Some patients with rapidly progressive diseases can also benefit from enteral tube feeding in earlier stages of their disease progress. For example, enteral tube feeding can relieve pressure to eat at mealtimes and can reduce mealtime duration [41]. It may also reduce accelerated weight loss due to poor caloric intake [230]. However, enteral feeding may also have downsides, for example enteral feeding can result in dependency upon the feeding tube [216]. The subsequent psychological and social aspects [216, 229] of this dependency may affect the patient's HRQoL in a negative way [229]. For instance, patients receiving enteral feeding sometimes report feelings of discomfort while dressing and washing, as well as uncomfortable feelings about their body image [44]. Also, loss of social interaction associated with feeding (e.g. not visiting their family and not going out because of tube feeding) can take place [216]. In a study of Roberge et al. [231], feelings of depression were reported in some patients receiving enteral feeding. Though the feeding technique is typically physically well tolerated [231], some consequences of enteral feeding may occur such as local wound complications, leakage around the insertion site, tube occlusion and increased reflux [232, 233]. Aspiration pneumonia, which refers to the inhalation of oropharyngeal material into the larynx and lower respiratory tract [234], can also be a common complication [235]. Finally, in some populations, enteral feeding is associated with an increased mortality rate, both in hospital [236] and in the community [237]. Evaluating the impact of enteral feeding on the patients' HRQoL is therefore essential and should be taken into account when considering enteral feeding.

A literature review related to the topic of this systematic review has been previously published [44]. Brotherton et al. [44] reviewed studies that measured the impact of enteral tube feeding on adult patient's HRQoL to identify feeding-related factors that influence patient's HRQoL. This literature review exposed a lack of specific HRQoL assessment tools [44]. Our systematic review will evaluate the psychometric properties of PROMs and proxy-reported outcome measures that assess HRQoL among patients receiving enteral feeding. Based upon an initially limited search in MEDLINE, the Cochrane Database of Systematic Reviews and JBI COnNECT+ in February 2016, to our knowledge there are no systematic reviews that have reviewed the psychometric properties and the clinical utility of PROMs and proxy-reported outcome measures assessing HRQoL among patients receiving enteral feeding.

The objective of this systematic review is to evaluate the psychometric properties (i.e. validity, reliability and responsiveness) and the clinical utility of PROMs and proxy-reported outcome measures

(questionnaires or structured interviews) that assess HRQoL among patients receiving enteral feeding to make recommendations for its use in clinical practice and research.

Definitions

To fully understand the psychometric properties, some terms should be defined [59, 78]: (Note that these descriptions are also applicable for the proxy-reported outcome measures.)

Validity [59] can be defined in terms of:

- *Content validity* refers to the extent to which the different domains of the PROM are comprehensively sampled by the items in the questionnaire.
- *Criterion-related validity* measures the extent to which scores on a particular PROM are related to a gold standard. For most PROMs, it is not possible to measure criterion validity because there is no gold standard [53]. A gold standard only exists when a shorter version of the PROM is tested against its complete version. This entire version of the PROM is then considered the gold standard [130].
- *Construct validity* examines the extent to which the subscales on a particular PROM are correlated with subscales from other PROMs that assess similar constructs. These comparisons are based on theoretically derived hypotheses concerning the constructs that are being measured. These constructs and hypotheses should be defined in the studies of the PROMs being reviewed. Therefore, the constructs will be used as defined in the studies to evaluate construct validity.
- *Floor and ceiling effects:* In the case of floor and ceiling effects, extreme items are missing in the lower or upper end of the scale. Consequences of these effects are reduced reliability (no ability of the questionnaire to distinguish between patients with highest and lowest scores) and limited responsiveness (changes in subjects cannot be measured).

Reliability encompasses “reproducibility” [59] and “internal consistency” [78]:

Reproducibility can be defined in terms of “agreement” and “reliability” [59]:

- Agreement refers to the degree to which the scores on repeated measures are close to each other, more specifically the absolute measurement error. The following estimates of agreement are frequently used:

  – Standard error of measurement (SEM) = measurement error.

  - Systematic differences should be considered as part of the SEM because they should be distinguished from real changes (i.e. due to treatment). Therefore, $SEM_{agreement}$ is preferred.
  - Standard error of measurement could be converted in the smallest detectable change (SDC). The SDC is the smallest within-person change in score that – with $P < 0.05$ – can be interpreted as a "real change" in one individual (SDC_{ind}) with $SDC_{ind} = 1.96 \times \sqrt{2} \times SEM$ or for a group SDC_{group} with $SDC_{ind}/\sqrt{n}$.

  – Limits of agreement (LOA): "equal the mean change in scores of repeated measurements (mean change) ± $1.96 \times SD_{change}$" [59 (p.37)]. SD_{change} = standard deviation of these changes.

  – Minimal important change (MIC): "the minimal amount of change in the (sub)scale that is considered to be important" [59 (p.37)].

- Reliability refers to the degree to which patients can be distinguished from each other, despite measurement errors (relative measurement error). Reliability is defined as "the proportion of the total variance in the measurements which is because of true differences among patients" [78 (p.743)]. The "true" does not refer to the accuracy of the score, but to its consistency. As previously reported [78], it is important to note that "true" should be seen in the context of classical test theory, as one of the two components of any observation, namely the "true score" and the "error associated with the observation" [78 (p.743)]. "True" is defined by Mokkink et al. as "the average score that would be obtained if the scale were given an infinite number of times" [78 (p.743)]. The preferred estimates of reliability are the intraclass correlation coefficient (ICC) for continuous measures [59, 78, 85] and weighted Cohen's kappa for ordinal measures [59, 78].

Internal consistency refers to the extent to which items in a (sub)scale are intercorrelated and thus measuring the same concept. Factor analysis is applied to assess whether the items form one scale/dimension or more than one. After assessing the number of (homogeneous) (sub)scales, each (sub)scale should be measured by Cronbach's alpha [59].

Responsiveness [59] assesses the ability of a PROM to detect clinically important changes over time, even if these changes are small. Responsiveness is a measure of longitudinal validity and should therefore be examined by testing predefined hypotheses about expected correlations between changes in measures or expected differences in changes between known groups. The PROM should be able to distinguish clinically important change from measurement error. The following estimates of responsiveness are frequently used:

- Responsiveness ratio (Guyatt's RR): The degree in which the clinically important change (MIC) is related to the between-subject variability in within-subject changes in stable subjects (SD_{change}). Terwee et al. [59] recommend the RR to be at least 1.96. "At the value of 1.96, the MIC equals the SDC_{ind} (= 1.96 × SD_{change})"[59 (p.37)].
- Area under the receiver operating characteristics curve or AUC: The ability of a questionnaire to distinguish between patients who have changed and who have not, based on an external criterion.

Clinical utility is defined in terms of "interpretability" and "feasibility" to complete the PROM and the proxy-reported outcome measure. It is considered as an important characteristic of an instrument for its use in research or clinical practice [60].

- *Interpretability* is the extent to which a qualitative meaning can be assigned to quantitative scores. Interpretability can be determined by mean and SD for various subgroups and by determining MIC [59].
- *Feasibility* is the facilitation in completing the PROM, the way of administration, time duration, the patient's burden to complete the PROM and/or the proxy's burden to complete the proxy-reported outcome measure.

6.2 INCLUSION CRITERIA

Types of participants

The current review will consider all studies including patients of all ages (infants, children, adults and elderly), who receive enteral tube feeding as a consequence of dysphagia or because of other conditions, and their proxies.

Intervention/phenomena of interest

This review will consider primary studies that report on:

- Multidimensional PROMs that assess HRQoL among patients receiving enteral feeding.
- Multidimensional proxy-reported outcome measures that assess HRQoL among patients receiving enteral feeding.

Batteries of scales will be excluded because these are not multidimensional HRQoL instruments.

Outcomes

The current review will consider studies that report on the clinical utility and psychometric properties concerning validity, reliability, responsiveness of the PROMs and the proxy-reported outcome measures of interest. It is likely that not all of these outcomes will be present in each of the included articles. This review will include studies that report on at least one of the following outcomes involving validity, reliability, responsiveness and/or clinical utility.

For validity, studies reporting on at least one of the following outcomes will be included:

- *Content validity:* Purpose of the instrument: definition of HRQoL related to enteral feeding and the constructs being measured, target population: adults/children/infants/elderly, mode of administration: self-administered/interview based, number of domains, number of items per scale/domain, scales (single-item and multiple-item scales), items related to enteral feeding, item selection (target population and experts or investigators were involved in item selection), range of scores, response

options and interpretability of the items (clearly, simple items; recall time = the time period to which the questions refer).

- *Criterion-related validity:* Pearson's or Spearman correlation coefficient or other appropriate statistics.
- *Construct validity:* Constructs and hypotheses are specified in advance, correlation coefficients (i.e. Spearman or Pearson's correlation coefficient) or appropriate statistics according to the stated hypotheses.
- *Floor and ceiling effects:* Percentage (%) of respondents that achieve the lowest or highest possible score.

For reliability, studies reporting on at least one of the following outcomes will be included:

- *Reproducibility:* As defined in the background section, reproducibility encompasses ''agreement'' and ''reliability''. Therefore, studies reporting on at least one of the following outcomes of agreement and/or reliability will be included:

  – Agreement: SDC (SDC_{ind} when applied in individuals and SDC_{group} when applied in groups), LOA, convincing arguments that agreement is acceptable, and MIC.

  – Reliability: ICC for continuous measures and weighted Cohen's kappa coefficient for ordinal measures.
- *Internal consistency:* Factor analysis, sample size and Cronbach's alpha.

For responsiveness, studies reporting on at least one of the following outcomes will be included: SDC, MIC, LOA, RR and AUC.

Clinical utility is defined in terms of interpretability and feasibility to complete the PROM and the proxy-reported outcome measure. Studies that report on at least one of the following outcomes will be included:

- *Interpretability:* Mean, SD and MIC.
- *Feasibility:* Time to administer, mode of administration, facilitation required to complete the PROM, patient's burden to complete the PROM and proxy's burden to complete the proxy-reported outcome measure.

Types of studies

The current review will consider any study that reports on multidimensional PROMs and proxy-reported outcome measures that assess HRQoL among patients receiving enteral feeding and addresses validity, reliability, responsiveness and clinical utility.

6.3 SEARCH STRATEGY

The search strategy aims to find both published and unpublished studies. A three-step search strategy will be utilized in this review. An initial limited search of MEDLINE and CINAHL will be undertaken followed by analysis of the text words contained in the title and abstract, and of the index terms used to describe the article. A second search, using all identified keywords and index terms, will then be undertaken across all included databases. Third, citations and the reference list of all selected studies and relevant reviews will be checked to identify any studies that were missed in our searches.

Four concepts will be used to determine the search terms: (1) enteral feeding, (2) PROM, (3) QoL and (4) validity. For each concept, both keywords (e.g. MeSH terms for PubMed) and free-text words will be used. All search terms within each concept will be combined with "OR". Then, all concepts will be combined by means of "AND". The full search strategy in PubMed is represented in Appendix 6.1.

The databases to be searched include: MEDLINE (via PubMed), Embase, CINAHL and PROQOLID.

The search for unpublished studies will include: Google Scholar, MedNar and relevant homepages of patient-reported and proxy-reported outcome tools. Homepages of international and national societies for PROMs, such as the International Society for QoL Research, will be searched. Dissertations and theses will be hand searched; more specifically, three libraries of university colleges for dieticians (VIVES University College Bruges, HoGent campus libraries and Thomas More University College) will be hand searched. The search for dissertations and theses will also be performed by using the ProQuest database.

The search for unpublished studies and screening references will be independently performed by two reviewers (I.S. and A.W.).

Selecting studies

Selecting studies will also be independently performed by the two reviewers (I.S. and A.W.). The selection will be conducted in two stages: first, both reviewers will screen titles and abstracts against the inclusion criteria. All articles that potentially fulfill all criteria will be retrieved. Subsequently, the full texts will be screened using the same inclusion criteria. Studies published in English, German, French and Dutch will be considered for inclusion in this review, as these languages are mastered by the review team. All studies published prior to the start of this systematic review will be considered for inclusion in this review. Another search, as an update, will be conducted between the period of starting the review and finishing it.

Differences in opinion will be resolved by discussion. If no consensus is reached, a third reviewer will be asked to judge whether or not the article fulfills the inclusion criteria. A PRISMA flow chart will be used to present included and excluded studies [238].

Assessment of methodological quality

Papers selected for retrieval will be assessed by two independent reviewers for methodological quality prior to inclusion in the review.

Study quality will be assessed using the COnsensus-based standards for the selection of health status Measurement INstruments (COSMIN) checklist [58, 239]. This is a validated and internationally accepted tool for the assessment of the methodological quality of studies [58]. The checklist focusses on Health-Related Patient-Reported Outcomes (HR-PROs). The COSMIN tool consists of 12 separate sections ("boxes"), of which 10 will be used to evaluate whether a study meets the standard for good methodological quality. Of these, nine boxes represent psychometric properties, and one box focusses on the methodological quality assessment of studies on interpretability of HR-PRO tools [58]. The remaining two boxes contain general requirements (i.e. Generalizability box) and will not be used in this review.

The COSMIN tool is a modular tool, which means that the number of boxes to fill out depends on the number of psychometric characteristics reported in a study. It may not be necessary to complete the whole checklist when assessing the methodological quality of a particular study. When a study, for example, only assesses responsiveness and criterion validity, two boxes need to be scored. Concerning

the validation assessment of an instrument in different language groups, the same box may need to be scored multiple times in case the design of the study was different among countries [58].

The COSMIN tool uses a 4-point scale (excellent, good, fair and poor) to rate each psychometric property for a particular study [239]. A score can be obtained per box by taking the lowest rating of any item in a box ("worse score counts"). Finally, an overall score, which is determined by the lowest box score, can be obtained to rate the methodological quality of the study. The Interpretability box should be completed only for studies that investigate the interpretability of an instrument. The Interpretability box is also recommended to be used in systematic reviews of psychometric properties as a data extraction tool to extract all available information on interpretability issues (i.e. MIC), and therefore, no scoring was developed for that box [239]. Appendix 6.2 outlines the COSMIN tool to be used.

Any disagreements concerning assessment of study quality that arise between the reviewers will be resolved by discussion with a third reviewer.

Studies with low quality, defined as "poor" scores in all boxes, will be excluded from the review.

6.4 DATA EXTRACTION

Data will be extracted from papers included in the review by means of an attached self-developed data extraction tool (Appendix 6.3). The data extracted will include specific details about the study characteristics, instrument characteristics and the outcomes of significance to the review question and specific objectives. The "Generalizability" box from the COSMIN tool will also be included in the data extraction tool [58].

The following data will be extracted from included studies:

- *Study characteristics:* Authors, title and year of publication, aim of the study, study design, patient population characteristics (age, gender, medical history, etiology and/or comorbidities, method of enteral feeding, time on receiving enteral feeding and underlying etiology of receiving enteral feeding), characteristics of controls (age and gender), characteristics of proxy (type of proxy, age and gender) and sample size (= size of the population [patient, controls and proxy] in which the instrument was applied).
- *Descriptive characteristics of the instrument:* Name of the instrument, purpose of the instrument (definition of HRQoL related to enteral feeding and definition of the

constructs), study/target population, mode of administration (self-administered or interview based), facilitation required to complete the PROM, number and type of scales/domains, number of items per scale/domain, total number of items, response options, range of scores and time to administer, and items related to enteral feeding.

- *Outcomes:* The relevant data on the psychometric properties and the clinical utility will be extracted, as previously reported, under the "Inclusion criteria" and the "Background".

Two reviewers will extract all data independently. Any differences in opinion will be discussed and resolved in a consensus meeting between the two reviewers. Authors of primary studies will be contacted by electronic mail in case of missing information or to clarify unclear data.

6.5 DATA SYNTHESIS

To judge the psychometric properties of the PROMs and the proxy-reported outcome measures, the quality criteria proposed by Terwee et al. [59] will be used. These criteria are based on the criteria of the ''Scientific Advisory Committee of the Medical Outcomes Trust'' [240] and have explicit criteria for what constitutes good psychometric properties. The checklist was applied by Barten et al. [241] for the psychometric evaluations of both questionnaires and semi-structured interviews. All psychometric properties of the PROMs will be rated as ''+'' (positive rating), ''–'' (negative rating), ''?'' (indeterminate rating) and ''0'' (no information available) [59].

For validity, the following quality criteria will be used:

- *Content validity:* Purpose of the instrument: definitions of HRQoL in enteral feeding and the constructs being measured, target population: adults/children/infants/elderly, mode of administration: self-administered/interview based, multiple domains, multiple items per scale/domain, items related to enteral feeding, item selection: target population and experts or investigators were involved in item selection, large range of scores, interpretability of the items (clearly, simple items), recall time (= the time period to which the questions refer) is stated and justified.

- *Criterion-related validity:* Good validity if correlation with gold standard ≥0.70. Convincing arguments that gold standard is gold should be described.
- *Construct validity:* Constructs and hypotheses are specified in advance, correlation coefficients (i.e. Spearman or Pearson's correlation coefficient) or appropriate statistics according to the stated hypotheses. Good validity if statistics (i.e. Spearman's rho) show that at least 75% of the results are in correspondence with these hypotheses, in (sub)groups of at least 50 patients or 50 proxies.
- *Floor and ceiling effects:* Absence of floor and ceiling effects: ≤15% of respondents achieve the lowest or highest possible score in a sample of at least 50 patients or 50 proxies.

For reliability, the following quality criteria will be used:

- *Reproducibility*

  – Agreement: Good agreement if MIC < SDC or MIC outside the LOA or convincing arguments that agreement is acceptable. A sample size of at least 50 patients or 50 proxies is needed to assess agreement.

  – Reliability: Good reliability if ICC or Weighted Cohen's kappa is at least 0.70 in a sample size of at least 50 patients or 50 proxies.
- *Internal consistency:* Good internal consistency if factor analysis is first applied on adequate sample size (7 × number of items and >100) AND Cronbach's alpha(s) calculated per dimension AND Cronbach's alpha(s) between 0.70 and 0.95.

For responsiveness, good responsiveness is obtained if SDC < MIC OR MIC outside the LOA OR RR > 1.96 OR AUC ≥ 0.70.

Clinical utility is defined in terms of interpretability and feasibility to complete the PROM and the proxy-reported outcome measure. Therefore, the following criteria will be used:

- *Interpretability:* Interpretability is good if mean and SD of scores for at least four relevant subgroups of patients are presented AND MIC is defined in a sample size of at least 50 patients or 50 proxies.

- *Feasibility:* Based on the arguments of the studies, it will be clear if the PROM and the proxy-reported outcome measure is feasible in terms of time to administer, mode of administration (interview based or questionnaire), facilitation required to complete the PROM, patient's burden to complete the PROM and proxy's burden to complete the proxy-reported outcome measure.

The data will be presented in a narrative text including the psychometric properties and the contextual data, both presented in the included studies. A meta-analysis will not be performed, because it is inappropriate for determining the psychometric properties of the PRO-tools.

The results of the review will be presented for each review question separately. Tables will be used in the presentation of the results.

Appendix 6.1 Search strategy

Concepts	**Search terms (MEDLINE)**
Enteral feeding	*MeSH:* nutritional support; Gastrostomy; Parenteral Nutrition Solutions; Deglutition disorders; Deglutition; Pneumonia, Aspiration
	Title/abstract: "nutritional support"; "Artificial Feeding"; "Enteral nutrition"; "Enteral Feeding"; "Force Feeding*"; "Tube Feeding"; "Feeding Tube*"; "Gastric Feeding"; "Parenteral Nutrition"; "Parenteral Feeding*"; "Intravenous Feeding*"; "Parenteral Hyperalimentation"; "Intravenous Hyperalimentation"; gastrostomy; Gastrostomies; "percutaneous endoscopic"; nasogastric; Dysphagia; "nihil by mouth"; "nihil per os"; "nothing by mouth"; Deglutition*; Swallowing*; "Aspiration Pneumonia*"; "Acid Aspiration*"; "Respiratory Aspiration*"
PROM	*MeSH:* Quality of Health Care; Outcome Assessment (Health Care); Health Care Surveys; Health Impact Assessment; Interviews as Topic; Nutrition Assessment; Nutritional Status; Questionnaires; Proxy
	Title/abstract: "Quality of Health Care"; "Health Care Quality"; "Healthcare Quality"; "Quality of Healthcare"; "Outcome* Assessment*"; "Outcome Studies"; "Outcome Study"; "Outcome* Research*"; "Outcome Measure*"; "Patient Outcome*"; "Patient Centered Outcome*"; "Patient-Centered Outcome*"; "Treatment* Outcome*"; "Treatment* Effectiveness"; "Treatment* Efficacy"; "Rehabilitation Outcome*"; "Process Assessment*"; "Process Measure*"; "Health Care Survey*"; "Healthcare Survey*"; "Health Impact"; "Impact Assessment*"; Interview*; Interviewer*; "Focus Group*"; "Nutrition* Assessment*"; PNI; "Nutritional Index"; "Nutritional Indices"; "Nutritional Indexes"; "Nutrition Index"; "Nutrition Indices"; "Nutrition Indexes"; "deglutition handicap index"; DHI; "Nutrition Survey*"; "Nutritional Survey*"; "nutrition examination*"; "nutritional examination*"; "NHANES"; "Nutritional status"; "Nutrition Status"; "Questionnaire*"; "Self Report*"; proxy*; "Proxies"; "Patient Agent*"
Quality of life	*MeSH:* Quality of Life
	Title/abstract: "Quality of Life"; "Life Qualities"; "Life Quality"
Validity	*MeSH:* Social Validity, Research; Validation Studies as Topic; Validation Studies [Publication Type]; Reproducibility of Results; Program Evaluation; Nursing Evaluation Research
	Title/abstract: validation; Reproducibility; Reliability; Reliabilities; Validity; Validities; "Measurement* Accuracy"; "Measurement Accuracies"; "Program Evaluation*"; "Program Sustainability"; "Program Sustainabilities"; "Program Effectiveness"; "Program Appropriateness"; "Nursing Evaluation*"; "Evaluation Research*"; "Evaluation Studies"; "Evaluation Study"

PNI = prognostic nutritional index, DHI = deglutition handicap index, NHANES = National Health and Nutrition Examination Survey.

Appendix 6.2 Critical appraisal tool (COSMIN checklist)
(Terwee et al. [239])

Reviewer_____Date_____

Author_____Year_____Record Number_____

Box A. Internal consistency	**Excellent**	**Good**	**Fair**	**Poor**
1 Does the scale consist of effect indicators, i.e. is it based on a reflective model?				
Design requirements				
1 Was the percentage of missing items given?	Percentage of missing items described	Percentage of missing items NOT described		
2 Was there a description of how missing items were handled?	Described how missing items were handled	Not described but it can be deduced how missing items were handled	Not clear how missing items were handled	
3 Was the sample size included in the internal consistency analysis adequate?	Adequate sample size (≥100)	Good sample size (50-99)	Moderate sample size (30-49)	Small sample size (<30)
4 Was the unidimensionality of the scale checked, i.e. was factor analysis or IRT model applied?	Factor analysis performed in the study population	Authors refer to another study in which factor analysis was performed in a similar study population	Authors refer to another study in which factor analysis was performed, but not in a similar study population	Factor analysis was NOT performed and no reference to another study
5 Was the sample size included in the unidimensionality analysis adequate?	7* number of items and ≥100	5* number of items and ≥100 OR 6-7* number of items but <100	5* number of items but <100	<5* number of items
6 Was an internal consistency statistic calculated for each (unidimensional) (sub)scale separately?	Internal consistency statistic calculated for each subscale separately			Internal consistency statistic NOT calculated for each subscale separately
7 Were there any important flaws in the design or methods of the study?	No other important methodological flaws in the design or execution of the study		Other minor methodological flaws in the design or execution of the study	Other important methodological flaws in the design or execution of the study

(continued)

Box A. Internal consistency	**Excellent**	**Good**	**Fair**	**Poor**
Statistical methods				
8 For Classical Test Theory (CTT), continuous scores: was Cronbach's alpha calculated?	Cronbach's alpha calculated		Only item-total correlations calculated	No Cronbach's alpha and no item-total correlations calculated
9 For CTT, dichotomous scores: was Cronbach's alpha or KR-20 calculated?	Cronbach's alpha or KR-20 calculated		Only item-total correlations calculated	No Cronbach's alpha or KR-20 and no item-total correlations calculated
10 For IRT: was a goodness of fit statistic at a global level calculated for example χ^2, reliability coefficient of estimated latent trait value (index of (subject or item) separation)?	Goodness of fit statistic at a global level calculated			Goodness of fit statistic at a global level NOT calculated

NB. Item 1 is used to determine whether internal consistency is relevant for the instrument under study. It is not used to rate the quality of the study.

Box B. Reliability: relative measures (including test-retest reliability, inter-rater reliability and intra-rater reliability)	**Excellent**	**Good**	**Fair**	**Poor**
Design requirements				
1 Was the percentage of missing items given?	Percentage of missing items described	Percentage of missing items NOT described		
2 Was there a description of how missing items were handled?	Described how missing items were handled	Not described but it can be deduced how missing items were handled	Not clear how missing items were handled	
3 Was the sample size included in the analysis adequate?	Adequate sample size (≥100)	Good sample size (50-99)	Moderate sample size (30-49)	Small sample size (<30)
4 Were at least two measurements available?	At least two measurements			Only one measurement

(continued)

Box B. Reliability: relative measures (including test-retest reliability, inter-rater reliability and intra-rater reliability)	**Excellent**	**Good**	**Fair**	**Poor**
5 Were the administrations independent?	Independent measurements	Assumable that the measurements were independent	Doubtful whether the measurements were independent	Measurements NOT independent
6 Was the time interval stated?	Time interval stated		Time interval NOT stated	
7 Were patients stable in the interim period on the construct to be measured?	Patients were stable (evidence provided)	Assumable that patients were stable	Unclear if patients were stable	Patients were NOT stable
8 Was the time interval appropriate?	Time interval appropriate		Doubtful whether time interval was appropriate	Time interval NOT appropriate
9 Were the test conditions similar for both measurements for example type of administration, environment and instructions?	Test conditions were similar (evidence provided)	Assumable that test conditions were similar	Unclear if test conditions were similar	Test conditions were NOT similar
10 Were there any important flaws in the design or methods of the study?	No other important methodological flaws in the design or execution of the study		Other minor methodological flaws in the design or execution of the study	Other important methodological flaws in the design or execution of the study
Statistical methods				
11 For continuous scores: was an intraclass correlation coefficient (ICC) calculated?	ICC calculated and model or formula of the ICC is described	ICC calculated but model or formula of the ICC not described or not optimal. Pearson or Spearman correlation coefficient calculated with evidence provided that no systematic change has occurred	Pearson or Spearman correlation coefficient calculated WITHOUT evidence provided that no systematic change has occurred or WITH evidence that systematic change has occurred	No ICC or Pearson or Spearman correlations calculated
12 For dichotomous/nominal/ordinal scores: was Kappa calculated?	Kappa calculated			Only percentage agreement calculated

(continued)

Box B. Reliability: relative measures (including test-retest reliability, inter-rater reliability and intra-rater reliability	**Excellent**	**Good**	**Fair**	**Poor**
13 For ordinal scores: was a weighted kappa calculated?	Weighted Kappa calculated		Unweighted Kappa calculated	Only percentage agreement calculated
14 For ordinal scores: was the weighting scheme described for example linear and quadratic?	Weighting scheme described	Weighting scheme NOT described		

Box C. Measurement error: absolute measures	**Excellent**	**Good**	**Fair**	**Poor**
Design requirements				
1 Was the percentage of missing items given?	Percentage of missing items described	Percentage of missing items NOT described		
2 Was there a description of how missing items were handled?	Described how missing items were handled	Not described but it can be deduced how missing items were handled	Not clear how missing items were handled	
3 Was the sample size included in the analysis adequate?	Adequate sample size (≥100)	Good sample size (50-99)	Moderate sample size (30-49)	Small sample size (<30)
4 Were at least two measurements available?	At least two measurements			Only one measurement
5 Were the administrations independent?	Independent measurements	Assumable that the measurements were independent	Doubtful whether the measurements were independent	Measurements NOT independent
6 Was the time interval stated?	Time interval stated		Time interval NOT stated	

(continued)

Box C. Measurement error: absolute measures	**Excellent**	**Good**	**Fair**	**Poor**
7 Were patients stable in the interim period on the construct to be measured?	Patients were stable (evidence provided)	Assumable that patients were stable	Unclear if patients were stable	Patients were NOT stable
8 Was the time interval appropriate?	Time interval appropriate		Doubtful whether time interval was appropriate	Time interval NOT appropriate
9 Were the test conditions similar for both measurements for example type of administration, environment and instructions	Test conditions were similar (evidence provided)	Assumable that test conditions were similar	Unclear if test conditions were similar	Test conditions were NOT similar
10 Were there any important flaws in the design or methods of the study?	No other important methodological flaws in the design or execution of the study		Other minor methodological flaws in the design or execution of the study	Other important methodological flaws in the design or execution of the study
Statistical methods				
11 For CTT: was the standard error of measurement (SEM), smallest detectable change (SDC) or limits of agreement (LoA) calculated?	SEM, SDC, or LoA calculated	Possible to calculate LoA from the data presented		SEM calculated based on Cronbach's alpha or on SD from another population

Box D. Content validity (including face validity)	**Excellent**	**Good**	**Fair**	**Poor**
General requirements				
1 Was there an assessment of whether all items refer to relevant aspects of the constructs to be measured?	Assessed if all items refer to relevant aspects of the construct to be measured		Aspects of the construct to be measured poorly described AND this was not taken into consideration	NOT assessed if all items refer to relevant aspects of the construct to be measured
2 Was there an assessment of whether all items are relevant for the study population for example age, gender, disease characteristics, country and setting?	Assessed if all items are relevant for the study population in adequate sample size (≥10)	Assessed if all items are relevant for the study population in moderate sample size (5-9)	Assessed if all items are relevant for the study population in small sample size (<5)	NOT assessed if all items are relevant for the study population OR target population not involved
3 Was there an assessment of whether all items are relevant for the purpose of the measurement instrument? (discriminative, evaluative and/or predictive)	Assessed if all items are relevant for the purpose of the application	Purpose of the instrument was not described but assumed	NOT assessed if all items are relevant for the purpose of the application	
4 Was there an assessment of whether all items together comprehensively reflect the construct to be measured?	Assessed if all items together comprehensively reflect the construct to be measured		No theoretical foundation of the construct and this was not taken into consideration	NOT assessed if all items together comprehensively reflect the construct to be measured
5 Were there any important flaws in the design or methods of the study?	No other important methodological flaws in the design or execution of the study		Other minor methodological flaws in the design or execution of the study	Other important methodological flaws in the design or execution of the study

Box E. Structural validity	Excellent	Good	Fair	Poor
1 Does the scale consist of effect indicators, i.e. is it based on a reflective model?				
Design requirements				
2 Was the percentage of missing items given?	Percentage of missing items described	Percentage of missing items NOT described		
3 Was there a description of how missing items were handled?	Described how missing items were handled	Not described but it can be deduced how missing items were handled	Not clear how missing items were handled	
4 Was the sample size included in the analysis adequate?	7* number of items and ≥100	5* number of items and ≥100 OR 5-7* number of items but <100	5* number of items but <100	<5* number of items
5 Were there any important flaws in the design or methods of the study?	No other important methodological flaws in the design or execution of the study		Other minor methodological flaws in the design or execution of the study (e.g. rotation method not described)	Other important methodological flaws in the design or execution of the study (e.g. inappropriate rotation method)
Statistical methods				
6 For CTT: was exploratory or confirmatory factor analysis performed?	Exploratory or confirmatory factor analysis performed and type of factor analysis appropriate in view of existing information	Exploratory factor analysis performed while confirmatory would have been more appropriate		No exploratory or confirmatory factor analysis performed
7 For IRT: were IRT tests for determining the (uni)dimensionality of the items performed?	IRT test for determining (uni)dimensionality performed			IRT test for determining (uni)dimensionality NOT performed

Box F. Hypotheses testing	Excellent	Good	Fair	Poor
Design requirements				
1 Was the percentage of missing items given?	Percentage of missing items described	Percentage of missing items NOT described		
2 Was there a description of how missing items were handled?	Described how missing items were handled	Not described but it can be deduced how missing items were handled	Not clear how missing items were handled	
3 Was the sample size included in the analysis adequate?	Adequate sample size (≥100 per analysis)	Good sample size (50-99 per analysis)	Moderate sample size (30-49 per analysis)	Small sample size (<30 per analysis)
4 Were hypotheses regarding correlations or mean differences formulated a priori (i.e. before data collection)?	Multiple hypotheses formulated a priori	Minimal number of hypotheses formulate a priori	Hypotheses vague or not formulated but possible to deduce what was expected	Unclear what was expected
5 Was the expected *direction* of correlations or mean differences included in the hypotheses?	Expected direction of the correlations or differences stated	Expected direction of the correlations or differences NOT stated		
6 Was the expected absolute or relative *magnitude* of correlations or mean differences included in the hypotheses?	Expected magnitude of the correlations or differences stated	Expected magnitude of the correlations or differences NOT stated		
7 For convergent validity: was an adequate description provided of the comparator instrument(s)?	Adequate description of the constructs measured by the comparator instrument(s)	Adequate description of most of the constructs measured by the comparator instrument(s)	Poor description of the constructs measured by the comparator instrument(s)	NO description of the constructs measured by the comparator instrument(s)
8 For convergent validity: were the measurement properties of the comparator instrument(s) adequately described?	Adequate measurement properties of the comparator instrument(s) in a population similar to the study population?	Adequate measurement properties of the comparator instrument(s) but not sure if these apply to the study population	Some information on measurement properties (or a reference to a study on measurement properties) of the comparator instrument(s) in any study population	No information on the measurement properties of the comparator instrument(s)

(continued)

Box F. Hypotheses testing	**Excellent**	**Good**	**Fair**	**Poor**
9 Were there any important flaws in the design or methods of the study?	No other important methodological flaws in the design or execution of the study		Other minor methodological flaws in the design or execution of the study (e.g. only data presented on a comparison with an instrument that measures another construct)	Other important methodological flaws in the design or execution of the study
Statistical methods				
10 Were design and statistical methods adequate for the hypotheses to be tested?	Statistical methods applied appropriate	Assumable that statistical methods were appropriate, for example Pearson correlations applied, but distribution of scores or mean (SD) not presented	Statistical methods applied NOT optimal	Statistical methods applied NOT appropriate

Box G. Cross-cultural validity	**Excellent**	**Good**	**Fair**	**Poor**
Design requirements				
1 Was the percentage of missing items given?	Percentage of missing items described	Percentage of missing items NOT described		
2 Was there a description of how missing items were handled?	Described how missing items were handled	Not described but it can be deduced how missing items were handled	Not clear how missing items were handled	
3 Was the sample size included in the analysis adequate?	CTT:7* number of items and ≥100 IRT: ≥200 per group	CTT: 5* number of items and ≥100 OR 5-7* number of items but <100	CTT:5* number items but <100 IRT:100-199 per group	CTT:<5* number of items IRT: <100 in one or both groups

(continued)

Box G. Cross-cultural validity	Excellent	Good	Fair	Poor
		IRT: ≥200 in one group and 100-199 in one group		
4 Were both the original language in which the HR-PRO instrument was developed, and the language in which the HR-PRO instrument was translated described?	Both source language and target language described			Source language NOT known
5 Was the expertise of the people involved in the translation process adequately described for example expertise in the disease(s) involved, expertise in the construct to be measured, expertise in both languages?	Expertise of the translators described with respect to disease, construct and language	Expertise of the translators with respect to disease or construct poor or not described	Expertise of the translators with respect to language not described	
6 Did the translators work independently from each other?	Translators worked independent	Assumable that the translators worked independent	Unclear whether translators worked independent	Translators worked NOT independent
7 Were items translated forward and backward?	Multiple forward and multiple backward translations	Multiple forward translations but one backward translation	One forward and one backward translation	Only a forward translation
8 Was there an adequate description of how differences between the original and translated versions were resolved?	Adequate description of how differences between translators were resolved	Poorly or NOT described how differences between translators were resolved		
9 Was the translation reviewed by a committee (e.g. original developers)?	Translation reviewed by a committee (involving other people than the translators, e.g. the original developers)	Translation NOT reviewed by (such) a committee		

(continued)

Box G. Cross-cultural validity	**Excellent**	**Good**	**Fair**	**Poor**
10 Was the HR-PRO instrument pre-tested (e.g. cognitive interviews) to check interpretation, cultural relevance of the translation and ease of comprehension?	Translated instrument pretested in the target population	Translated instrument pretested, but unclear if this was done in the target population	Translated instrument pretested, but NOT in the target population	Translated instrument NOT pre-tested
11 Was the sample used in the pre-test adequately described?	Sample used in the pre-test adequately described		Sample used in the pre-test NOT (adequately) described	
12 Were the samples similar for all characteristics except language and/or cultural background?	Shown that samples were similar for all characteristics except language/culture	Stated (but not shown) that samples were similar for all characteristics except language/culture	Unclear whether samples were similar for all characteristics except language/culture	Samples were NOT similar for all characteristics except language/culture
13 Were there any important flaws in the design or methods of the study?	No other important methodological flaws in the design or execution of the study		Other minor methodological flaws in the design or execution of the study	Other important methodological flaws in the design or execution of the study
Statistical methods				
14 For CTT: was confirmatory factor analysis performed?	Multiple-group confirmatory factor analysis performed			Multiple-group confirmatory factor analysis NOT performed
15 For IRT: was differential item function (DIF) between language groups assessed?	DIF between language groups assessed			DIF between language groups NOT assessed

Box H. Criterion validity	Excellent	Good	Fair	Poor
Design requirements				
1 Was the percentage of missing items given?	Percentage of missing items described	Percentage of missing items NOT described		
2 Was there a description of how missing items were handled?	Described how missing items were handled	Not described but it can be deduced how missing items were handled	Not clear how missing items were handled	
3 Was the sample size included in the analysis adequate?	Adequate sample size (≥100)	Good sample size (50-99)	Moderate sample size (30-49)	Small sample size (<30)
4 Can the criterion used or employed be considered as a reasonable "gold standard"?	Criterion used can be considered an adequate "gold standard" (evidence provided)	No evidence provided, but assumable that the criterion used can be considered an adequate "gold standard"	Unclear whether the criterion used can be considered an adequate "gold standard"	Criterion used can NOT be considered an adequate "gold standard"
5 Were there any important flaws in the design or methods of the study?	No other important methodological flaws in the design or execution of the study		Other minor methodological flaws in the design or execution of the study	Other important methodological flaws in the design or execution of the study
Statistical methods				
6 For continuous scores: were correlations, or the area under the receiver operating curve calculated?	Correlations or AUC calculated			Correlations or AUC NOT calculated
7 For dichotomous scores: were sensitivity and specificity determined?	Sensitivity and specificity calculated			Sensitivity and specificity NOT calculated

Box I. Responsiveness	Excellent	Good	Fair	Poor
Design requirements				
1 Was the percentage of missing items given?	Percentage of missing items described	Percentage of missing items NOT described		
2 Was there a description of how missing items were handled?	Described how missing items were handled	Not described but it can be deduced how missing items were handled	Not clear how missing items were handled	
3 Was the sample size included in the analysis adequate?	Adequate sample size (≥100)	Good sample size (50-99)	Moderate sample size (30-49)	Small sample size (<30)
4 Was a longitudinal design with at least two measurements used?	Longitudinal design used			No longitudinal design used
5 Was the time interval stated?	Time interval adequately described			Time interval NOT described
6 If anything occurred in the interim period (e.g. intervention, other relevant events), was it adequately described?	Anything that occurred during the interim period (e.g. treatment) adequately described	Assumable what occurred during the interim period	Unclear or NOT described what occurred during the interim period	
7 Was a proportion of the patients changed (i.e. improvement or deterioration)?	Part of the patients were changed (evidence provided)	NO evidence provided, but assumable that part of the patients were changed	Unclear if part of the patients were changed	Patients were NOT changed
Design requirements for hypotheses testing				
For constructs for which a gold standard was not available:				
8 Were hypotheses about changes in scores formulated a priori (i.e. before data collection)?	Hypotheses formulated a priori		Hypotheses vague or not formulated but possible to deduce what was expected	Unclear what was expected

(continued)

Box I. Responsiveness	**Excellent**	**Good**	**Fair**	**Poor**
9 Was the expected *direction* of correlations or mean differences of the change scores of HR-PRO instruments included in these hypotheses?	Expected direction of the correlations or differences stated	Expected direction of the correlations or differences NOT stated		
10 Were the expected absolute or relative *magnitude* of correlations or mean differences of the change scores of HR-PRO instruments included in these hypotheses?	Expected magnitude of the correlations or differences stated	Expected magnitude of the correlations or differences NOT stated		
11 Was an adequate description provided of the comparator instrument(s)?	Adequate description of the constructs measured by the comparator instrument(s)		Poor description of the constructs measured by the comparator instrument(s)	NO description of the constructs measured by the comparator instrument(s)
12 Were the measurement properties of the comparator instrument(s) adequately described?	Adequate measurement properties of the comparator instrument(s) in a population similar to the study population	Adequate measurement properties of the comparator instrument(s) but not sure if these apply to the study population	Some information on measurement properties (or a reference to a study on measurement properties) of the comparator instrument(s) in any study population	NO information on the measurement properties of the comparator instrument(s)
13 Were there any important flaws in the design or methods of the study?	No other important methodological flaws in the design or execution of the study		Other minor methodological flaws in the design or execution of the study (e.g. only data presented on a comparison with an instrument that measures another construct)	Other important methodological flaws in the design or execution of the study
Statistical methods				
14 Were design and statistical methods adequate for the hypotheses to be tested?	Statistical methods applied appropriate		Statistical methods applied NOT optimal	Statistical methods applied NOT appropriate

(continued)

Box I. Responsiveness	**Excellent**	**Good**	**Fair**	**Poor**
Design requirements for comparison with a gold standard For constructs for which a gold standard was available:				
15 Can the criterion for change be considered as a reasonable gold standard?	Criterion used can be considered an adequate "gold standard" (evidence provided)	No evidence provided, but assumable that the criterion used can be considered an adequate "gold standard"	Unclear whether the criterion used can be considered an adequate "gold standard"	Criterion used can NOT be considered an adequate "gold standard"
16 Were there any important flaws in the design or methods of the study?	No other important methodological flaws in the design or execution of the study		Other minor methodological flaws in the design or execution of the study	Other important methodological flaws in the design or execution of the study
Statistical methods				
17 For continuous scores: were correlations between change scores, or the area under the receiver operator characteristic (ROC) curve calculated?	Correlations under Area under the ROC Curve (AUC) calculated			Correlations or AUC NOT calculated
18 For dichotomous scales: were sensitivity and specificity (changed versus not changed) determined?	Sensitivity and specificity calculated			Sensitivity and specificity NOT calculated

Box J. Interpretability	**Yes**	**No**	**?**
1 Was the percentage of missing items given?	□	□	
2 Was there a description of how missing items were handled?	□	□	
3 Was the sample size included in the analysis adequate?	□	□	□
4 Was the distribution of the (total) scores in the study sample described?	□	□	
5 Was the percentage of the respondents who had the lowest possible (total) score described?	□	□	
6 Was the percentage of the respondents who had the highest possible (total) score described?	□	□	
7 Were scores and change scores (i.e. means and SD) presented for relevant (sub) groups for example for normative groups, subgroups of patients or the general population?	□	□	
8 Was the minimal important change (MIC) or the minimal important difference (MID) determined?	□	□	
9 Were there any important flaws in the design or methods of the study?	□	□	

The Interpretability box will be used to extract all relevant information on the interpretability issues from the included studies. All information available regarding interpretability issues will be further described in that box.

Appendix 6.3 Data extraction tool

		Study
Authors, title and year of publication		
Aim of the study		
Study design		
Patient population characteristics	Age	
	Gender	
	Medical history	
	Etiology	
	Comorbidities	
	Method of enteral feeding (PEG, …)	
	Time on receiving enteral feeding	
	Underlying etiology of receiving enteral feeding	
	Sample size	
Characteristics of control group	Age	
	Gender	
	Sample size	
Characteristics of proxy	Type of proxy	
	Age	
	Gender	
	Sample size	
Instrument characteristics	Name	
	Purpose (definitions of HRQoL in enteral feeding and the constructs being measured)	
	Study/target population	
	Mode of administration (self-administered or interview-based)	
	Facilitation required to complete the PROM	
	Number and type of scales/domains	
	Number of items per scale/domain	
	Total number of items	
	Response options	
	Range of scores	
	Time to administer	
	Items related to enteral feeding	
Outcome domain evaluated	Validity	
	Reliability and evaluation of agreement/disagreement between patient and proxy	
	Responsiveness	
	Interpretability	

Appendix 6.3 continued

		Study
Outcome domain evaluated	Patient's burden to complete the PROM	
	Proxy's burden to complete the proxy-reported outcome measure	
Statistical method(s) used		
Results		
Generalizability	Was the sample in which the PROM/proxy-reported outcome measure was evaluated adequately described?	Yes No NA
	In terms of:	
	1 Median or mean age (with standard deviation or range)?	□ □
	2 Distribution of sex?	□ □
	3 Important disease characteristics (e.g. severity, status and duration) and description of treatment?	□ □ □
	4 Setting(s) in which the study was conducted for example general population, primary care or hospital/rehabilitation care?	□ □
	5 Countries in which the study was conducted?	□ □
	6 Language in which the PROM/proxy-reported outcome measure was evaluated?	□ □
	7 Was the method used to select patients/proxies adequately described for example convenience, consecutive or random?	□ □
		Yes No ?
	8 Was the percentage of missing responses (response rate) acceptable?	□ □ □
Author conclusions		
Reviewer's comments		

CHAPTER VII: SUMMARY, CONCLUSIONS AND FUTURE PERSPECTIVES

7.1 SUMMARY AND CONCLUSIONS

Stroke has been recognized worldwide as a leading cause of disability and death [1, 2]. Post-stroke oropharyngeal dysphagia, which is defined as the presence of an abnormal swallowing physiology of the upper aerodigestive tract [16], is a common complication after stroke [1]. The reported short- and long-term outcomes associated with PSOD are often not consistent among studies [18], which can be explained by variations in the assessment methods of dysphagia, the timing of the initial evaluation after stroke and lesion location [18]. Outcome predictions are also influenced by methodological variations. Most studies have used retrospective methods [108-110], chart reviews [108-110] and/or interview methods [15, 19, 109] to document the outcomes of PSOD and/or feeding status. Furthermore, the interval between successive measurements was often too long [15, 116]. Consistency in assessing swallowing and evaluating functional independence at regular intervals during follow-up [15] enables better descriptions of and predictions regarding who will have a persistent swallowing disorder, who will require a modified oral intake, who will have decreased functional independence, and who is at risk of long-term institutionalization. Furthermore, these improvements will help patients and their families to better understand the course of the disease [11]. Therefore, longitudinal datasets with comprehensive measurements performed at different time points are required to build models that allow accurate predictions of mortality and short- and long-term outcomes [116] to enable the planning of future care and support [15].

The main purpose of this thesis was therefore to conduct a multi-centre prospective longitudinal study with frequent assessments of swallowing, oral health status, feeding status and functional independence at specific time points during one year post-stroke (i.e., at 1 month, 3 months, 6 months, and 1 year). The aim of this prospective study (cf. study II, Chapter III) was to study the role of oropharyngeal dysphagia in predicting place of residence and functional outcomes (i.e., feeding status, functional independence, pneumonia, survival and HRQoL) after stroke. An essential methodological issue that needed to be solved prior to the actual start of this study was to select appropriate standardized clinical assessment tools that enabled the prediction of the outcomes of interest. We were particularly interested in the value of the MASA and of an oral health assessment instrument to predict functional outcomes after stroke. The MASA is a standardized CSE that has been specifically validated for assessing

PSOD [87-89]. However, a standardized oral health assessment tool for use by SLPs was not available. In **STUDY I** of this thesis, we investigated therefore the feasibility and reliability of the OHAT as used by SLPs (cf. Chapter II). A multi-centre study in 132 subjects was conducted by three SLPs. The feasibility was assessed by reporting the time to complete the OHAT and by the completion of a semi-structured questionnaire with statements on the self-perceived scoring ability. The inter-rater, test-retest and intra-rater reliabilities of the OHAT were assessed. The study demonstrated that the OHAT is a feasible instrument and quick and simple to administer. Total OHAT scores showed good inter-rater, intra-rater and test-retest agreement. A high level of inter-rater reliability was achieved with almost perfect agreement for seven of the eight individual categories and the need for referral, as well as with perfect agreement for 'dental pain'. Excellent test-retest agreement was reached for 'natural teeth' and 'dentures'. Lower levels of test-retest agreement were found for 'gums and tissues', 'lips', 'dental pain' and the overall referral decision. All individual categories showed excellent intra-rater agreement, except for 'gums and tissues'. This was the first study to examine the feasibility and reliability of the OHAT as used by SLPs. As the results showed both good feasibility and reliability, we can conclude that the OHAT has the potential to add to the CSE.

After establishing the feasibility and reliability of the OHAT as used by SLPs, **STUDY II** was performed. This multi-centre prospective long-term study was performed in 151 stroke patients (cf. Chapter III). The main aim of **STUDY II** was to determine whether the presence of PSOD at baseline (i.e., during the first swallowing examination after the onset of stroke) could predict place of residence after discharge. The secondary aims of **STUDY II** were to evaluate whether baseline MASA and OHAT scores could predict the occurrence of pneumonia during hospitalization and follow-up and to examine whether the baseline MASA score could predict patient outcomes over time in terms of feeding status (i.e., measured by the FOIS), functional independence (i.e., measured by the Barthel index), and place of residence. Because the MASA was performed at different time points during the subsequent year, we also investigated whether the relationships among the MASA score and feeding status, functional independence and place of residence changed over time by studying the associations among these factors at each time point. Survival was investigated as a function of baseline MASA and/or OHAT scores while correcting for the other variables. Dysphagia was present in 76.2% of patients at baseline. The results of this study – with demonstrated sufficient statistical power – showed that having PSOD at baseline could predict failure to return home after discharge. The baseline MASA and OHAT scores were both

significant predictors of pneumonia during hospitalization and follow-up. The baseline MASA score was a significant predictor of feeding status and functional independence over time, particularly in the short term (for a detailed description, we refer to Chapter III). As the year progressed, the MASA score at a particular time point (i.e., the MASA score from 1 month to 1 year for the FOIS and the MASA score from 3 months to 1 year for the Barthel index) became more informative than the baseline MASA score for predicting the feeding status and functional independence at that time point. The baseline MASA score was shown to be a significant predictor of the place or residence during follow-up when combined with time and the MASA scores at specific time points. The MASA repeated at specific time points was not a significant predictor of place of residence during follow-up. The baseline MASA score and age were shown to be significant predictors of survival after correcting for the occurrence of pneumonia and the presence of stroke expansion/recurrence during the one-year follow-up. The baseline OHAT score nearly reached significance when it was included instead of the MASA in the statistical model. Severe PSOD, as assessed by the MASA, was associated with poor survival. This study confirmed that PSOD patients experience poor outcomes.

Although not reported in Chapter III, one of the outcomes of study II was the observation that PSOD patients with additional language and/or cognitive impairment often struggled with the DSWAL-QoL. In **STUDY III** (cf. Chapter IV), we therefore developed the aDSWAL-QoL to increase self-reporting in DysLC patients. We developed the aDSWAL-QoL based on aphasia- and cognitive-friendly recommendations from previous literature [55, 70, 71, 120-123]. Furthermore, the feasibility and CTT-psychometric properties (i.e., internal consistency, test-retest reliability, and criterion validity) of the aDSWAL-QoL were evaluated. A secondary aim of **STUDY III** was to predict the need for assistance to complete the aDSWAL-QoL based on cognitive impairment, language comprehension, age, group, and functional dependency. We developed the aDSWAL-QoL during a five-step investigation (further specified in Chapter IV) including an DSWAL-QoL scale adjustment process, the involvement of an expert panel, initial trial assessments, further refinement of the instrument and preparation of the actual investigation, and additional trial assessments. A cross-sectional study was performed in 78 dysphagic patients, among 43 were DysLC patients. The aDSWAL-QoL was more feasible than the DSWAL-QoL for the DysLC group: they required less assistance, missed fewer responses, and gave more reliable responses. Even with the retest, the DysLC patients showed greater ease in completing the aDSWAL-QoL. We obtained high internal consistency, excellent test-retest agreement and good criterion validity

for the total aDSWAL-QoL scale. Almost all subscales showed high internal consistency, moderate to good test-retest agreement and criterion validity. However, the psychometric properties of the 'Food selection' subscale were inadequate; to a lesser extent, this was also true for the Fear of eating and Eating desire subscales. Therefore, caution is required when interpreting the results from these subscales. With respect to the secondary aim of **STUDY III**, we found moderate negative correlations between the need for assistance in the DysLC group and cognitive impairment, language comprehension, and functional independency. This highlighted the increasing need for assistance as functional levels decrease. We found that in our sample population, an MMSE ≤20, DysLC group membership, and a Barthel Index <7 could predict the need for assistance in completing the aDSWAL-QoL. In conclusion, the aDSWAL-QoL is a feasible, reliable, and valid tool for use with DysLC patients. Hence, the instrument may be used by a large population.

To (further) evaluate the construct validity of DSWAL-QoL and aDSWAL-QoL, a modern psychometric approach, called the Rasch methodology, was applied. This model has been considered the gold standard against which PRO scales summarizing item responses should be tested [80]. The aim of **STUDY IV** (cf. Chapter V) was therefore to evaluate the structural validity (i.e., an aspect of construct validity) and objectivity of both the DSWAL-QoL and aDSWAL-QoL total scales and subscales and the statistical sufficiency of the total scores and subscales scores using item analysis with the Rasch model. The Rasch-item analysis was performed using RUMM2030 software with previously collected data from a validation study of 108 patients (among whom 78 were involved in study III). The assessment included evaluations of overall model fit, reliability, unidimensionality, threshold ordering, individual item and person fits, DIF, LID and targeting. The analysis could not establish the psychometric properties of either of the scales or their subscales because they did not fit the Rasch model, and multidimensionality, disordered thresholds, DIF, and/or LID were found. The reliability and power of fit were high for the total scales (PSI = 0.93) but low for most of the subscales (PSI < 0.70). The targeting of persons and items was suboptimal. The main source of misfit was disordered thresholds for both the total scales and subscales. Based on the results of the analysis, adjustments to improve the scales were implemented as follows: disordered thresholds were rescaled, misfit items were removed and items were split for DIF. However, the multidimensionality and LID could not be resolved. The reliability and power of fit remained low for most of the subscales. This study represented the first analyses of the DSWAL-QoL and aDSWAL-QoL scales with the Rasch model. We could conclude that relying on the DSWAL-QoL and aDSWAL-QoL

total and subscale scores to make conclusions regarding dysphagia-related HRQoL should be treated with caution before the structural validity and objectivity of both scales have been established. One major limitation of this study was the relatively limited sample size. Our study sample of 108 patients was within the recommended sample size [200]. However, the effective sample size (i.e., without subjects showing extreme scores) in this study was smaller than the original sample size. Therefore, this study had to be interpreted as a pilot study. To derive definitive conclusions about the items and the scales, well-targeted and sample sizes ≥ 250 people are recommended [196, 200]. So, until the psychometric properties have been established in a larger sample, we suggest to use the proposed scoring structure (cf. Chapter V) for each individual item, taking into account to derive only qualitative information from that item. This study illustrated the added value of a Rasch analysis in the detection of the psychometric strengths and weaknesses of these rating scales. Therefore, this study can be viewed as an essential step forward toward the further improvement of these scales.

Studies II, III and IV showed that (1) not all items of the DSWAL-QoL and aDSWAL-QoL were suitable for patients receiving enteral feeding, and (2) in patients with severe language impairment and/or cognitive disorders (MMSE < 5 and AAT < 37) self-reporting was impossible. With regard to these findings, caution should therefore be undertaken in the use of these DSWAL-QoL and aDSWAL-QoL scales with tube-fed patients. A disease-specific PRO that is specifically developed and validated for this population group is recommended. In patients where self-reporting is impossible, valuable information about their HRQoL can be obtained by proxy-reported outcome measures. As a means to select the most suitable PRO and proxy-reported outcome measure for assessing HRQoL among patients receiving enteral feeding, it is important to evaluate the psychometric properties and the clinical utility of these measurement instruments [58]. **STUDY V** involved the development of a systematic review protocol to systematically review the clinical utility and psychometric properties of (1) measures to assess HRQoL among patients receiving enteral feeding, regardless the cause of receiving enteral feeding, (2) measures to assess HRQoL in the subgroup 'patients receiving enteral feeding as a consequence of dysphagia' and (3) measures (1 and 2) that are completed by patients' proxy to assess HRQoL among patients receiving enteral feeding. The systematic review protocol (cf. Chapter VI) includes a detailed description of the objectives, inclusion criteria, search strategy (including the selection of the studies and the evaluation of their methodological quality), data extraction and data synthesis.

In summary, the conclusions of this thesis were:

1. The OHAT is a feasible and reliable oral health assessment tool that can be used by SLPs to screen oral health in a standarized manner (Chapter II).
2. The presence of PSOD at baseline could predict failure to return home after discharge (Chapter III).
3. The baseline MASA and OHAT scores were both significant predictors of pneumonia during hospitalization and follow-up (Chapter III).
4. The baseline MASA score was a significant predictor of feeding status and functional independence over time, but its effect was stronger in the short term than in the long term. The baseline MASA was also a significant predictor of the place of residence during follow-up; its effect was similar throughout the year (Chapter III).
5. With regard to the effect of the baseline MASA score combined with the MASA scores assessed at specific time points during the subsequent year, the following can be concluded: (1) the baseline MASA score was a significant predictor of the feeding status, functional independence and place of residence during follow-up; (2) the MASA score at a specific time point was a significant predictor of feeding status and functional independence, and as the year progressed, it became more informative for the feeding status and functional independence than the MASA at baseline; (3) the MASA score at a specific time point was not a significant predictor of place of residence during follow-up (Chapter III).
6. The baseline MASA score was shown to be a significant predictor of survival following stroke. Severe PSOD, as assessed by the MASA, was associated with poor survival. The OHAT nearly reached significance for the prediction of survival, when it was included instead of the MASA in the statistical model (Chapter III).
7. The aDSWAL-QoL has been developed, based on aphasia- and cognitive-friendly recommendations from literature (Chapter IV).
8. The aDSWAL-QoL is a feasible and – according to CTT – reliable and valid tool for use with DysLC patients. However, the psychometric properties of the Food selection and – to a lesser extent – the Fear of eating and Eating desire subscales were insufficient (Chapter IV).
9. The analysis with the Rasch model could not establish the structural validity and objectivity of the DSWAL-QoL and aDSWAL-QoL total scales and subscales and the statistical

sufficiency of the total scores and subscale scores. The analysis with the Rasch model identified areas that require further investigation. Potential solutions suggested by the Rasch model induced scale improvement. We suggest to use the proposed scoring structure for the items and to derive only qualitative information from that item until the psychometric properties have been established in a larger and well-targeted sample (Chapter V).

10. A systematic review protocol was developed to systematically review the psychometric properties and the clinical utility of PROs and proxy-reported outcome measures that assess HRQoL among patients receiving enteral feeding to make recommendations for use in clinical practice and research (Chapter VI).

7.2 FUTURE PERSPECTIVES

In study I, we established the feasibility and reliability of the OHAT as used by SLPs and implemented the use of the OHAT in study II to screen oral health in a standardized manner in stroke patients with and without dysphagia. Future research is necessary to evaluate whether the implementation of the OHAT within a CSE can promote oral health care awareness and improve oral care management in daily clinical practice. Following some small adjustments (e.g., providing a section on the scoresheet where the pairs of teeth in the chewing position can be reported), the instrument will be usable in clinical practice. This study I was limited by not examining the concurrent (i.e., criterion) validity of the tool as determined by comparing the OHAT results obtained from SLPs with a dental examination completed by a qualified dentist. Concurrent validity of the OHAT warrants further assessment to demonstrate the diagnostic accurracy of the tool. Furthermore, it would be interesting to induce more differentation in the referral strategies and to refine the care strategies following OHAT assessment. In study II, we explored the value of the OHAT in predicting pneumonia and survival. However, future research is needed. As pneumonia is a multi-factorial phenomenon, repeating binary logistic regression analyses while including more clinical predictors that may contribute to pneumonia occurrence is warranted. We will also report on the value of the OHAT in predicting place of residence, feeding status and functional independence. As described in study II (cf. Chapter III), assessments of swallowing function and type of diet were not repeated at discharge for practical reasons. In future studies, it would be interesting to repeat these assessments at

discharge to determine the extent to which these parameters influence discharge destination and to plan future care and support. We collected more data in study II than we reported in Chapter III. We will analyze and report on these data (e.g., analysis of dysphagia-related QoL and oral health-related QoL in stroke patients). In study III, we developed the aDSWAL-QoL to be used by a large population DysLC patients. However, it would be interesting to investigate whether patients could be served more effectively by being divided into further subgroups, i.e. one group having exclusively language impairment, and modifying the tool according to their needs. Future research should also investigate whether conversion of the pen-and-paper self-administered aDSWAL-QoL into a computer-assisted self-administered format could decrease patient reliance on the investigators for assistance. We have developed a proxy version of the DSWAL-QoL and a preliminary analysis on patient-proxy agreement has already been performed. However, we will further analyze and report on these data. In study III, we included patients who had dysphagia for at least one month and not those in the immediate post-stroke phase, because of validation requirements. If the patient's condition permits, it would be interesting to assess dysphagia-related HRQoL in the acute (and/or subacute) phase following stroke and to evaluate whether the aDSWAL-QoL can be applied at this stage. Study IV included a sample size of 108 patients to perform item analysis with the Rasch model on the DSWAL-QoL and aDSWAL-QoL. Although this sample size was within the recommended sample size [200], the effective sample was smaller due to the presence of extreme scores. Therefore, study IV was considered a pilot study. In order to derive definitive conclusions about the items and the scales, a large empirical study is recommended. We recommend using IRT for further validation of the original SWAL-QoL and all of its translations. Furthermore, it would be beneficial to create and validate shorter versions of the DSWAL-QoL and aDSWAL-QoL. Regarding the observations that some of the items of the DSWAL-QoL and aDSWAL-QoL questionnaires were not suitable for dysphagic patients receiving enteral feeding and that self-reporting was impossible among dysphagic patients with severe language and/or cognitive impairment, we are now performing a systematic review to examine the psychometric properties and the clinical utility of PROs and proxy-reported outcome measures that assess HRQoL among patients receiving enteral feeding. Taking into account the importance of patient involvement in the development and validation of PROs [242] and the weaknesses of existing scales (cf. study III and IV), we decided to develop our own disease-specific PRO for patients receiving enteral feeding as a consequence of oropharyngeal dysphagia. This research is ongoing.

HOOFDSTUK VII: SAMENVATTING, CONCLUSIES EN TOEKOMSTPERSPECTIEVEN

7.1 SAMENVATTING EN CONCLUSIES

Cerebrovasculair accident (CVA) wordt wereldwijd beschouwd als een hoofdoorzaak van functionele beperkingen en mortaliteit [1, 2]. Orofaryngeale dysfagie (OD), meer bepaald de aanwezigheid van een abnormale slikfunctie van het bovenste spijsverteringskanaal [16], is een vaak voorkomende complicatie na CVA [1]. De gerapporteerde korte- en langetermijnoutcomes met betrekking tot dysfagie na CVA zijn vaak uiteenlopend [18]. Dit kan verklaard worden door de verschillende methodes die gebruikt worden om dysfagie te detecteren of te diagnosticeren, het tijdstip waarop de eerste slikevaluatie na het CVA gebeurt en de lokalisatie van het letsel [18]. Ook verschillen in de gehanteerde methodologie kunnen de langetermijnoutcome beïnvloeden. De meeste studies hebben gebruik gemaakt van retrospectieve methodes [108-110], dossierstudies [108-110] en/of interviewmethodes [15, 19, 109] om de outcome van OD en de voedingsstatus na CVA te documenteren. Bovendien was het interval tussen de opeenvolgende meetmomenten vaak te lang [15, 116]. De consistente afname van slikonderzoeken en beoordelingen van functionele onafhankelijkheid op regelmatige tijdstippen post-CVA [15] is belangrijk omdat het toelaat om beter te beschrijven en voorspellingen te doen inzake wie de kans loopt op een persistente slikstoornis, wie een aangepast dieet zal nodig hebben, wie blijvend functioneel afhankelijk zal zijn en wie een risico op institutionalisering op lange termijn vertoont. Dit zorgt er ook voor dat patiënten en hun familie meer inzicht krijgen in het verloop van de ziekte [11]. Longitudinale datasets met uitgebreide metingen op specifieke tijdstippen zijn nodig om modellen te construeren die het mogelijk maken accurate voorspellingen te doen aangaande korte- en langetermijnoutcomes en mortaliteit [116] en vervolgens in staat stellen het zorgpad zo goed mogelijk uit te bouwen [12].

Het hoofddoel van deze thesis was om een multicentrische, prospectieve, longitudinale studie uit te voeren waarbij de slikfunctie, de mondgezondheid, de voedingsstatus en de functionele onafhankelijkheid op regelmatige momenten gedurende één jaar post-CVA (d.i. op 1 maand, 3 maanden, 6 maanden en 1 jaar post-CVA) werden onderzocht. Het doel van die prospectieve studie (cfr. studie II, Hoofdstuk III) was om na te gaan of orofaryngeale dysfagie een rol speelt bij het voorspellen van de verblijfplaats en de functionele outcomes na een CVA. Een belangrijke methodologische kwestie hierbij was het selecteren van geschikte, gestandaardiseerde, klinische onderzoeksmethoden die toelieten voorspellingen met betrekking tot die outcomes te maken. De MASA is een gestandaardiseerd klinisch

diagnostisch slikonderzoek dat specifiek gevalideerd werd voor gebruik bij CVA-patiënten om orofaryngeale dysfagie te diagnosticeren [87-89]. Een gestandaardiseerd instrument om de mondgezondheid door logopedisten te laten beoordelen was echter niet beschikbaar. **STUDIE I** had dan ook tot doel om de haalbaarheid en de betrouwbaarheid van de OHAT te evalueren voor gebruik door logopedisten (cfr. Hoofdstuk II). Een multicentrische studie werd uitgevoerd door drie logopedisten op 132 proefpersonen. De haalbaarheid van de OHAT werd geëvalueerd aan de hand van een tijdmeting en door middel van het invullen van een semi-gestructureerde vragenlijst die peilde naar de bekwaamheid in het scoren van het instrument. De interbeoordelaarsbetrouwbaarheid, de test-hertestbetrouwbaarheid en intrabeoordelaarsbetrouwbaarheid van de OHAT werden onderzocht. De studie toonde aan dat de OHAT een haalbaar instrument is dat snel en gemakkelijk af te nemen is. De totale OHAT-scores vertoonden een goede interbeoordelaars-, intrabeoordelaars- en test-hertestbetrouwbaarheid. Wat de interbeoordelaarsbetrouwbaarheid betreft, een hoge proportie van overeenstemming werd gevonden voor zeven van de acht individuele categorieën en voor de doorverwijzingsbeslissing en een perfecte overeenkomst werd gevonden voor 'dental pain'. Een uitstekende test-hertest overeenkomst werd gevonden voor 'natural teeth' en 'dentures'. Lagere niveaus van test-hertestovereenkomst werden gevonden voor 'gums and tissues', 'lips', 'dental pain', en de doorverwijzingsbeslissing. Alle individuele categorieën vertoonden een uitstekende intrabeoordelaarsbetrouwbaarheid, behalve de categorie 'gums and tissues'. Deze studie was de eerste die de haalbaarheid en de betrouwbaarheid van de OHAT voor gebruik door logopedisten heeft onderzocht. We kunnen op basis van de resultaten (d.i. haalbaar en betrouwbaar) concluderen dat de OHAT potentieel heeft om deel uit te maken van het klinisch slikonderzoek.

Na het vaststellen van de haalbaarheid en de betrouwbaarheid van de OHAT voor gebruik door logopedisten, werd **STUDIE II** uitgevoerd. Deze multicentrische, prospectieve, langetermijnstudie werd uitgevoerd op 151 CVA-patiënten (cfr. Hoofdstuk III). Het hoofddoel van **STUDIE II** was om te bepalen of de aanwezigheid van OD baseline (d.w.z., tijdens het eerste slikonderzoek dat plaatsvond post-CVA) de verblijfplaats na ontslag kon voorspellen. Secundaire doelen van **STUDIE II** hadden betrekking op de predictieve waarde van de baseline MASA- en OHAT-scores bij het voorspellen van pneumonie tijdens hospitalisatie en follow-up alsook op de predictieve waarde van de baseline MASA-score bij het voorspellen van patiëntenoutcomes in termen van voedingsstatus (d.i. gemeten door de FOIS), functionele onafhankelijkheid (d.i. gemeten door de Barthel index) en verblijfplaats over de tijd heen.

Omdat de MASA ook op verschillende tijdstippen gedurende het opeenvolgende jaar werd afgenomen, werd op ieder tijdstip onderzocht of er veranderingen optraden in de samenhang tussen de MASA scores en de voedingsstatus, functionele onafhankelijkheid en verblijfplaats. **STUDIE II** onderzocht eveneens overleving als functie van baseline MASA- en/of OHAT-scores, terwijl er rekening gehouden werd met andere variabelen. De resultaten van deze studie gaven aan dat dysfagie baseline aanwezig was in 76.2% van de patiënten. Deze studie – die voldoende statistische power behaalde – toonde aan dat de baseline aanwezigheid van OD na CVA van voorspellende betekenis was voor het niet naar huis gaan na ontslag uit het ziekenhuis. De baseline MASA- en OHAT-scores waren beide significante voorspellers van pneumonie, zowel tijdens hospitalisatie als tijdens follow-up. De baseline MASA-score was een significante voorspeller van de voedingsstatus en functionele onafhankelijkheid over de tijd heen en dit vooral op korte termijn (voor een gedetailleerde beschrijving verwijzen we graag naar Hoofdstuk III). Naarmate het jaar vorderde, werd de MASA-score op een specifiek meetmoment (d.i. de MASA-score van 1 maand tot 1 jaar voor de FOIS en de MASA-score van 3 maanden tot 1 jaar voor de Barthel index) van meer informatieve betekenis dan de baseline MASA-score bij het voorspellen van de voedingsstatus en de functionele onafhankelijkheid op dat tijdstip. De baseline MASA-score was een significante voorspeller van de verblijfplaats tijdens follow-up, zowel wanneer die beschouwd werd in combinatie met tijd als met de MASA-scores op de verschillende meetmomenten. De MASA op de verschillende meetmomenten was geen significante voorspeller van de verblijfplaats tijdens follow-up. De baseline MASA-score en leeftijd waren beide significante voorspellers van overleving na het corrigeren voor pneumonie en voor de aanwezigheid van een uitbreiding/recidief van het CVA tijdens de éénjarige follow-up. De baseline OHAT-score bereikte net geen significantie, wanneer het – in de plaats van de MASA – werd geïncludeerd in het statistisch model. De aanwezigheid van ernstige OD na CVA, als vastgesteld met behulp van de MASA, werd geassocieerd met slechtere overlevingskansen. Deze studie bevestigde voor CVA-patiënten dat het hebben van OD geassocieerd is met een slechte prognose.

Hoewel het niet gerapporteerd werd in Hoofdstuk III, bleek een van de outcomes van studie II te zijn dat CVA-patiënten met OD en met bijkomende taal- en/of cognitieve stoornissen vaak moeilijkheden vertoonden met het invullen van de Nederlandse versie van de SWAL-QoL (DSWAL-QoL). Om die reden hebben we in **STUDIE III** (cfr. Hoofdstuk IV) de aangepaste DSWAL-QoL (aDSWAL-QoL) ontwikkeld voor de populatie dysfagiepatiënten met bijkomende taal- en/of cognitieve stoornissen – de zogenaamde DysLC-patiënten – met de bedoeling zelfrapportering te faciliteren. De aDSWAL-QoL werd

ontwikkeld op basis van afasie- en cognitiefvriendelijke aanbevelingen vanuit de literatuur [55, 70, 71, 120-123]. Ook de haalbaarheid van de aDSWAL-QoL werd onderzocht, evenals zijn psychometrische eigenschappen volgens klassieke testtheorie (d.i. interne consistentie, test-hertest betrouwbaarheid en criteriumvaliditeit). Een tweede doel van **STUDIE III** was om de nood aan assistentie voor het invullen van de aDSWAL-QoL te voorspellen op basis van de variabelen cognitieve stoornis, taalbegrip, leeftijd, groep en functionele afhankelijkheid. De aDSWAL-QoL werd ontwikkeld aan de hand van een 5-stappen onderzoek (voor verdere informatie verwijzen we graag naar Hoofdstuk IV), die het volgende omvatte: een DSWAL-QoL adaptatieproces, het betrekken van een expert panel, initiële procesevaluaties, verdere verfijning van het instrument en voorbereiding van het eigenlijke onderzoek, en bijkomende procesevaluaties. Een cross-sectioneel onderzoek werd uitgevoerd bij 78 dysfagiepatiënten, waaronder 43 DysLC-patiënten. Voor de DysLC-groep was de haalbaarheid van de aDSWAL-QoL groter dan die van de DSWAL-QoL: minder assistentie was nodig, minder ontbrekende responsen en meer betrouwbare responsen kwamen voor. Ook tijdens het hertestmoment was het invullen van de aDSWAL-QoL gemakkelijker voor de DysLC-patiënten. Een hoge interne consistentie, excellente test-hertest overeenkomst en een goede criteriumvaliditeit werden vastgesteld voor de totale aDSWAL-QoL schaal. Bijna alle subschalen vertoonden een hoge interne consistentie en een matige tot goede test-hertest overeenkomst en criteriumvaliditeit. De psychometrische eigenschappen van de 'Food selection' subschaal waren echter ontoereikend. Weliswaar in mindere mate was dit ook het geval voor de Fear of eating en Eating desire subschalen. Om die reden moeten de resultaten van deze subschalen met enige voorzichtigheid geïnterpreteerd worden. Wat het tweede onderzoeksdoel van **STUDIE III** betrof, we vonden voor de DysLC-groep matige negatieve correlaties tussen de nood aan assistentie en de variabelen cognitieve stoornis, taalbegrip en functionele onafhankelijkheid. Dit benadrukt dat de nood aan assistentie toeneemt wanneer de functionele mogelijkheden verminderen. Op basis van de steekproef van deze studie konden we besluiten dat een MMSE $\leq$20, het behoren tot de DysLC-groep en een Barthel index <7 de nood aan assistentie bij het invullen van de aDSWAL-QoL kon voorspellen. Concluderend konden we stellen dat de aDSWAL-QoL een haalbaar, betrouwbaar en valide instrument is voor gebruik door DysLC-patiënten; bijgevolg kan het instrument gebruikt worden door een grote groep.

Een moderne psychometrische benadering, meer bepaald de Rasch methodologie, werd toegepast om de constructvaliditeit van de DSWAL-QoL en aDSWAL-QoL (verder) te onderzoeken. Dit model wordt beschouwd als de gouden standaard waartegen PRO-schalen, die de responsen van de

verschillende items samenvoegen, dienen getest te worden [80]. Het doel van **STUDIE IV** (cfr. Hoofdstuk V) was om de structurele validiteit (d.i. een aspect van constructvaliditeit) en objectiviteit van zowel de DSWAL-QoL als de aDSWAL-QoL totale schalen en subschalen te evalueren alsook de statistische doeltreffendheid van de totale scores en subschaal scores door gebruik te maken van item analyse met behulp van het Rasch model. De analyse van de items volgens het Rasch model werd uitgevoerd door middel van RUMM2030 software op eerdere data, verzameld vanuit een validatiestudie van 108 patiënten (waaronder 78 proefpersonen betrokken waren in studie III). Onderzoek werd uitgevoerd naar de algehele model fit, betrouwbaarheid, unidimensionaliteit, threshold ordering, individuele item en person fit, DIF, LID en targeting. De analyse kon de psychometrische eigenschappen noch van de totale schalen noch van hun subschalen aantonen: ze pasten niet in het Rasch model en multidimensionaliteit, gestoorde thresholds, DIF en/of LID werden gevonden. De betrouwbaarheid en power of fit waren hoog voor de totale schalen (PSI = 0.93), maar laag voor de meeste subschalen (PSI < 0.70). De targeting van personen en items was suboptimaal. De hoofdreden voor misfit waren de aanwezigheid van gestoorde thresholds voor zowel de totale schalen als voor de subschalen. Er werden aanpassingen, die gesuggereerd werden door de Rasch methodologie, geïmplementeerd: gestoorde thresholds werden herschaald, misfit items werden verwijderd en items werden opgesplitst voor DIF. Voor de multidimensionaliteit en LID werden geen oplossingen gevonden. De betrouwbaarheid en power of fit bleven laag voor de meeste subschalen. Dit was de eerste studie die een analyse met behulp van het Rasch model heeft uitgevoerd op de DSWAL-QoL en de aDSWAL-QoL. Op basis van deze studie werd besloten dat enige voorzichtigheid moet geboden worden bij het trekken van conclusies met betrekking tot dysfagie-gerelateerde levenskwaliteit op basis van de DSWAL-QoL en aDSWAL-QoL totale en subschaalscores totdat de structurele validiteit en objectiviteit van beide schalen worden vastgesteld. Een grote beperking van deze studie was de relatief beperkte steekproefomvang. De steekproefomvang van 108 patiënten in deze studie was binnen de aanbevolen omvang [200]. De effectieve steekproefgrootte (d.i. zonder proefpersonen met extreme scores) in deze studie was echter kleiner dan de originele steekproefomvang. Om die reden moet deze studie dan ook beschouwd worden als een pilootstudie. Correct gebalandeerde steekproeven die ≥ 250 mensen includeren [196, 200] worden aanbevolen om definitieve conclusies te kunnen trekken met betrekking tot de items en de schalen. We suggereren dus om momenteel de voorgestelde scoringstructuur (cfr. Hoofdstuk V) voor ieder individueel item te gebruiken totdat de psychometrische eigenschappen met behulp van een grotere steekproef worden

aangetoond. Bovendien is het belangrijk om alleen kwalitatieve informatie van de items af te leiden. Deze studie toonde de meerwaarde van een Rasch analyse aan bij het detecteren van de psychometrische sterktes en zwaktes van dergelijke beoordelingsschalen. Om die reden kan deze studie gezien worden als een essentiële stap vooruit bij de verdere verbetering van deze schalen.

Studies II, III en IV toonden aan dat (1) niet alle items van de DSWAL-QoL en aDSWAL-QoL geschikt waren voor patiënten die op enterale voeding aangewezen zijn en (2) dat zelfrapportering onmogelijk was bij patiënten met ernstige taal- en/of cognitieve stoornissen (MMSE < 5 en AAT < 37). Voortbouwend op deze bevindingen moeten we vooreerst concluderen dat enige voorzichtigheid dient geboden te worden bij het gebruik van de DSWAL-QoL en aDSWAL-QoL bij patiënten die op sondevoeding aangewezen zijn. Het gebruik van een ziektespecifieke PRO die specifiek ontwikkeld en gevalideerd werd voor deze patiëntenpopulatie is dan ook aanbevolen. Bij patiënten waar zelfrapportering onmogelijk is, kan waardevolle informatie over hun HRQoL verworven worden aan de hand van proxy-reported outcome measures (noteer dat proxies mensen zijn die in nauw verband staan tot de patient zoals famielieleden en partners). Om de meest geschikte PRO en proxy-reported outcome measure te kunnen selecteren bij het onderzoek naar de HRQoL van patiënten met enterale voeding is het belangrijk om de psychometrische eigenschappen en de klinische bruikbaarheid van deze meetinstrumenten te evalueren [58]. **STUDIE V** had betrekking op de ontwikkeling van een systematisch review protocol om de klinische bruikbaarheid en de psychometrische eigenschappen te beoordelen van (1) meetinstrumenten die HRQoL evalueren van patiënten die op enterale voeding aangewezen zijn, waarbij er geen rekening gehouden wordt met de reden voor enterale voeding, (2) meetinstrumenten die HRQoL evalueren van patiënten die als gevolg van dysfagie op enterale voeding aangewezen zijn en (3) meetinstrumenten (1 en 2) die door proxies van patiënten met enterale voeding vervolledigd worden om de HRQoL van dergelijke patiënten te kunnen evalueren. Dit systematisch review protocol (cfr. Hoofdstuk VI) geeft een gedetailleerde beschrijving weer van de doelen van de review, de inclusiecriteria, de zoekstrategie (inclusief de selectie van de studies en de evaluatie van hun methodologische kwaliteit) en de manier waarop de data-extractie en de datasynthese zullen gebeuren.

Samengevat, de conclusies van deze thesis waren:

1. De OHAT is een haalbaar en betrouwbaar instrument dat kan gebruikt worden door logopedisten om de mondgezondheid op een gestandaardiseerde manier te screenen (Hoofdstuk II).
2. De aanwezigheid van OD tijdens de eerste slikevaluatie post-CVA is van voorspellende betekenis voor het niet naar huis gaan na ontslag uit het ziekenhuis (Hoofdstuk III).
3. De baseline MASA- en OHAT-scores waren beide significante voorspellers van pneumonie, zowel tijdens hospitalisatie als tijdens follow-up (Hoofdstuk III).
4. De baseline MASA-score was een significante voorspeller van voedingsstatus en functionele onafhankelijkheid over de tijd heen, maar het effect van de baseline MASA was sterker op korte termijn dan op lange termijn. De baseline MASA was ook een significante voorspeller van de verblijfplaats tijdens follow-up; het effect van de baseline MASA was hetzelfde voor het gehele jaar (Hoodstuk III).
5. Wanneer we het effect van de baseline MASA gecombineerd met de MASA-scores op de opeenvolgende meetmomenten bestuderen, kunnen we het volgende concluderen: (1) de baseline MASA-score was een significante voorspeller van voedingsstatus, functionele onafhankelijkheid en verblijfplaats tijdens follow-up; (2) de MASA-score op een bepaald meetmoment was een significante voorspeller van voedingsstatus en functionele onafhankelijkheid en naarmate het jaar vorderde werd die van meer informatieve betekenis voor de voedingsstatus en functionele onafhankelijkheid dan de baseline MASA; (3) de MASA-score op een specifiek meetmoment was geen significante voorspeller van de verblijfplaats tijdens follow-up (Hoofdstuk III).
6. De baseline MASA was een significante voorpeller van overleving post-CVA. Ernstige OD na CVA, als vastgesteld door de MASA, werd geassocieerd met slechtere overlevingskansen. De OHAT was bijna significant als voorspeller van overleving wanneer het in plaats van de MASA geïncludeerd werd in het statistisch model (Hoofdstuk III).
7. De aDSWAL-QoL werd ontwikkeld, gebaseerd op afasie- en cognitiefvriendelijke aanbevelingen vanuit de literatuur (Hoofdstuk IV).

8. De aDSWAL-QoL is een haalbaar en – volgens de klassieke testtheorie – een betrouwbaar en valide instrument voor gebruik door DysLC-patiënten. De psychometrische eigenschappen van de Food selection en – hetzij in mindere mate – de psychometrische eigenschappen van de Fear of eating en Eating desire subschalen waren onvoldoende (Hoofdstuk IV).
9. De analyse aan de hand van het Rasch model kon noch de structurele validiteit en objectiviteit van de DSWAL-QoL en aDSWAL-QoL totale schalen en subschalen noch de statistische toereikendheid van de totale scores en subschaal scores vaststellen. De analyse met behulp van het Rasch model kon areas identificeren die verder onderzoek vereisen. Potentiële oplossingen, die door het Rasch model gesuggereerd werden, hebben tot schaalverbetering geleid. Tot de psychometrische eigenschappen van de schalen in een grotere en beter gebalanceerde steekproef worden vastgesteld, suggereren we om de voorgestelde scoringstructuur voor de items te hanteren en om alleen kwalitatieve informatie van de items af te leiden (Hoofdstuk V).
10. Er werd een systematisch review protocol ontwikkeld om systematisch de psychometrische eigenschappen en de klinische bruikbaarheid te beoordelen van PROs en proxy-reported outcome measures die HRQoL meten bij patiënten met enterale voeding om aanbevelingen te kunnen doen voor het gebruik ervan in de klinische praktijk en in onderzoek (Hoofdstuk VI).

7.2 TOEKOMSTPERSPECTIEVEN

In studie I hebben we de haalbaarheid en betrouwbaarheid van de OHAT voor gebruik door logopedisten vastgesteld. De OHAT werd geïmplementeerd in studie II om de mondgezondheid van CVA-patiënten met en zonder dysfagie op een gestandaardiseerde wijze te screenen. Toekomstig onderzoek is nodig om te evalueren of het implementeren van de OHAT binnen een klinisch diagnostisch slikonderzoek het belang van mondzorg kan promoten en het mondzorgbeleid in de dagdagelijkse klinische praktijk kan verbeteren. Na het aanbrengen van nog een paar kleine aanpassingen (bv. het voorzien van ruimte op het scoreformulier waar paren van kiezen kunnen gedocumenteerd worden), is het instrument bruikbaar in de klinische praktijk. Een beperking van studie I was dat we de concurrente (d.i. criteriumvaliditeit) validiteit

van de OHAT niet onderzocht hebben, wat betekent dat de OHAT-resultaten die logopedisten bekomen zouden vergeleken moeten worden met resultaten die tandartsen bekomen op basis van een grondig tandartsonderzoek. Onderzoek naar concurrente validiteit van de OHAT is wenselijk om de diagnostische accuraatheid van het instrument aan te tonen. Verder zou het ook interessant zijn om meer te differentiëren in de verwijzingsstrategieën. De strategieën die volgen op een OHAT-afname zouden bovendien verder verfijnd kunnen worden. De waarde van de OHAT bij het voorspellen van pneumonie en overleving werd in studie II bestudeerd. Echter, verder onderzoek is vereist. Pneumonie is een multifactorieel probleem en daarom is het herhalen van een binaire logistische regressie-analyse met inclusie van meer klinische variabelen die een mogelijke voorspeller van pneumonie kunnen zijn nodig. We zullen in de toekomst ook rapporteren over de waarde van de OHAT bij het voorspellen van de verblijfplaats, voedingsstatus en functionele onafhankelijkheid. Zoals beschreven in studie II (cfr. Hoofdstuk III), werden omwille van praktische redenen het onderzoek van de slikfunctie en het bepalen van het type dieet niet herhaald op het moment van ontslag uit het ziekenhuis. In toekomstige studies zou het interessant zijn om deze onderzoeken ook op dit tijdstip (d.i. bij ontslag uit het ziekenhuis) te herhalen om te bepalen in hoeverre deze parameters de verblijfplaats na ontslag en het verdere zorgpad kunnen beïnvloeden. Via studie II werd meer data verzameld dan dat we gerapporteerd hebben in Hoofdstuk III. We zullen die data (bv. analyse van dysfagiegerelateerde levenskwaliteit en mondgezondheidgerelateerde levenskwaliteit van CVA-patiënten) verder analyseren en rapporteren. In studie III hebben we de aDSWAL-QoL ontwikkeld voor het gebruik door een grote populatie DysLC-patiënten. Echter, het zou interessant zijn te onderzoeken of die patiënten voordeel zouden halen uit de situatie waarin ze verdeeld worden in verdere subgroepen (d.i. een groep die bijvoorbeeld alleen taalstoornissen vertoont) waarbij het instrument wordt aangepast aan hun specifieke noden. Verder onderzoek naar het omzetten van de pen-en-papier-methode voor het invullen van de aDSWAL-QoL in een meer geautomatiseerde, technologische toepassing is ook zinvol. Dit laatste zou er immers voor kunnen zorgen dat de patiënt minder beroep moet doen op anderen om assistentie te verkrijgen bij het invullen van de vragenlijst. Er werd eveneens een proxyversie van de DSWAL-QoL ontwikkeld en een preliminaire analyse met betrekking tot een onderzoek naar de patient-proxy overeenkomst inzake dysfagiegerelateerde HRQoL werd reeds uitgevoerd. We zullen die data verder analyseren en rapporteren. In studie III hebben we omwille van validiteitscriteria alleen die patiënten geïncludeerd die dysfagie vertoonden voor ten minste één maand en niet diegene die zich in de acute (en/of subacute) fase post-CVA bevonden. Als de toestand

van de patiënt dit toelaat, zou het interessant zijn om dysfagiegerelateerde HRQoL te onderzoeken in de acute (en/of subacute) fase post-CVA en na te gaan of de aDSWAL-QoL al tijdens die fase kan toegepast worden. Studie IV omvatte een steekproef van 108 patiënten om item analyse met behulp van het Rasch model te kunnen uitvoeren op de DSWAL-QoL en aDSWAL-QoL. Hoewel die steekproefgrootte zich binnen de aanbevolen omvang bevond [200], was de effectieve steekproefgrootte kleiner door de aanwezigheid van extreme scores. Om die reden moest studie IV beschouwd worden als een pilootstudie. Om definitieve conclusies met betrekking tot de items en de schalen te kunnen trekken, wordt een grote empirische studie aanbevolen. We bevelen IRT aan voor de verdere validering van de originele SWAL-QoL en al zijn vertalingen. Het zou bovendien gunstig zijn om kortere versies van de DSWAL-QoL en aDSWAL-QoL te creëren en te valideren. We voeren momenteel een systematische review uit naar de psychometrische eigenschappen en klinische bruikbaarheid van PROs en proxy-reported outcome measures die HRQoL onderzoeken bij patiënten met enterale voeding omdat we gemerkt hebben dat sommige items van de DSWAL-QoL en aDSWAL-QoL vragenlijsten niet specifiek genoeg waren voor een dergelijke populatie en omdat zelfrapportering onmogelijk was door patiënten met ernstige taal- en/of cognitieve stoornissen. Omdat het zo belangrijk is om patiënten te betrekken in het ontwikkelen en valideren van PROs [242] en gezien de zwaktes van bestaande schalen (cfr. studie III en IV), hebben we beslist om onze eigen ziektespecifieke PRO te ontwikkelen voor patiënten die als gevolg van orofaryngeale dysfagie op enterale voeding aangewezen zijn. Dit onderzoek is lopende.

CHAPTER VIII

SWOT ANALYSIS

In the following tables, we will summarize the main strengths, weaknesses, opportunities and threats of the performed research, as presented in this thesis.

S	**MAIN STRENGTHS**
	The development and/or psychometric evaluation of tools (i.e., aDSWAL-QoL and OHAT) that can be used in daily clinical practice and research
	Multi-centre research trials
	The prospective characteristic of the studies
	Evaluation of long-term outcome of stroke patients with frequent follow-up
	Relatively large sample sizes
	The inclusion of patient groups that are at times excluded from HRQoL measurement studies
	Detection of psychometric weaknesses in validated clinical tools (i.e., DSWAL-QoL and aDSWAL-QoL) by the introduction of a modern, innovative psychometric approach (i.e., Rasch measurement model)
	A first, specific solution for scale improvement of DSWAL-QoL and aDSWAL-QoL, as induced by the adjustments suggested by the Rasch model
W	**MAIN WEAKNESSES**
	The missing values –which are inherent to multi-centre studies– related to data that should have been provided by professionals that were not directly involved in the research team
	Concurrent validity (i.e., criterion validity) of the OHAT for use by SLPs was not established
	The limited number of raters involved in the reliability study of the OHAT
	Although the sample sizes in the different studies were within the recommended sample sizes (i.e., based on a sample size calculation or on internationally qualified criteria), the effective sample size (e.g., in study IV) was sometimes smaller
	Equally balanced and more homogeneous groups could have benefit the results
	Convenience sampling in most of the studies
	We did not use a cognitive assessment battery to evaluate the cognitive abilities, but a screening tool (i.e., MMSE)
O	**MAIN OPPORTUNITIES**
	Further refinement (e.g., incorporating additional illustrations in the publicly available visual training resources) and validation of the OHAT, and evaluating if the implementation of this tool in a clinical swallowing examination would promote oral health care awareness and improve oral care management in daily clinical practice
	Facilitating the implementation of DSWAL-QoL and aDSWAL-QoL in clinical practice and increasing the feasibility for a broad population (e.g. by conversion of the PRO into a computer-assisted self-administered format)
	Validation of the proxy version of the DSWAL-QoL
	Further analysis of collected data (i.e., patient-proxy agreement for DSWAL-QOL and aDSWAL-QOL; data related to study II (e.g., analysis of the dysphagia-related HRQoL in stroke patients))
	Finalizing the systematic review on the clinical utility and psychometric properties of existing PROs and proxy-reported outcome measures among patients receiving enteral feeding
	Further development and validation of a disease-specific PRO for patients receiving enteral feeding as a consequence of oropharyngeal dysphagia
T	**MAIN THREATS**
	Financial and time constraints may tax the further development, refining, and validation of clinical tools
	The implementation of predictive models in clinical practice is limited and should always be balanced against the ethical issues in (stroke) care

REFERENCES

1. Baroni AF, Fábio SR, Dantas RO. Risk factors for swallowing dysfunction in stroke patients. Arq Gastroenterol. 2012;49(2):118-24. http://www.scielo.br/pdf/ag/v49n2/05.pdf. Accessed May 2017.
2. Cohen DL, Roffe C, Beavan J, Blackett B, Fairfield CA, Hamdy S, et al. Post-stroke dysphagia: A review and design considerations for future trials. Int J Stroke. 2016;11(4):399-411. doi:10.1177/1747493016639057.
3. European Registers of Stroke (EROS) Investigators, Heuschmann PU, Di Carlo A, Bejot Y, Rastenyte D, Ryglewicz D, et al. Incidence of Stroke in Europe at the Beginning of the 21ste Century. Stroke. 2009;40(5):1557-63.
4. Katsnelson MJ, Della-Morte D, Rundek T. Stroke in young. Periodicum biologorum. 2012;114(3):347-53. File:///C:/Users/Notebook/Downloads/katsnelson_dellamorte_rundek_347_353%20(3).pdf. Accessed Septembre 2017.
5. Zhang Y, Chapman AM, Plested M, Jackson D, Purroy F. The incidence, prevalence, and mortality of stroke in France, German, Italy, Spain, the UK, and the US: A Literature Review. Stroke Res. Treat. 2012;2012:436125. doi:10.1155/2012/436125.
6. Intego. Registratienetwerk van huisartsen in Vlaanderen. http://www.intego.be. Accessed August 2017.
7. Buntinx F, Devroey D, Van Casteren V. The incidence of stroke and transient ischaemic attacks is falling: a report from the Belgian sentinel stations. Br J Gen Pract. 2002;52(483):813-7. http://www.europepmc.org/articles/pmc1316084/pdf/12392121.pdf. Accessed Septembre 2017.
8. Feigin VL, Lawes CMM, Bennett DA, Anderson CS. Stroke epidemiology: a review of population-based studies of incidence, prevalence, and case-fatality in the late 20th century. Lancet Neurol. 2003;2(1):43-53.
9. Devroey D, Van Casteren V, Buntinx F. De epidemiologie van cerebrovasculaire incidenten in België. Registratie door het Belgische netwerk van huisartsenpeilpraktijken in 1998 en 1999. IPH/EPI Reports. 2003. https://www.wiv-isp.be/Epidemio/epinl/medvnl/cva_9899_n.pdf. Accessed Septembre 2017.

10. Devroey D, Van Casteren V, Buntinx F. Registration of stroke through the Belgian sentinel network and factors influencing stroke mortality. Cerebrovasc Dis. 2003;16(3):272-9.
11. Edwardson MA, Dromerick AW. Ischemic stroke prognosis in adults. http://www.uptodate.com/contents/ischemic-stroke-prognosis-in-adults. Accessed April 2017.
12. Arnold M, Liesirova K, Broeg-Morvay A, Meisterernst J, Schlager M, Mono ML, et al. Dysphagia in acute stroke: incidence, burden and impact on clinical outcome. PLoS One. 2016;11(2). doi:10.1371/journal.pone.0148424.
13. Kernan WN, Obviagele B, Black HR, Bravata DM, Chimowitz MI, Ezekowitz MD, et al. Guidelines for the prevention of stroke in patients with stroke and transient ischemic attack: a guideline for healthcare professionals from the American Heart Association/American Stroke Association. Stroke. 2014;45(7):2160-236. doi:10.1161/STR.0000000000000024.
14. Maeshima S, Osawa A, Miyazaki Y, Seki Y, Miura C, Tazawa Y, et al. Influence of dysphagia on short-term outcome in patients with acute stroke. Am J Phys Med Rehabil. 2011;90(4):316-20.
15. Smithard DG, Smeeton NC, Wolfe CD. Long-term outcome after stroke: does dysphagia matter? Age Ageing. 2007;36(1):90-4. doi: 10.1093/ageing/afl149.
16. Martino R, Pron G, Diamant N. Screening for oropharyngeal dysphagia in stroke: insufficient evidence for guidelines. Dysphagia. 2000;15(1):19-30. doi:10.1007/s004559910006.
17. Langdon PC, Lee AH, Binns CW. Dysphagia in acute ischaemic stroke: severity, recovery and relationship to stroke subtype. J Clin Neurosci. 2007;14(7):630-4.
18. Martino R, Foley N, Bhogal S, Diamant N, Speechley M, Teasell R. Dysphagia after stroke. Incidence, Diagnosis, and Pulmonary Complications. Stroke. 2005;36:2756-63. doi:10.1161/01.STR.0000190056.76543.eb.
19. Ickenstein GW, Stein J, Ambrosi D, Goldstein R, Horn M, Bogdahn U. Predictors of survival after severe dysphagic stroke. J Neurol. 2005;252(12):1510-6.
20. Armstrong JR, Mosher BD. Aspiration pneumonia after stroke. Intervention and prevention. Neurohospitalist. 2011;1(2):85-93. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3726080/. Accessed April 2017.
21. Bonilha HS, Simpson AN, Ellis C, Mauldin P, Martin-Harris B, Simpson K. The one-year attributable cost of post-stroke dysphagia. Dysphagia. 2014;29(5):545-552.

22. Langmore SE, Terpenning MS, Schork A, Chen Y, Murray JT, Lopatin D, et al. Predictors of aspiration pneumonia: how important is dysphagia? Dysphagia. 1998;13(2):69-81.
23. Marik PE, Kaplan D. Aspiration pneumonia and dysphagia in the elderly. Chest. 2003;124(1):328-36.
24. Ortega O, Parra C, Zarcero S, Nart J, Sakwinska O, Clavé P. Oral health in older patients with oropharyngeal dysphagia. Age Ageing. 2014;43:132–37. doi:10.1093/ageing/aft164.
25. Terpenning M. Geriatric oral health and pneumonia risk. Clin Infect Dis. 2005;40(12):1807-10. doi:10.1086/430603.
26. Tada A, Miura H. Prevention of aspiration pneumonia (AP) with oral care. Arch Gerontol Geriatr. 2012;55(1):16-21.
27. Yoon MN, Steele CM. Health care professionals' perspectives on oral care for long-term care residents: nursing staff, speech-language pathologists and dental hygienists. Gerodontology. 2012;29(2):e525-35. doi:10.1111/j.1741-2358.2011.00513.x.
28. van der Maarel-Wierink CD, Vanobbergen JN, Bronkhorst EM, Schols JM, de Baat C. Risk factors for aspiration pneumonia in frail older people: a systematic literature review. J Am Med Dir Assoc. 2011;12(5):344-54.
29. Yoneyama T, Yoshida M, Ohrui T, Mukaiyama H, Okamoto H, Hoshiba K, et al. Oral care reduces pneumonia in older patients in nursing homes. J Am Geriatr Soc. 2002; 50(3):430-3.
30. Bassim CW, Gibson G, Ward T, Paphides BM, Denucci DJ. Modification of the risk of mortality from pneumonia with oral hygiene care. J Am Geriatr Soc. 2008;56(9):1601-7.
31. Shay K, Ship JA. The importance of oral health in the older patient. J Am Geriatr Soc. 1995;43(12):1414-22.
32. Logemann JA, Curro FA, Pauloski B, Gensler G. Aging effects on oropharyngeal swallow and the role of dental care in oropharyngeal dysphagia. Oral Dis. 2013; 19(8):733-7.
33. Ueno M, Yanagisawa T, Shinada K, Ohara S, Kawaguchi Y. Masticatory ability and functional tooth units in Japanese adults. J Oral Rehabil. 2008;35(5):337-44.
34. Hildebrandt GH, Dominguez BL, Schork MA, Loesche WJ. Functional units, chewing, swallowing, and food avoidance among the elderly. J Prosthet Dent. 1997;77(6):588-95.
35. N'Gom PI, Woda A. Influence of impaired mastication on nutrition. J Prosthet Dent. 2002;87(6):667-73.

36. Papas AS, Joshi A, Giunta JL, Palmer CA. Relationships among education, dentate status, and diet in adults. Spec Care Dentist. 1998;18(1): 26-32.
37. Papas AS, Palmer CA, Rounds MC, Russell RM. The effects of denture status on nutrition. Spec Care Dentist. 1998;18(1): 17-25.
38. World Health Organization. Programme on mental health: WHOQOL user manual World Health Organization: Geneva; 1998; 3. Revision 2012. http://www.who.int/iris/handle/10665/77932. Accessed December 2014.
39. Romero M, Vivas-Consuelo D, Alvis-Guzman N. Is Health Related Quality of Life (HRQoL) a valid indicator for health systems evaluation? SpringerPlus. 2013;2(1):664. doi:10.1186/2193-1801-2-664.
40. Skevington SM, Lotfy M, O'Connell KA. WHOQOL Group. The World Health Organization's WHOQOL-BREF quality of life assessment: psychometric properties and results of the international field trial. A report from the WHOQOL group. Qual Life Res. 2004;13(2):299-310.
41. Brotherton A, Abbott J, Aggett P. The impact of percutaneous endoscopic gastrostomy feeding upon daily life in adults. J Hum Nutr Diet. 2006;19(5):355-67.
42. Testa MA, Simonson DC. Assessments of quality-of-life outcomes. N Engl J Med. 1996;334(13):835-40.
43. Barnett DB. Assessment of quality of life. Am J Cardiol. 1991;67(12):41C-4C.
44. Brotherton AM, Judd PA. Quality of life in adult enteral tube feeding patients. J Hum Nutr Diet. 2007;20(6):513-22.
45. Deshpande PR, Rajan S, Sudeepthi BL, Nazir CPA. Patient-reported outcomes: a new era in clinical research. Perspect Clin Res 2011;2(4):137-44. doi:10.4103/2229-3485.86879.
46. Patrick D, Erickson P. Health policy, quality of life: health care evaluation and resource allocation. In Romero M, Vivas-Consuelo D, Alvis-Guzman N. Is Health Related Quality of Life (HRQoL) a valid indicator for health systems evaluation? SpringerPlus. 2013;11(2):664. doi:10.1186/2193-1801-2-664.
47. Sneeuw KC, Sprangers MA, Aaronson NK. The role of health care providers and significant others in evaluating the quality of life of patients with chronic disease. J Clin Epidemiol. 2002;55(11):1130-43.

48. Vanderwegen J, Van Nuffelen G, De Bodt M. The validation and psychometric properties of the Dutch version of the Swallowing Quality-of-Life Questionnaire (SWAL-QoL). Dysphagia. 2013;28(1):11-23.
49. Zraick RI, Atcherson SR, Ham BK. Readability of patient-reported outcome questionnaires for use with persons with swallowing disorders. Dysphagia. 2012;27(3):346–52.
50. de Haan R, Aaronson N, Limburg M, Hewer RL, van Crevel H. Measuring quality of life in stroke. Stroke. 1993;24(2): 320–27.
51. Lemmens J, Bours GJ, Limburg M, Beurskens AJ. The feasibility and test-retest reliability of the Dutch Swal-QoL adapted interview version for dysphagic patients with communicative and/or cognitive problems. Qual Life Res. 2013;22(4):891-5.
52. Timmerman AA, Speyer R, Heijnen BJ, Klijn-Zwijnenberg IR. Psychometric characteristics of health-related quality-of-life questionnaires in oropharyngeal dysphagia. Dysphagia. 2014;29(2):183-98.
53. US Department of Health and Human Services, Food and Drug Administration. Guidance for industry patien t-reported outcome measures: use in medical product development to support labeling claims. 2009. http://www.fda.gov/downloads/Drugs/Guidances/ UCM193282.pdf. Accessed August 2015.
54. Hilari K, Byng S, Lamping DL, Smith SC. Stroke and Aphasia Quality of Life Scale-39 (SAQOL-39): evaluation of acceptability, reliability, and validity. Stroke. 2003;34(8):1944-50. doi:10.1161/01.STR.0000081987.46660.ED.
55. Chue WL, Rose ML, Swinburn K. The reliability of the communication disability Profile: a patient-reported outcome measure for aphasia. Aphasiology. 2010;24(6–8):940-56.
56. Orlandoni P, Jukic PN. Health-related quality of life and functional health status questionnaires in oropharyngeal dysphagia. J Aging Res Clin Pract. 2016;5(1):31-7. http://dx.doi.org/10.14283/jarcp.2016.87.
57. Finizia C, Rudberg I, Bergqvist H, Rydén A. A cross-sectional validation study of the Swedish version of SWAL-QoL. Dysphagia. 2012;27(3):325-35.
58. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of

health status measurement instruments: an international Delphi study. Qual Life Res 2010;19(4):539-49.

59. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol 2007;60(1):34-42.
60. Glad J, Jergeby U, Gustafsson C, Sonnander K. Social work practitioners' experience of the clinical utility of the home observation for measurement of the environment (HOME) Inventory. Child Fam Soc Work. 2012;17(1):23-33.
61. McHorney CA, Bricker DE, Kramer AE, Rosenbek JC, Robbins J, Chignell KA, et al. The SWAL-QOL outcomes tool for oropharyngeal dysphagia in adults: I. Conceptual foundation and item development. Dysphagia. 2000;15(3):115-21.
62. McHorney CA, Bricker DE, Robbins J, Kramer AE, Rosenbek JC, Chignell KA. The SWAL-QOL outcomes tool for oropharyngeal dysphagia in adults: II. Item reduction and preliminary scaling. Dysphagia. 2000;15(3):122-33.
63. McHorney CA, Robbins J, Lomax K, Rosenbek JC, Chignell K, Kramer AE, et al. The SWAL-QOL and SWAL-CARE outcomes tool for oropharyngeal dysphagia in adults: III. Documentation of reliability and validity. Dysphagia. 2002;17(2):97-114.
64. Speyer R, Heijnen BJ, Baijens LW, Vrijenhoef FH, Otters EF, Roodenburg N, et al. Quality of life in oncological patients with oropharyngeal dysphagia: validity and reliability of the Dutch version of the MD Anderson Dysphagia Inventory and the Deglutition Handicap Index. Dysphagia. 2011;26(4):407-14.
65. de Wit MPT, Kvien TK, Gossec L. Patient participation as an integral part of patient-reported outcomes development ensures the representation of the patient voice: a case study from the field of rheumatology. RMD Open. 2015;1:e000129. doi:10.1136/ rmdopen-2015-000129.
66. Falsetti P, Acciai C, Palilla R, Bosi M, Carpinteri F, Zingarelli A, et al. Oropharyngeal dysphagia after stroke: incidence, diagnosis, and clinical predictors in patients admitted to a neurorehabilitation unit. J Stroke Cerebrovasc Dis. 2009;18(5):329-35.
67. Daniels SK. Neurological disorders affecting oral, pharyngeal swallowing. GI Motility online. 2006. doi:10.1038/gimo34.

68. Pedersen PM, Vinter K, Olsen TS. Aphasia after stroke: type, severity and prognosis. The Copenhagen aphasia study. Cerebrovasc Dis. 2004;17:35-43. doi:10.1159/000073896.
69. McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jack CR, Kawas CH, et al. The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement. 2011;7(3):263-9. http:// dx.doi.org/10.1016/j.jalz.2011.03.005.
70. Dalemans RJ, de Witte LP, Beurskens AJ, van den Heuvel WJ, Wade DT. Psychometric properties of the community integration questionnaire adjusted for people with apahasia. Arch Phys Med Rehabil. 2010;91(3):395-9.
71. Logsdon RG, Gribbons LE, McCurry S, Linda T. Assessing quality of life in older adults with cognitive impairment. Psychosom Med. 2002;64(3):510-9.
72. Sneeuw KCA, Aaronson NK, de Haan RJ, Limburg M. Assessing quality of life after stroke. The value and limitations of proxy ratings. Stroke. 1997;28:1541-9.
73. Spaccavento S, Craca A, Del Prete M, Falcone R, Colucci A, Di Palma A, et al. Quality of life measurement and outcome in aphasia. Neuropsychiatr Dis Treat. 2014;10:27–37. doi:10.2147/ NDT.S52357.
74. Hepworth LR, Rowe FJ, Harper R, Jarvis K, Shipman T, Rodgers H. Patient reported outcome measures for visual impairment after stroke: a systematic review. Health Qual Life Outcomes. 2015;13:146. doi:10.1186/s12955-015-0338-x.
75. Bowling A. Mode of questionnaire administration can have serious effects on data quality. J Public Health (Oxf). 2005;27(3):281-91. doi:10.1093/pubmed/fdi031.
76. Kroll T, Wyke S, Jahagirdar D, Ritchie K. If patient-reported outcome measures are considered key health-care quality indicators, who is excluded from participation? Health Expect. 2014;17(5):605-7.
77. Cella D, Hahn EA, Jensen SE, Butt Z, Nowinski J, Rothrock N. Methodological issues in the selection, administration and use of patient-reported outcomes in performance measurement in health care settings: PRO commissioned paper. Washington, DC: National Quality Forum (NQF); 2012. file:///C:/Users/Notebook/Downloads/PRO_Commissioned-Paper-1%20(1).pdf. Accessed December 2015.

78. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol. 2010;63(7):737-45.
79. Hagquist C, Bruce M, Gustavsson JP. Using the Rasch model in nursing research: an introduction and illustrative sample. Int J Nurs Stud. 2009;46(3):380-93.
80. Kreiner S. Validity and objectivity: reflections on the role and nature of Rasch models. Nordic Psychol. 2007;59(3):268-98.
81. Tennant A, Conaghan PG. The Rasch measurement model in rheumatology: What is it and why use it? When should it be applied, and what should one look for in a Rasch paper? Arthritis Rheum. 2007;57(8):1358-62.
82. Hagell P. Testing rating scale unidimensionality using the principal component analysis (PCA)/t-test protocol with the Rasch model: the primacy of theory over statistics. Open J Stat. 2014;4:456-65. doi:10.4236/ojs.2014.46044.
83. de Vet HCW, Terwee CB, Mokkink LB, Knol DL. Measurements in medicine. A practical guide. 1st ed. Cambridge: Cambridge University Press; 2011.
84. Novella JL, Jochum C, Jolly D, Morrone I, Ankri J, Bureau F, et al. Agreement between patients' and proxies' reports of quality of life in Alzheimer's disease. Qual Life Res. 2001;10(5):443-52.
85. Sneeuw KC, Aaronson NK, Sprangers MA, Detmar SB, Wever LD, Schornagel JH. Comparison of patient and proxy EORTC QLQ-C30 ratings in assessing the quality of life of cancer patients. J Clin Epidemiol. 1998;51(7):617-31.
86. Varni JW, Limbers CA, Burwinkle TM. Parent proxy-report of their children's health-related quality of life: an analysis of 13,878 parents' reliability and validity across age subgroups using the PEDsQL^{TM} 4.0 Generic Core Scales. Health Qual Life Outcomes. 2007;5:2.
87. Crary MA, Carnaby GD, Sia I, Khanna A, Waters M. Spontaneous Swallowing Frequency [has potential to] identify dysphagia in acute stroke. Stroke. 2013; 44(12):3452-57. doi:10.1161/STROKEAHA.113.003048.
88. Mann G. The Mann Assessment of Swallowing Ability. Clifton Park: Thomson Delmar Learning; 2002.

89. Crary MA, Carnaby-Mann GD, Miller L, Antonios N, Silliman S. Dysphagia and nutritional status at the time of hospital admission for ischemic stroke. J Stroke Cerebrovasc Dis. 2006; 15(4):164-71. doi:10.1016/j.jstrokecerebrovasdis.2006.05.006.

90. Chalmers JM, King PL, Spencer AJ, Wright FA, Carter KD. The oral health assessment tool – validity and reliability. Aust Dent J. 2005;50(3):191-9.

91. College of Audiologists & Speech-Language Pathologists of Ontario. CASLPO PSG practice standards and guidelines for dysphagia intervention by speech language pathologists, approved September 2007; 2007. http://www.caslpo.com/sites/default/uploads/files/PSG_EN_Dysphagia.pdf. Accessed January 2015.

92. Heylen L, Blux T. Gezondheidswetgeving en sociale zekerheid voor logopedisten. Antwerpen: Garant; 2008.

93. American Speech-Language-Hearing Association. Scope of practice in speech-language pathology [Scope of Practice]; 2007. http://www.asha.org/policy. Accessed July 2015.

94. American Speech-Language-Hearing Association. Knowledge and skills needed by speech-language pathologists providing services to individuals with swallowing and/or feeding disorders [Knowledge and Skills]; 2002. http://www.asha.org/policy. Accessed July 2015.

95. McCullough GH, Martino R. Clinical evaluation of patients with dysphagia: importance of history taking and physical exam. Available from: http://www.springer.com/cda/content/document/cda_downloaddocument/9781461437789-c1.pdf. Accessed July 2015.

96. McCullough GH, Wertz RT, Rosenbek JC. Sensitivity and specificity of clinical/bedside examination signs for detecting aspiration in adults subsequent to stroke. J Commun Disord. 2001;34(1-2):55-72.

97. Pearson A, Chalmers J. Oral hygiene care for adults with dementia in residential aged care facilities. JBI Reports. 2004; 2(3):65-113.

98. Lin CY, Jones DB, Godwin K, Godwin RK, Knebl JA, Niessen L. Oral health assessment by nursing staff of Alzheimer's patients in a long-term-care facility. Spec Care Dentist. 1999;19(2):64-71.

99. Chalmers J, Pearson A. Oral hygiene care for residents with dementia: a literature review. J Adv Nurs. 2005;52(4):410-9.

100. Kayser-Jones J, Bird WF, Paul SM, Long L, Schell ES. An instrument to assess the oral health status of nursing home residents. Gerontologist. 1995;35:814-24.

101. Dickinson H, Watkins C, Leathley M. The development of THROAT: the holistic and reliable oral assessment tool. Clin. Eff Nurs. 2001; 5:104-10.

102. Andersson P, Persson L, Hallberg IR, Renvert S. Testing an oral assessment guide during chemotherapy in a Swedish care setting: a pilot study. J Clin Nurs. 1999;8:150-58.

103. Halton Region's Health Department. Oral Health Assessment Tool for long-term care. http://www.halton.ca. Accessed December 2014.

104. Regional Geriatric Program Central. Introducing the Oral Health Assessment Tool. http://www.rgpc.ca. Accessed December 2014.

105. Crary MA, Carnaby GD, Shabbir Y, Miller L, Silliman S. Clinical variables associated with hydration status in acute ischemic stroke patients with dysphagia. Dysphagia. 2016; 31(1):60-5.

106. Nakajima M, Takada T, Terasaki Y, Nagano L, Naritomi H, Minematsu K. Clinical significance of oral intake in patients with acute stroke. Dysphagia. 2010;25(3):192-7.

107. Daniels SK, Ballo LA, Mahoney MC, Foundas AL. Clinical predictors of dysphagia and aspiration risk: outcome measures in acute stroke patients. Arch Phys Med Rehabil. 2000;81(8):1030-3.

108. Wilkinson TJ, Thomas K, MacGregor S, Tillard G, Wyles C, Sainsbury R. Tolerance of early diet textures as indicators of recovery from dysphagia after stroke. Dysphagia. 2002; 17(3):227-32.

109. James A, Kapur K, Hawthorne AB. Long-term outcome of percutaneous endoscopic gastrostomy feeding in patients with dysphagic stroke. Age ageing. 1998;27(6):671-6. doi:10.1093/ageing/27.6.671.

110. Johnson ER, McKenzie SW, Sievers A. Aspiration pneumonia in stroke. Arch Phys Med Rehabil. 1993;74(9):973-6. http://www.archives-pmr.org/article/0003-9993(93)90277-H/pdf. Accessed April 2017.

111. Daniels SK, Anderson JA, Willson PC. Valid items for screening dysphagia risk in patients with stroke: a systematic review. Stroke. 2012;43(3): 892-7.

https://pdfs.semanticscholar.org/aa5a/3009c83b497918127f6696557658d142b706.pdf. Accessed April 2017.

112. Martino R, Pron G, Diamant N. Screening for oropharyngeal dysphagia in stroke: insufficient evidence for guidelines. Dysphagia. 2000;15(1):19-30. doi:10.1007/s004559910006.

113. Vanderwegen J, Guns C, Van Nuffelen G, De Bodt M, Van de Heyning P, Carnaby GD. The reliability of the MASA dysphagia screening protocol compared to FEES for patients in an acute stroke unit. Poster presented at the Dysphagia Research Society, Phoenix Arizona, March 2006.

114. González-Fernández M, Ottenstein L, Atanelov L, Christian AB. Dysphagia after stroke: an overview. Curr Phys Med Rehabil Rep. 2013;1(3):187-96. doi:10.1007/s40141-013-0017-y.

115. Maccarini AR, Filippini A, Padovani D, Limarzi M, Loffredo M, Casolino D. Clinical non-instrumental evaluation of dysphagia. Acta Otorhinolaryngol Ital. 2007; 27(6):299-305. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2640056/. Accessed May 2017.

116. Kissela B, Lindsell CJ, Kleindorfer D, Alwell K, Moomaw CJ, Woo D, et al. Clinical prediction of functional outcome after ischemic stroke: the surprising importance of periventricular white matter disease and race. Stroke. 2009;40(2):530-6. doi:10.1161/STROKEAHA.108.521906.

117. Simpelaere IS, Van Nuffelen G, Vanderwegen J, Wouters K, De Bodt M. Oral health screening: feasibility and reliability of the oral health assessment tool as used by speech pathologists. Int Dent J. 2016;66(3):178-89.

118. Khaldoun E, Woisard V, Verin E. Validation in French of the SWAL-QOL scale in patients with oropharyngeal dysphagia. Gastroenterol Clin Biol. 2009;33(3):167-71.

119. Lam PM, Lai CKY. The validation of the Chinese version of the Swallow Quality-of-Life Questionnaire (SWAL-QOL) using exploratory and confirmatory factor analysis. Dysphagia. 2011;26(2):117-24.

120. Logsdon RG, Gibbons LE, McCurry SM, Teri L. Quality of life in Alzheimer's Disease: patient and caregiver reports. J Ment Health Aging. 1999;5(1):21-32.

121. Hilari K, Byng S. Measuring quality of life in people with aphasia: the Stroke Specific Quality of Life Scale. Int J Lang Commun Disord. 2001;36(suppl):86-91.

122. Worrall L, Rose T, Howe T, Brennan A, Egan J, Oxenham D, et al. Access to written information for people with aphasia. Aphasiology. 2005;19:923-9.

123. Boers J. Measuring Quality of Life in Aphasics – Development of the QoL-Aphasia 28 Items. Theses: Utrecht University; 2009.

124. Simpelaere IS, Vanderwegen J, Wouters K, De Bodt M, Van Nuffelen G. Feasibility and Psychometric Properties of the Adjusted DSWAL-QoL Questionnaire for Dysphagic Patients with Additional Language and/or Cognitive Impairment: Part I. Dysphagia. 2017;32(3):401-19.

125. Hendriks J, Fyfe S, Styles I, Skinner SR, Merriman G. Scale construction utilising the Rasch unidimensional measurement model: A measurement of adolescent attitudes towards abortion. Australas Med J. 2012;5(5):251-61. doi:10.4066/ AMJ.2012.952.

126. Petrillo J, Cano SJ, McLeod LD, Coon CD. Using classical test theory, item response theory, and Rasch measurement theory to evaluate patient-reported outcome measures: a comparison of worked examples. Value Health. 2015; 18(1):25-34. doi:10.1016/j.jval.2014.10.005.

127. Lo C, Liang WM, Hang LW, Wu TC, Chang YJ, Chang CH. A psychometric assessment of the St. George's respiratory questionnaire in patients with COPD using Rasch model analysis. Health Qual Life Outcomes. 2015;13:131. doi:10.1186/s12955-015-0320-7.

128. Chang YJ, Liang WM, Wu HC, Lin HC, Wang JY, Li TC, et al. Psychometric evaluation of the Taiwan Chinese version of the EORTC QLQ-PR25 for HRQoL assessment in prostate cancer patients. Health Qual Life Outcomes. 2012;10:96. doi:10.1186/1477-7525-10-96.

129. Vincent JI, MacDermid JC, King GJW, Grewal R. Rasch analysis of the patient-rated elbow evaluation questionnaire. Health Qual Life Outcomes. 2015;13:84. doi:10.1186/s12955-015-0275-8.

130. Abma IL, van der Wees PJ, Veer V, Westert GP, Rovers M. Measurement properties of patient-reported outcome measures (PROMs) in adults with obstructive sleep apnea (OSA): a systematic review. Sleep Med Rev. 2016;28:14-27.

131. Rofes L, Arreola V, Almirall J, Cabré M, Campins L, García-Peris P, et al. Diagnosis and management of oropharyngeal dysphagia and its nutritional and respiratory complications in the elderly. Gastroenterol Res Pract. 2011. 2011:818979. doi:10.1155/2011/818979.

132. Pace CC, McCullough GH. The association between oral microorgansims and aspiration pneumonia in the institutionalized elderly: review and recommendations. Dysphagia. 2010; 25(4):307-22.

133. Kayser-Jones J, Schell ES. Nursing staff can give dental exams. Provider. 1995;21(6):75-76.

134. Kayser-Jones J, Bird WF, Redford M, Schell ES, Einhorn SH. Strategies for conducting dental examinations among cognitively impaired nursing home residents. Spec Care Dent.1996;16(2):46-52.

135. Brushing up on mouth care. Annual Oral Health Assessment sheet for nursing staff: Oral Health Assessment Tool for long-term care. http://www.ahprc.dal.ca/projects/oral-care/pdfs/OHAT-AnnualOHAssessment.pdf. Accessed December 2014.

136. Halton Region's Health Department. Microsoft Power Point – Dysphagia – Oral health care tips. http://www.halton.ca/common/pages/UserFile.aspx?fileId=100073. Accessed December 2014.

137. Bowes D, van der Horst M-L, Kirkpatrick T. Oral Health Assessment Tool. Series 3. https://www.youtube.com/watch?v=MHdVralbWYQ. Presented January 2011. Voice Over Re-Recording May 2012. Accessed December 2014.

138. Halton Region's Health Department. Oral Hygiene Care Plan for long-term care. http://www.halton.ca. Accessed December 2014.

139. Brushing Up on Mouth Care. Oral Hygiene Care Plan for nursing staff. http://www.ahprc.dal.ca/projects/oral-care/pdfs/OHAT-AnnualOHAssessment.pdf. Accessed December 2014.

140. Kiyak HA, Reichmuth M. Barriers to and enablers of older adults' use of dental services. J Dent Educ.2005;69(9):975-86.

141. Pajukoski H, Meurman JH, Snellman-Gröhn S, Sulvaka R. Oral health in hospitalized and nonhospitalized community-dwelling elderly patients. Oral Surg Oral Med Oral Pathol Oral Radiol Endod.1999;88(4):473-43.

142. Griffin SO, Jones JA, Brunson D, Griffin PM, Bailey WD. Burden of oral disease among older adults and implications for public health priorities. Am J Public Health. 2012;102(3):411-18.

143. Folstein MF, Folstein SE, McHughPR. Mini-Mental State: a practical method for grading the cognitive state of patients for the clinician. J psychiat Res.1975;12:189-98. http://home.uchicago.edu/~/tmurray1/research/articles/printed%20and%20read/mini%20mental%20state_a%20practical%20method%20for%20grading%20the%20cognitive%20state%20of%20patients%20for%20the%20clinician.pdf. Accessed January 2015.

144. Portney LG, Watkins MP. Foundations of clinical research. Applications to practice. New Jersey: Prentice-Hall; 2000.

145. Nichols DP. Choosing an intraclass correlation coefficient. http://www.ats.ucla.edu/stat/spss/library/whichicc.htm. Accessed February 2015.

146. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33(1):159-74.

147. Herrera D, Alonso B, de Arriba L, Santa Cruz I, Serrano C, Sanz L. Acute periodontal lesions. Periodontol 2000. 2014;65(1):149-77.

148. Scully C, Felix DH. Oral medicine: update for the dental practitioner. Aphthous and other common ulcers. Br Dent J. 2005;199: 259-64. http://www.oralcancerfoundation.org/dental/pdf/oral_ulcers.pdf . Accessed February 2015.

149. Werneke MW, Hart DL, Deutscher D, Stratford PW. Clinician's ability to identify neck and low back interventions: an inter-rater chance-corrected agreement pilot study. J Man Manip Ther. 2011;19(3):172-81.

150. Chalmers JM, Robinson J, Nankivell N. The practical oral care video: evaluation of a dental awareness month initiative. Aust Dent J. 2005;50(2):75-80.

151. Olubunmi B, Olushola I. Effects of information dissemination using video of indigenous language on 11-12 years children's dental health. Ethiop J Health Sci. 2013;23(3):201-8. doi:10.4314/ejhs.v23i3.2.

152. Crary MA, Mann GD, Groher ME. Initial psychometric assessment of a functional oral intake scale for dysphagia in stroke patients. Arch Phys Med Rehabil. 2005; 86(8):1516-20.

153. Langmore S. Endoscopic evaluation of oral and pharyngeal phases of swallowing. GI Motility online. 2006; doi:10.1038/gimo28.

154. Mahoney FI, Barthel DW. Functional evaluation: the Barthel Index. Md State Med J. 1965;14:61-5.

155. de Haan R, Limburg M, Schuling J, Broeshart J, Jonkers L, van Zuylen P. Klinimetrische evaluatie van de Barthel-index, een maat voor beperkingen in het dagelijks functioneren. Ned Tijdschr Geneeskd. 1993; 137(18):917-21 . https://www.ntvg.nl/system/files/publications/1993109170001a.pdf. Accessed May 2017.

156. Mann G, Hankey GJ, Cameron D. Swallowing function after stroke: prognosis and prognostic factors at 6 months. Stroke. 1999;30:744-8.

157. Chua KS, Kong KH. Functional outcome in brainstem stroke patients after rehabilitation. Arch Phys Med Rehabil. 1996;77(2):194-7.

158. Hamdy S, Aziz Q, Rothwell JC, Crone R, Hughes D, Tallis RC, et al. Explaining oropharyngeal dysphagia after unilateral hemispheric stroke. Lancet. 1997;350(9079):686-92.

159. Parker C, Power M, Hamdy S, Bowen A, Tyrrell P, Thompson DG. Awareness of dysphagia by patients following stroke predicts swallowing performance. Dysphagia. 2004;19(1):28-35.

160. Daniels SK, Foundas AL. Lesion localization in acute stroke patients with risk of aspiration. J Neuroimaging. 1999;9(2):91-8.

161. Schelp AO, Cola PC, Gatto AR, Silva RG, Carvalho LR. [Incidence of oropharyngeal dysphagia associated with stroke in a regional hospital in São Paulo State – Brazil]. Arq Neuropsiquiatr. 2004;62(2B):503-6.

162. Smithard DG, O'Neill PA, Parks C, Morris J. Complications and outcome after acute stroke. Does dysphagia matter? Stroke. 1996;27(7):1200-4.

163. Odderson IR, Keaton JC, McKenna BS. Swallow management in patients on an acute stroke pathway: quality is cost effective. Arch Phys Med Rehabil. 1995;76(12):1130-3.

164. Ellenbecker CH, Samia L, Cushman MJ, Alster K (2008). Patient safety and quality in home health care. In Hughes RG (ed). Patient safety and quality. An Evidence-Based Handbook for Nurses. Rockville (MD): Agency for Healthcare Research and Quality (US), chapter 13.

165. Martin BJ, Corlew MM, Wood H, Olson D, Golopol LA, Wingo M, Kirmani N. The association of swallowing dysfunction and aspiration pneumonia. Dysphagia. 1994;9(1):1-6.

166. Croghan JE, Burke EM, Caplan S, Denman S. Pilot study of 12-month outcomes of nursing home patients with aspiration on videofluoroscopy. Dysphagia. 1994;9(3):141-6.

167. Gordon C, Hewer RL, Wade DT. Dysphagia in acute stroke. Br Med J (Clin Res Ed). 1987;295:411-14.

168. Sellars C, Bowie L, Bagg J, Sweeney MP, Miller H, Tilston J, et al. Risk factors for chest infection in acute stroke: a prospective cohort study. Stroke. 2007;38(8):2284-2291. doi:10.1161/STROKEAHA.106.478156.

169. Thorstensson H, Johansson B. Does oral health say anything about survival in later life? Findings in a Swedish cohort of 80+ years at baseline. Community Dent Oral Epidemiol. 2009; 37(4):325-32.

170. Appollonio I, Carabellese C, Frattola A, Trabucchi M. Influence of dental status on dietary intake and survival in community-dwelling elderly subjects. Age Ageing. 1997;26(6):445-56.

171. Andersen KK, Olsen TS, Dehlendorff C, Kammersgaard LP. Hemorrhagic and ischemic strokes compared: stroke severity, mortality, and risk factors. Stroke. 2009;40(6):2068-72. doi:10.1161/STROKEAHA.108.540112.

172. Jørgensen HS, Nakayama H, Raaschou HO, Olsen TS. Intracerebral hemorrhage versus infarction: stroke severity, risk factors, and prognosis. Ann Neurol. 1995;38(1):45-50.

173. Gomes J, Wachsman AM (2013). Types of strokes. In Corrigan ML, Escuro AA, Kirby DF (eds). Handbook of clinical nutrition and stroke, Nutrition and Health. NY: Springer Science+Business Media, pp 15-31. doi:10.1007/978-1-62703-380-0_2.

174. Smith CJ, Kishore AK, Vail A, Chamorro A, Garau J, Hopkins SJ, et al. Diagnosis of Stroke-Associated Pneumonia: Recommendations From the Pneumonia in Stroke Consensus Group. Stroke. 2015;46(8):2335-40. doi:10.1161/STROKEAHA.115.009617.

175. Kadojić D, Bijelić BR, Radanović R, Porobić M, Rimac J, Dikanović M. Aphasia in patients with ischemic stroke. Acta Clin Croat. 2012;51(2):221-5.

176. Nys GMS, van Zandvoort MJE, de Kort PLM, Jansen BPW, de Haan EHF, Kappelle LJ. Cognitive disorders in acute stroke: prevalence and clinical determinants. Cerebrovasc Dis. 2007;23(5-6):408-16.

177. Rasquin SMC, Verhey FRJ, van Oostenbrugge RJ, Lousberg R, Lodder J. Demographic and CT scan features related to cognitive impairment in the first year after stroke. J Neurol Neurosurg Psychiatry. 2004;75:1562-7. doi:10.1136/jnnp.2003.024190.

178. Graetz P, De Bleser R, Willmes K. De Akense Afasie Test. Nederlandse versie. Lisse: Swets & Zeitlinger; 1992.

179. O'Bryant SE, Humphreys JD, Smith GE, Ivnik RJ, Graff-Radford NR, Petersen RC, et al. Detecting dementia with the Mini-Mental State Examination (MMSE) in highly educated individuals. Arch Neurol. 2008;65(7):963-67. doi:10.1001/archneur.65.7.963.

180. Kukull WA, Larson EB, Teri L, Bowen J, McCormick W, Pfanschmidt ML. The Mini-Mental State Examination score and the clinical diagnosis of dementia. J Clin Epidemiol. 1994;47(9):1061-7.

181. Dutch vocabulary profile. http://www.zoekeenvoudigewoorden.nl/index.php. Accessed January 2016.

182. Council of Europe, Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR). Strasbourg: Language Policy Unit. http://www.coe.int/lang-CEFR. Accessed January 2016.

183. Jewell G, McCourt ME. Pseudoneglect: a review and meta-analysis of performance factors. Neuropsychologia. 2000;38:93-110.

184. Bour A, Rasquin S, Boreas A, Limburg M, Verhey F. How predictive is the MMSE for cognitive performance after stroke? J Neurol. 2010; 257(4):630-37. doi: 10.1007/s00415-009-5387-9.

185. Bogaardt HC, Speyer R, Baijens LW, Fokkens WJ. Cross-cultural adaptation and validation of the Dutch version of SWAL-QoL. Dysphagia. 2009;24(1):66-70.

186. Preston CC, Colman AM. Optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent preferences. Acta Psychol. 2000;104(1):1-15.

187. Somers A, Robays H, Audenaert K, Van Maele G, Bogaert M, Petrovic M. The use of hypnosedative drugs in a university hospital: has anything changed in 10 years? Eur J Clin Pharmacol. 2011;67(7):723-9.

188. Van Excel NJ, Scholte op Reimer WJ, Koopmanschap MA. Assessment of post-stroke quality of life in cost-effectiveness studies: the usefulness of the Barthel Index and the EuroQoL-5D. Qual Life Res. 2004;13(2):427-33.

189. Wheeler A, McKenna B, Madell D, Harrison J, Prebble K, Larsson E, et al. Self-reported health-related quality of life of mental health service users with serious mental illness in New Zealand. J Prim Health Care. 2015;7(2):117-23. https://www.diva-portal.org/smash/get/diva2:1033313/FULLTEXT01.pdf. Accessed January 2017.

190. Cano SJ, Hobart JC. The problem with health measurement. Patient Prefer Adherence. 2011;5:279-90. doi:10.2147/PPA.S14399.

191. Christensen KB, Kreiner S, Mesbah M. Rasch models in health. 1st ed. Hoboken: Wiley; 2013. doi:10.1002/9781118574454.

192. Hansen T, Lambert H, Faber J. Validation of the Danish version of the McGill ingestive skills assessment using classical test theory and the Rasch model. Dis Rehabil. 2012;34(10): 859-68.

193. Wright BD. Model selection: rating scale model (RSM) or partial credit model (PCM)? Rasch Meas Trans. 1998; 12(3):641-2. https://www.rasch.org/rmt/rmt1231.htm. Accessed August 2016.

194. Andrich D, Lyne A, Sheridan B, Luo G. RUMM2030. Perth: RUMM Laboratory Pty; 2010. http://www.rummlab.com.au/. Accessed August 2016.

195. Pallant JF, Tennant A. An introduction to the Rasch measurement model: an example using the hospital anxiety and depression scale (HADS). Br J Clin Psychol. 2007;46(pt1): 1-18.

196. Chen WH, Lenderking W, Jin Y, Wyrwich W, Gelhorn H, Revicki DA. Is Rasch model analysis applicable in small sample size pilot studies for assessing item characteristics? An example using PROMIS pain behavior item bank data. Qual Life Res. 2014;23:485-93.

197. Kutlay Ş, Küçükdeveci AA, Elhan AH, Öztuna D, Koç N, Tennant A. Validation of the World Health Organization disability assessment schedule II (WHODAS-II) in patients with osteoarthritis. Rheumatol Int. 2011;31(3):339-46.

198. Zucca A, Lambert SD, Boyes AW, Pallant JF. Rasch analysis of the mini-mental adjustment to cancer scale (Mini-MAC) among a heterogeneous sample of long-term cancer survivors: a cross-sectional study. Health Qual Life Outcomes. 2012;10(1):55.

199. Marais I, Andrich D. Formalizing dimension and response violations of local independence in the unidimensional Rasch model. J Appl Meas. 2008;9(3):200-15. file:///C:/Users/Notebook/Downloads/ARC-Report11DistinguishingMDandRD%20(2).pdf. Accessed February 2017.

200. Linacre JM. Sample size and item calibration stability. Rasch Meas Trans. 1994;7:328. http://www.rasch.org/rmt/rmt74m.htm. Accessed June 2016.

201. Chimi CJ, Russell DL. The Likert scale: a proposal for improvement using quasi-continuous variables. In: The proceedings of the information systems education conference. 2009. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.467.646&rep=rep1&type=pdf. Accessed August 2016.

202. González-Romá V, Espejo B. Testing the middle response categories «Not sure», «In between» and «?» in polytomous items. Psicothema. 2003; 15(2):278-84. http://www.psicothema.com/pdf/1058.pdf. Accessed August 2016.

203. Whitley BE, Kyte ME. Principles of research in behavioural science. 3rd ed. NY: Taylor & Francis; 2013.

204. Rummlab. Cronbach's alpha and the Person Separation Index (PSI). http://www.rummlab.com.au/rmrelidx2030. Accessed June 2016.

205. Holgado-Tello FP, Chacón-Moscoso S, Barbero-García I, Vila-Abad E. Polychoric versus Pearson correlations in exploratory and confirmatory factor analysis of ordinal variables. Qual Quant. 2010;44(1):153-66. doi:10.1007/s11135-008-9190-y.

206. Cordier R, Joosten A, Clavé P, Schindler A, Bülow M, Demier N, et al. Evaluating the Psychometric Properties of the Eating Assessment Tool (EAT-10) Using Rasch Analysis. Dysphagia. 2016. doi:10.1007/s00455-016-9754-2.

207. Hilari K, Boreham LD. Visual analogue scales in stroke: what can they tell us about health-related quality of life? BMJ Open. 2013;3(9):e003309. doi:10.1136/bmjopen-2013-003309.

208. Peeters JB, Daudey L, Heijdra YF, Molema J, Dekhuijzen PN, Vercoulen JH. Development of a battery of instruments for detailed measurement of health status in patients with COPD in routine care: the Nijmegen Clinical Screening Instrument. Qual Life Res. 2009;18(7):901-12.

209. Fayers PM, Machin D. Quality of life: the assessment, analysis and reporting of patient-reported outcomes. 3rd. ed. UK: Wiley; 2015.

210. Marshall S, Haywood K, Fitzpatrick R. Impact of patient-reported outcome measures on routine practice: a structured review. J Eval Clin Pract. 2006;12(5):559-68.

211. Garrett A, Schmidt L, Mackintosh A, Fitzpatrick R. Quality of life measurement: bibliographic study of patient assessed health outcome measures. BMJ. 2002; 324(7351): 1417.

212. Pickard AS, Knight SJ. Proxy-evaluation of health-related quality of life: a conceptual framework for understanding multiple proxy perspectives. Med Care. 2005;43(5):493-9.

213. Oozeer NB, Corsar K, Glore RJ, Penney S, Patterson J, Paleri V. The impact of enteral feeding route on patient-reported long term swallowing outcome after chemoradiation for head and neck cancer. Oral Oncol. 2011;47(10):980-3.

214. Needham DM, Dinglas VD, Bienvenu OJ, Colantuoni E, Wozniak AW, Rice TW, et al. One year outcomes in patients with acute lung injury randomised to initial trophic or full enteral feeding: prospective follow-up of EDEN randomised trial. BMJ. 2013;346:f1532.

215. Martin L, Blomberg J, Lagergren P. Patients' perspectives of living with a percutaneous endoscopic gastrostomy (PEG). BMC Gastroenterol. 2012;12(1):126. doi:10.1186/1471-230X-12-126.

216. Sura L, Madhavan A, Carnaby G, Crary MA. Dysphagia in the elderly: management and nutritional considerations. Clin Interv Aging. 2012;7:287-98.

217. Bakheit AMO. Management of neurogenic dysphagia. Postgrad Med J. 2001;77(913):694-9. doi:10.1136/bmj.77.913.694.

218. National Collaborating Centre for Acute Care (UK). Nutrition support for adults: oral nutrition support, enteral tube feeding and parenteral nutrition. London: National Collaborating Centre for Acute Care (UK); 2006; http://www.nice.org.uk/guidance/cg32. Accessed December 2015.

219. Janssens A, Rogers M, Thompson Coon J, Allen K, Green C, Jenkinson C, et al. A systematic review of generic multi-dimensional patient-reported outcome measures for children, Part II: Evaluation of psychometric performance of English-language versions in a general population. Value Health. 2015;18(2):334-45. doi:10.1016/j.jval.2015.01.004.

220. Teckle P, Peacock S, McTaggart-Cowan H, van der Hoek K, Chia S, Melosky B, et al. The ability of cancer-specific and generic preference-based instruments to discriminate across clinical and self-reported measures of cancer severities. Health Qual Life Outcomes. 2011;9(1):106. doi:10.1186/1477-7525-9-106.

221. Peters JB, Daudey L, Heijdra YF, Molema J, Dekhuijzen PN, Vercoulen JH. Development of a battery of instruments for detailed measurement of health status in patients with COPD in routine care: the Nijmegen Clinical Screening Instrument. Qual Life Res. 2009;18(7):901-12.

222. Sprangers MA, Aaronson NK. The role of health care providers and significant others in evaluating the quality of life of patients with chronic disease: a review. J Clin Epidemiol. 1992;45(7):743-60.

223. Anis MK, Abid S, Jafri W, Abbas Z, Shah HA, Hamid S, et al. Acceptability and outcomes of the percutaneous endoscopic gastrostomy (PEG) tube placement – patients' and caregivers' perspectives. BMC Gastroenterology. 2006;6:37. doi:10.1186/1471-230X-6-37.

224. Crary MA, Groher ME. Introduction to adult swallowing disorders. Philadelphia, PA: Butterworth Heinemann; 2003.

225. Ickenstein GW. Diagnosis and treatment of neurogenic dysphagia. 1st ed. Bremen: UNI-MED Verlag AG; 2011.

226. Ney DM, Weiss JM, Kind AJ, Robbins J. Senescent swallowing: impact, strategies, and interventions. Nutr Clin Pract. 2009;24(3):395-413.

227. Flidel-Rimon O, Friedman S, Lev E, Juster-Reicher A, Amitay M, Shinwell E. Early enteral feeding and nosocomial sepsis in very low birthweight infants. Arch Dis Child Fetal Neonatal Ed. 2004; 89(4):F289-F92. doi:10.1136/adc.2002.021923.

228. Mahant S, Friedman JN, Connolly B, Goia C, Macarthur C. Tube feeding and quality of life in children with severe neurological impairment. Arch Dis Child. 2009;94(9):668-73.

229. Schneider SM, Pouget I, Staccini P, Rampal P, Hebuterne X. Quality of life in long-term home enteral nutrition patients. Clin Nutr. 2000;19(1):23-8.

230. Miller RG, Jackson CE, Kasarskis EJ, England JD, Forshew D, Johnston W, et al. Practice parameter update: the care of the patient with amyotrophic lateral sclerosis: multidisciplinary care, symptom management, and cognitive/behavioral impairment (an evidence-based review): report of the Quality Standards Subcommittee of the American Academy of Neurology. Neurology. 2009;73(15):1227-33. doi:10.1212/WNL.0b013e3181bc01a4.

231. Roberge C, Tran M, Massoud C, Poirée B, Duval N, Damecour E, et al. Quality of life and home enteral feeding: a French prospective study in patients with head and neck or oesophageal cancer. Br J Cancer. 2000;82(2):263-9.

232. Finucane TE, Christmas C, Travis K. Tube feeding in patients with advanced dementia: a review of the evidence. JAMA. 1999;282(14):1365-70.

233. Roche V. Percutaneous endoscopic gastrostomy. Clinical care of PEG tubes in older adults. Geriatrics. 2003;58(11):22-6; 28-9.

234. Marik, PE. Aspiration pneumonitis and aspiration pneumonia. N Engl J Med. 2001;344(9):665-71.

235. Gomes GF, Pisani JC, Macedo ED, Campos AC. The nasogastric feeding tube as a risk factor for aspiration and aspiration pneumonia. Curr Opin Clin Nutr Metab Care. 2003;6(3):327-33.

236. Grant MD, Rudberg MA, Brody JA. Gastrostomy placement and mortality among hospitalized Medicare beneficiaries. JAMA. 1998;279(24):1973-6.

237. Elia M, Russell CA, Stratton RJ, Holden CE, Micklewright A, Barton A (Eds):et al. Trends in Home Artificial Nutrition Support in the UK during 1996-2000. A report by the British Artificial Nutrition Survey (BANS). London: British Association Parenteral and Enteral Nutrition, 2001.
238. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med. 2009;6(7):e1000097.
239. Terwee CB, Mokkink LB, Knol DL, Ostelo RW, Bouter LM, de Vet HC. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. Qual Life Res. 2012;21(4):651-7.
240. Scientific Advisory Committee of the Medical Outcomes Trust. Assessing health status and quality-of-life instruments: Attributes and review criteria. Qual Life Res. 2002;11(3):193-205.
241. Barten JA, Pisters MF, Huisman PA, Takken T, Veenhof C. Measurement properties of patient-specific instruments measuring physical function. J Clin Epidemiol. 2012;65(6):590-601.
242. Wiering B, de Boer D, Delnoij D. Patient involvement in the development of patient-reported outcome measures: a scoping review. Health expect. 2017;20(1):11-23. doi:10.1111/hex.12442.

CURRICULUM VITAE

Ingeborg Simpelaere was born in Ostend (Belgium) in 1979. In 2001, she obtained a Master's degree in Speech Language Pathology and Audiology Sciences at Ghent University with a master's thesis entitled 'Effectiveness of Lee Silverman Voice Treatment (LSVT) on nasality and voice in Parkinson's disease' (supervisors: Prof. dr. Kristiane Van Lierde, Prof. dr. Marc De Bodt). She also obtained a Master's degree in Teaching in Secondary Education in 2004 at Ghent University and an additional Bachelor's degree in Audiology at the Catholic University College of Bruges (KHBO) in 2003.

In 2001, she started working as a speech-language pathologist in AZ Delta Hospital Roeselare-Menen. She has been working for several years with patients suffering from neurological diseases and has a special interest in the diagnosis and treatment of neurogenic dysphagia and neurogenic communication disorders (i.e., aphasia, dysarthria and apraxia) as well.

Since 2004, she has been working in the VIVES University College of Bruges, where she is engaged as a researcher and lecturer. Her main research interests are swallowing disorders and neurogenic communication disorders and she also provides lectures about these topics.

In Octobre 2013, she officially started her PhD at the University of Antwerp, supervised by Prof. dr. Marc De Bodt and Prof. dr. Gwen Van Nuffelen. Her core research is neurogenic dysphagia, in particular post-stroke oropharyngeal dysphagia and its associated functional outcomes, and the adapation and psychometric validation of patient-reported outcome measures measuring health-related quality of life in dysphagic patients.

SCIENTIFIC CURRICULUM

1. National first author publications

Simpelaere I, Hallaert C, Mahieu S, Verbrugghe K, Deklerck J. Dysfagie: een multidisciplinaire benadering. Een bevraging van verpleegkundigen binnen de acute intramurale zorg. Logopedie. 2014; 27(4):84-98.

Simpelaere I, Van Lierde KM, De Bodt M, Van Cauwenberge P. Logopedische behandeling bij de ziekte van Parkinson: een literatuuroverzicht. Tijdschrift voor Logopedie en Audiologie. 2002;32(3):86-96.

2. National second author publication

White A, Simpelaere I, Vanmaele C, Vanhooren G, Deklerck J. Het ontwikkelen van een instrument voor het evalueren van de effectiviteit van rTMS-therapie bij personen met afasie. Logopedie. 2013;26(5):22-29.

3. Oral presentations

Levenskwaliteit bij dysfagiepatiënten

15de Symposium Logopedie & Audiologie UGent, Universiteit Gent, Gent (België), 20 oktober 2014

Slikmanoeuvres en houdingstechnieken bij patiënten met orofaryngeale dysfagie

Tweedaagse navorming Dysfagie: Onderzoek en behandeling. Een multidisciplinaire benadering, VIVES Hogeschool, Brugge (België), 20-21 november 2014

Levenskwaliteit bij dysfagiepatiënten

Tweedaagse navorming Dysfagie: Onderzoek en behandeling. Een multidisciplinaire benadering, VIVES Hogeschool, Brugge (België), 20-21 november 2014

Comparing patient and proxy agreement concerning impact of dysphagia on patients' quality of life (QoL) as measured by the Dutch version of the Swallowing Quality-of-Life (DSWAL-QoL) questionnaire

5th European Society for Swallowing Disorders (ESSD) Congress, Barcelona (Spain), 1-3 Octobre 2015

Application of the Rasch measurement model on the Dutch version of the Swallowing Quality-of-Life Questionnaire (DSWAL-QoL)

23rd Annual Conference of International Society for Quality of Life Research (ISOQOL), Copenhagen (Denmark), 19-22 Octobre 2016

Feasibility and psychometric properties of the adjusted DSWAL-QoL Questionnaire (aDSWAL-QoL) for dysphagic patients with additional language and/or cognitive impairment

3rd Annual Congress of Belgian Society for Swallowing Disorders (BSSD), Brussels (Belgium), 6 May 2017

Langetermijnoutcome van orofaryngeale dysfagie bij CVA-patiënten

18de Symposium Logopedie & Audiologie UGent, Universiteit Gent, Gent, 20 oktober 2017

4. First author poster presentations

Comparing patient and proxy agreement concerning impact of dysphagia on patients' quality of life (QoL) as measured by the Dutch version of the Swallowing Quality-of-Life (DSWAL-QoL) questionnaire

23rd Annual Meeting of the Dysphagia Research Society (DRS), Chicago (United States), 11-14 March 2015

Communicatiestimulatie onder de vorm van groepstherapie bij dementerende bejaarden – een verkennende studie

36ste Congres Vlaamse Vereniging voor Logopedisten (VVL), Gent (België), 20 maart 2015

Application of the Rasch measurement model on the Dutch version of the Swallowing Quality-of-Life Questionnaire (DSWAL-QoL)

5th ESSD Congress, Barcelona (Spain), 1-3 Octobre 2015

Comparing patient and proxy agreement concerning impact of dysphagia on patients' quality of life (QoL) as measured by the Dutch version of the Swallowing Quality-of-Life (DSWAL-QoL) questionnaire

22nd Annual Conference ISOQOL, Vancouver (Canada), 21-24 Octobre 2015

Application of the Rasch measurement model on the Dutch version of the Swallowing Quality-of-Life Questionnaire (DSWAL-QoL)

37ste VVL-congres, Gent (België), 20 mei 2016

Comparing patient and proxy agreement concerning impact of dysphagia on patients' quality of life (QoL) as measured by the Dutch version of the Swallowing Quality-of-Life (DSWAL-QoL) questionnaire

37ste VVL-congres, Gent (België), 20 mei 2016

5. Other poster presentations

Desmet B, Maertens M, Simpelaere I, et al. Assistive technology and communication problems. International conference of the International Association for Communication in Healthcare, St Andrews (United Kingdom), 5 September 2012.

Debaere K, Simpelaere I, et al. Empowerment in home care. International conference of the International Association for Communication in Healthcare, St Andrews (United Kingdom), 6 September 2012.

Cloet G, Simpelaere I. Ingedikte drank na een beroerte: dehydratie?
37ste VVL-congres, Gent (België), 20 mei 2016

6. Attended conferences

International conference of the International Association for Communication in Healthcare, St Andrews (Scotland), 4-7 Septembre 2012

2nd ESSD Congress, Barcelona (Spain), 25-27 Octobre 2012

15de Symposium Logopedie & Audiologie UGent, Universiteit Gent, Gent, 20 oktober 2014

4th ESSD Congress, Brussels (Belgium), 23-25 Octobre 2014

21st Annual Conference ISOQOL, Berlin (Germany), 15-18 Octobre 2014

Tweedaagse navorming Dysfagie: Onderzoek en behandeling. Een multidisciplinaire benadering, VIVES Hogeschool, Brugge (België), 20-21 november 2014

23rd Annual Meeting of the Dysphagia Research Society (DRS), Chicago (United States), 11-14 March 2015

36ste Congres Vlaamse Vereniging voor Logopedisten (VVL), Gent (België), 20 maart 2015

5th ESSD Congress, Barcelona (Spain), 1-3 Octobre 2015

22nd Annual Conference ISOQOL, Vancouver (Canada), 21-24 Octobre 2015

23rd Annual Conference of International Society for Quality of Life Research (ISOQOL), Copenhagen (Denmark), 19-22 Octobre 2016

3rd Annual Congress BSSD, Brussels (Belgium), 6 May 2017

37ste VVL-congres, Gent (België), 20 mei 2016

7th ESSD Congress, Barcelona (Spain), 20-22 Septembre 2017

18de Symposium Logopedie & Audiologie UGent, Universiteit Gent, Gent (België), 20 oktober 2017

7. Courses

Systematic Review Part I & Part II

CEBAM, Universiteit Leuven, Leuven, 2013

Taalopleiding Engels

Linguapolis, Universiteit Antwerpen, Antwerpen, 2014

Introduction to R

Flames statistics, Universiteit Gent, Gent, 2015

8. Educational tasks

Les over 'Levenskwaliteit bij dysfagiepatiënten'

Postgraduaat dysfagie, Arteveldehogeschool, Gent, 2016

Lecture on 'Evidence-based Dysphagia Assessment'

Université catholique de Louvain-la-Neuve, Louvain-la-Neuve, 2017

Supervision Bachelor thesis:

2012-2013: Sien Calcoen – Caroline Vansteelandt

Review van bestaande assessment-intrumenten en de geassocieerde logopedische problematiek bij het slik- en voedingsproces. Onderdeel van de ontwikkeling van een multidisciplinair onderzoeksinstrument voor gerichte behandeling van kinderen en adolescenten met dyskinetische CP.

2012-2013: Jolien Sentobin – Jolien Vangheluwe

Aanpassing van de DSWAL-QoL vragenlijst voor dysfagiepatiënten met afasie. Initiële modificatie en evaluatie.

2012-2013: Griet Vercruysse

Review en pilot-study omtrent de meerwaarde van het gebruik van de zuurstofsaturatiebepaling bij CVA-patiënten met dysfagie.

2012-2013: Ine Goethals, Charlotte Vandenbeeck, Steffi Maes

Presbyfagie of dysfagie? Literatuurstudie naar de functionele veranderingen in het eet- en slikproces van gezonde ouderen. Pilootstudie naar de invloed van de dentale status op de eet- en voedingstoestand van ouderen met CVA met en zonder dysfagie.

2013-2014: Jasmien Vanlerberghe – Elisa Van Ryssel – Sofie Van Houtte

Toegankelijkheid van de aangepaste Nederlandse versie van de Swallowing Quality-of-Life Questionnaire (DSWAL-QoL) voor dysfagiepatiënten met communicatieve en/of cognitieve problemen.

2013-2014: Eveline Mestdagh

Communicatiestimulatie onder de vorm van groepstherapie bij dementerende bejaarden – een verkennende studie.

2014-2015: Anneleen Meire – Sieg De Baerdemaeker

A randomized controlled trial naar het effect van neuromusculaire elektrostimulatie bij CVA-patiënten met orofaryngeale dysfagie: vergelijking van twee stimulatievormen 'Burst Mode Alternating Current' en 'Pulsed Current'. Vooronderzoek naar de optimale neurofysiologische parameters, eigen aan spierstimulatie, bij personen zonder neuromusculaire aandoeningen.

2014-2015: Annelien Gadeyne

Inventarisatie van sondevoeding-gerelateerde items bij patiënten met orofaryngeale dysfagie op basis van proxyrapportering.

2014-2015: Jolien Lammertijn

Inventarisatie van sondevoeding-gerelateerde items bij patiënten met orofaryngeale dysfagie op basis van patiëntrapportering en proxyrapportering.

2014-2015: Céline Bernard

Onderzoek naar de intra- en interbeoordelaarsbetrouwbaarheid van het logopedisch onderdeel van het multidisciplinair onderzoeksinstrument. Onderdeel van de ontwikkeling van een multidisciplinair onderzoeksinstrument voor gerichte behandeling van kinderen en adolescenten met dyskinetische cerebrale parese.

2014-2015: Sarah Van Belle

Vergelijking van de quality of life, aan de hand van de DSWAL-QoL, bij dysfagiepatiënten met en zonder sondevoeding.

2015-2016: Claire Streuve

Voorspellende factoren van dysfagie na CVA en outcomestudie

2016-2017: Tessa Debeuckelaere

Wat is de perceptie van de proxy (professional), vanuit het perspectief van de patiënt, op de quality of life bij patiënten met sondevoeding ten gevolg van een orofaryngeale dysfagie?

2016-2017: Febe Maene

Inventarisatie van sondevoeding-gerelateerde items bij patiënten met orofaryngeale dsyfagie op basis van patiëntenrapportering en proxyrapportering: vervolgproject.

2016-2017: Evi Verbeke

Outcome na CVA op korte en lange termijn: speelt de ernst van de dysfagie een rol?

2016-2017: Yenka Delaeter

HRQoL bij patiënten met orofaryngeale dysfagie en enterale voeding: de initiële ontwikkeling van een vragenlijst (bevraging bij patiënten)

2016-2017: Jieke Vanthuyne

HRQoL bij patiënten met orofaryngeale dysfagie en enterale voeding: de initiële ontwikkeling van een vragenlijst (bevraging bij verpleegkundigen)

Supervision Postgraduate thesis:

2016-2017: Elien De Cock

Multidisciplinaire samenwerking bij patiënten met sondevoeding in de thuissituatie: navraag bij zorgverleners

2016-2017: Sophie Desimpele

Bevraging van verpleeg- en zorgkundigen in woonzorgcentra omtrent detectie van dysfagie. Onderzoek naar kennis van detectie van dysfagie en nood aan vorming bij verpleeg- en zorgkundigen in woonzorgcentra in West- en Oost-Vlaanderen.

2016-2017: An Vandergunst

Aanpak van dysfagie in woonzorgcentra. Onderzoek naar kennis van aanpak van dysfagie en nood aan vorming bij verpleeg- en zorgkundigen in woonzorgcentra in West- en Oost-Vlaanderen

Supervision thesis 'Cursus Neurovasculaire Zorg':

2015-2016: Greet Cloet

Ingedikte drank na een beroerte: dehydratie?

2015-2016: Daisy Demasure

Initieel ontwerp van een tool als ondersteuning bij het informeren van de afasiepatiënt in de acute fase na CVA

9. Other

Member of the Ethics Committee of AZ Delta Hospital Roeselare-Menen-Torhout

Since 2015

Peer review A1 international publications:

Four articles (The titles of these articles cannot be presented because of privacy reasons):

2 articles for Dysphagia

1 article for Health Expectations

1 article for Health and Quality of Life Outcomes

Project Coordinator 'Projectmatig Wetenschappelijk Onderzoek' VIVES University College:

2012-2015: Het ontwikkelen van een multidisciplinair onderzoeksinstrument voor de inventarisatie van slik- en voedingsproblemen bij kinderen met cerebrale parese.

2016-2018: Ontwikkelen van een ziektespecifieke patient-reported outcome (PRO) instrument voor patiënten die via enterale voeding gevoed worden als gevolg van orofaryngeale dysfagie op basis van patiënt- en proxygebaseerde rapportering.

Conference organisator:

Tweedaagse navorming Dysfagie: Onderzoek en behandeling. Een multidisciplinaire benadering. VIVES Hogeschool Brugge, 20-21 november 2014.

Project participation:

2011-2012: Participatie in het Europees Fonds voor Regionale Ontwikkeling-project 655, afdeling Zorgcompetenties en –faciliteiten. Deelname aan de oprichting van het Centrum Empowerment in Ouderenzorg. Deelname aan het kwalitatief onderzoek met betrekking tot de empowerende hulpverlening in de thuiszorg en deelname aan de contextanalyse ouderen en tablets met betrekking tot empowerende technologie in de thuiszorg.

DANKWOORD

'Thank you words' seems small to what you have done for me…

Doctoreren… een woord, aanvankelijk nog ver van mijn bed…

Een zonnestraal tijdens een regenachtige namiddag; een van die namiddagen waarop er druk vergaderd werd met de collega's van de opleiding Logopedie & Audiologie. 'Doctoreren, zou dat niets voor jou zijn?' Een vraag die toen gesteld werd, in mijn hoofd bleef wentelen, die me tot diep nadenken aanzette en uiteindelijk drijfveren vond… Professor Dr. Marc De Bodt, een autoriteit op het logopedisch vakgebied, de promotor van mijn masterthesis, was niet alleen een bezield inspirator tijdens zijn lessen, maar evenzeer tijdens het eerste telefonisch contact om te bespreken wat het doctoraat zou inhouden. En het onderwerp? Dat wist ik vrijwel meteen: 'Dysfagie'. Zeer gelukkig omdat ik over dit onderwerp onderzoek mocht doen, kwam ik al vrij snel tot het idee om onderzoek te doen naar dysfagie in het kader van cerebrovasculaire accidenten. In de dagdagelijkse klinische praktijk kom ik immers frequent in contact met een dergelijke populatie. De toenemende aandacht voor kwaliteitseisen in de gezondheidszorg evenals de stimulerende samenwerking met de artsen zorgde ervoor dat ik me meer vragen ging stellen bij het gebruik van bepaalde meetinstrumenten voor een dergelijke populatie. Waren onze meetinstrumenten wel voldoende valide en bruikbaar in de klinische praktijk? Hoe zou de prognose zijn voor een CVA-patiënt met orofaryngeale dysfagie na een jaar? Zou mijn meetinstrument bijdragen aan het voorspellen van die prognose?

Een doctoraat doe je niet alleen. Omdat ik op zoveel steun van jullie kon rekenen, wil ik ieder van jullie dan ook zeer hartelijk danken:

In de eerste plaats wil ik Professor Dr. Marc De Bodt bedanken. Professor, het Nederlandse lexicon lijkt aanzienlijk gelimiteerd om uit te drukken wat u voor mij hebt betekend. Uw aanzienlijk optimisme en enthousiasme, uw zeer deskundige feedback, uw onvoorwaardelijk vertrouwen in mij! Het doctoraat ging

immers gepaard met het bereiken van hoge bergketens, maar ook met diepe dalen. Gelukkig hield u sterke klimtouwen vast en kon ik toch weer die berg op. Ik ben zeer blij dat ik die reis heb ondernomen, maar de volgende keer zou ik toch liever de diepe dalen vermijden en alleen de hoge bergtoppen verkennen. Ondanks uw zeer constructieve feedback, liet u me steeds vrij in het maken van keuzes. U liet me steeds zelf de knoop doorhakken, niettegenstaande ik me dikwijls heel onzeker heb gevoeld. Maar het was net die manier die ervoor zorgde dat ik steeds meer zelfvertrouwen kreeg in mijn eigen manier van denken en handelen. Ik kan u niet genoeg bedanken!

Professor Dr. Van Nuffelen, Gwen, u was evenzeer een grote steun en toeverlaat voor mij. Ik heb heel veel geleerd uit uw constructieve feedback en uit uw kritische vraagstelling naar bepaalde keuzes die ik maakte. Hartelijk dank voor u vele waardevolle ideeën en uw uitgebreide expertise op het domein van dysfagie waarop ik beroep mocht doen. U motiveerde me om deel te nemen aan internationale en nationale congressen. Gaandeweg voelde ik het zelfvertrouwen om te presenteren stijgen en kreeg ik gelukkig controle over mijn aanvankelijk veel te hoog spreektempo! Ik heb enorm genoten van de ontspannende momenten na een drukke congresdag. Wat waren onze ogen groot toen we die speciale en reusachtige pizza in Chicago aanvielen om dan na een paar stukken toch zo'n opgeblazen gevoel te ervaren… Of die gezellige wandeling doorheen het Millenium Park, waar we in maart toch nog omvangrijke ijsblokken konden aanschouwen….Dank je wel voor de morele steun tijdens die momenten waarop ik het even niet meer zag… Ik hoop dat we in de toekomst nog heel veel zullen samenwerken!

Professor Dr. Vanderveken en Doctor Boey, ik wil u hartelijk danken voor het kritisch nalezen van het proefschrift en de tussentijdse voortgangsrapporten! Professor Dr. Van Den Wyngaert, ik wil u bedanken voor het kritisch nalezen van de tussentijdse voortgangsrapporten en voor uw feedback tijdens het tussentijds evaluatiemoment. Professor Dr. Desuter en Professor Dr. Goeleven wens ik ook heel graag te bedanken om deel uit te maken van de jury!

Een heel speciaal woord van dank gaat ook uit naar Dokter Jan Vanderwegen. Jan, er zijn mensen met een kritisch en scherp oog, maar u hebt er zo twee! Echt waar! Ik ken er niet veel die zo kritisch en

tegelijk zo recht op de dingen zitten. Ik vond het een hele uitdaging om uw opmerkingen te verwerken, al moet ik toegeven dat die dikwijls lastig waren. Maar u had wel steeds een punt! Eens de publicaties door uw handen waren gegaan, voelde ik me dan ook steeds zekerder om die in te sturen. Maar ook voor uw persoonlijkheid moet ik u danken. Ik heb vaak achter mijn bureau, maar ook tijdens de congresdagen zitten lachen met uw guitige opmerkingen en grappige kanttekeningen. Dank u wel ook om mij zo aan te moedigen om deel te nemen aan nationale en internationale congressen. Samen met Gwen heb ik zo genoten van het congres in Chicago. Maar ook tijdens mijn oral presentation in Barcelona heb ik uw morele steun zo geapprecieerd. Stressgevoelens vooraf, maar oh zo'n fijn en opluchtend gevoel na afloop! Ik hoop dat we in de toekomst nog vaak samen aan die congressen mogen deelnemen!

Doctor Tina Hansen, I am deeply honoured that I got the chance to collaborate with you! You have learned me so much about psychometrics, about thinking straightforward, about representing the issues, … 'Be concise' and 'always think about "*what is need to know and what is nice to know"'* . These statements are adhered into my head and I really want to thank you for that. Honestly, I was not very fond of George Rasch in the beginning. At least, I did not like his mathematical theories very much. Such a complexity! Why did he make it so difficult? Or he was just Danish, and therefore difficult to understand??? Joking aside, by studying the literature, getting your explanation about some issues, by integrating myself in the world of the Rasch methodology, I am now completely convinced about the benefits of testing rating scales against this measurement model! I also would like to thank you for the person you are. You read and heard a lot 'Tina, one last question' or 'Tina, just one question left'. Thank you so much for so quickly responding my succeeding questions. Thank you so much for your immense patience in dealing with my undefatigable insistence to becoming familiar with the Rasch methodology and the doubts and critical reflections that were raising about this topic.

Doctor Ella Roelant, Ella, ik wil u ook heel hartelijk danken voor de inspanningen die u voor mij gedaan hebt. Dank u wel om zo snel mogelijk te antwoorden op mijn vragen. Ik heb op korte tijd heel veel geleerd over logistische regressie, over Kaplan-Meier curves, Cox-regression en wil u daarvoor dan ook heel graag bedanken.

Dank je wel Doctor Kristien Wouters, Kristien, voor de statistische begeleiding bij de aanvang van het doctoraat. U hebt me flink op weg geholpen bij de eerste twee publicaties. Hartelijk dank!

Doctor Trudy Bekkering, Doctor Bart Geurden, dank u wel voor uw begeleiding tijdens het tot stand komen van mijn systematic review protocol. Uw feedback was zeer waardevol; ik heb enorm veel geput uit uw feedback. Door het afsluiten van dit doctoraat kon ik pas later starten met het uitvoeren van de systematic review, maar ik kan u verzekeren dat dit in orde zal komen. Ik kijk er naar uit om samen een mooie publicatie af te leveren!

Een zeer gemeend woord van dank gaat ook uit naar de NKO-artsen van AZ Delta en het UZA. In het bijzonder wil ik graag Dokter Van Pelt en Dokter D. Van Rompaey bedanken voor hun interesse en deelname aan mijn onderzoek.

Ook het team van de neurologen van AZ Delta en het UZA wil ik heel graag danken. In het bijzonder wil ik Dokter Hasenbroekx, Dokter Merkx, Dokter Clement, Dokter Buyle, Dokter Bourgeois en Professor Dr. Cras bedanken om mijn onderzoek in het verleden te mogen voorstellen, voor het doorverwijzen van patiënten, voor het scoren van de NIHSS en GCS-schalen. Ook een woord van dank gaat uit naar Dokter. Vermeersch.

Hartelijk dank ook aan de overige artsen en directie van het AZ Delta en het UZA, aan de verpleegkundigen en paramedici van AZ Delta en het UZA, aan de verpleegkundigen en paramedici werkzaam in revalidatiecentra en woon- en zorgcentra in West-Vlaanderen en de provincie Antwerpen.

Ik zou ook graag een heel speciaal woord van danken richten tot mijn lieve collega's van AZ Delta en het UZA. Dank je wel om deel te nemen aan mijn onderzoek! Dit heeft enorm veel voor mij betekend! Dank je Greet Leenknecht, Delphine, Daisy, Greet Cloet, Tine, Ann-Sophie en Judit van AZ Delta. Dank je Cindy, Els en Leen van het UZA!

Dank je wel, Emke en Griet, voor jullie inspanningen en jullie laaiend optimisme tijdens de deelname aan het onderzoek! Dank jullie wel om tijdens de barre wintermaanden – en zelfs op tweede kerstdag!!! – klaar te staan aan de rusthuizen, serviceflats of aan het ziekenhuis om de betrouwbaarheid van de OHAT te kunnen evalueren. Anne, dank je wel om feedback te geven op het systematic review protocol en uiteraard hartelijk dank om in de toekomst de review te helpen uitvoeren.

Maar ook de collega's van de Hogeschool VIVES, in het bijzonder Hilde, Jo, Lut, Rudy, Ingrid, Ellen, Celine, Saartje, Ingeborg en Ann wens ik heel graag te bedanken! Dank je wel voor jullie inspirerende ideeën om aan een doctoraat te beginnen, dank je wel voor jullie begrip en aanmoediging om dit door te zetten.

Graag dank ik ook de studiegebieddirecteurs van de Hogeschool VIVES van de laatste jaren, namelijk Francis Decoster, Isabel Vanslembrouck en Jan Vandekerckhove, voor hun interesse en begrip voor dit doctoraat en voor de geboden kansen.

Ik wil zeker heel graag alle patiënten, wettelijke vertegenwoordigers en familieleden bedanken voor hun deelname aan de studies!

Dank je wel aan alle studenten Jolien (2x), Elisa, Jasmien, Sofie, Claire en Evi voor de inzet die jullie opbrachten voor jullie eindwerk over deelaspecten van dit onderzoek. Hopelijk hebben jullie veel geleerd uit jullie eindwerk en kijken jullie er met een tevreden blik op terug. Ik hoop dat jullie één voor één competente logopedisten mogen zijn.

Ik zou graag ook alle vrienden en vriendinnen bedanken (ik ga geen namen noemen om te vermijden dat ik er een paar schandelijk zou vergeten!) die er geweest zijn voor mij tijdens die 4 jaar! Dank je wel voor de gezellige avonden en plezante feestjes om er even tussenuit te zijn. Dank je wel voor jullie jarenlange vriendschap!

Ook zou ik graag mijn familie bedanken voor hun aanmoediging en belangstelling voor dit doctoraat. Een welgemeend woord van dank gaat vooral uit naar mijn schoonouders! Dank je wel om regelmatig op de kindjes te passen, zodat ik wat meer tijd had om in alle rust (!) te kunnen doorwerken.

Bram, dank je wel om de laatste aanpassingen aan mijn lay-out te doen op het moment dat ik in uiterste paniek zat! Gelukkig heb ik een schoonbroer die alles van tekstverwerking afweet.

Hannelore, zusje, dank je wel dat ik op je kon rekenen toen ik mocht gaan spreken op een internationaal congres! Aan de telefoon, aan je keukentafel, … moest je kritisch en telkens opnieuw luisteren naar mijn uitspraak. Hebben we toch ook wel een aantal keer gelachen met bepaalde accenten, niet? Of met woorden, die er toch wel de eerste maal bij mij zeer grappig uitkwamen. Dank je wel om mij te verbeteren, dank je wel om suggesties te doen op vlak van de Engelse taal. Maar dank je vooral voor de gezellige 'zus-momenten'!

Eveline, 'kleine zus', dank je dat ik bij jou terecht kon voor de onvoorwaardelijke morele steun! Je luisterde naar me als het plots eens te veel werd en motiveerde me om 'gewoon door te doen'. Dank je wel om me aanmoedigende berichtjes toe te sturen op het moment dat ik in het buitenland zat. Dank je wel omdat je mijn zus bent!

Mama en papa, ik wil jullie zeer graag bedanken om mij al die jaren zo aan te moedigen. Bedankt voor jullie liefde, jullie begrip, jullie grote hulp tijdens het realiseren van dit doctoraat. Ik hoop dat ik jullie vandaag trots kan maken. Wat als ik jou niet had gehad, mama! Ik zie het scenario zo voor me: lege kledingkasten, overvolle wasmanden, de kinderen in ongestreken kledij naar school, … Best dat jij er was om mijn huishouden zoveel mogelijk op te vangen!

Kristof, we kenden elkaar al vooraleer ik aan mijn masterstudies begon en door de vele jaren heen heb ik van jou heel veel steun mogen ontvangen. Je steunde me in mijn ideeën en moedigde me aan om mijn

dromen te realiseren. Nu zal je waarschijnlijk wel ongelooflijk blij zijn dat ik dit doctoraat vandaag kan afronden. Als ik denk aan de vele energie die naar het doctoraat gegaan is, de momenten van paniek en stress waarop je mij weer tot rede kon brengen, … Maar vooral zie ik je reikhalzend uitkijken naar twee volledig nette tafels binnen onze bureau. Wellicht moet je heel frequent het gevoel gehad hebben mij te moeten zoeken, bedolven onder een reusachtige papierberg, waar alleen ikzelf mijn weg in vond. Dank je wel voor het vele geduld en de ruimte – en dit zowel in de letterlijke als figuurlijke betekenis van het woord – die je me gaf om dit project tot een goed einde te brengen.

Een dikke zoen voor mijn twee schatten, Margaux en Marie-Julie.