NORTHWESTERN UNIVERSITY

Assessing the Physiology of Swallowing Impairment: Measuring the Measurement Method and Characterizing Diagnostic Impairment Profiles

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Communication Sciences and Disorders

By

Alexander E. Clain

EVANSTON, ILLINOIS

SEPTEMBER 2023

Abstract

Swallowing impairment or dysphagia has many possible causes with severe sequelae. One major mediator of the relationship between cause and sequelae is the physiologic impairment of the swallowing mechanism. The characterization of physiologic swallowing impairment is therefore of great importance in that treatment can target physiology to mitigate sequelae. The measurement of swallowing physiology is primarily conducted through Modified Barium Swallow Studies (MBSS), where Videofluoroscopy (VFS) of a patient swallowing an x-ray opaque bolus is recorded and interpreted by a Speech Language Pathologist (SLP). On the one hand, the goals of this assessment are objective in that they are aimed at understanding impairments to the physiologic biomechanics of the swallowing mechanism. On the other hand, the methods of this assessment are subjective in that clinicians must choose what boluses to give, what physiologic aspects of the swallow to assess, and how to score impairment for those physiologic aspects. The Modified Barium Swallow Impairment Profile (MBSImP™©) is a measurement method that standardized the subjective elements of the assessment, and subsequently enjoyed widespread clinical uptake. This widespread uptake of MBSImP resulted in the accumulation of over 50,000 patient records in a Swallowing Data Registry (SDR), a dataset that forms the basis for the analyses of this dissertation. Chapter 1 assesses the degree to which MBSImP's standardization of the subjective side of MBSS has resulted in a valid and reliable measurement tool as it is used in real-world and generalizable samples. Chapter 2 leverages MBSImP's standardized approach to conduct a high-level comparison of the physiologic impairment profiles of five diagnoses commonly associated with dysphagia, i.e. Stroke, Head and Neck Cancer, Dementia, Parkinson's Disease, and Chronic Obstructive Pulmonary Disorder.

List of Abbreviations

AHE	Anterior Hyoid Excursion
BH	Tongue Control During Bolus Hold
BP/M	Bolus Preparation/Mastication
BT	Bolus Transport/Lingual Motion
EM	Epiglottic Movement
IPS	Initiation Of Pharyngeal Swallow
LC	Lip Closure
LE	Laryngeal Elevation
LVC	Laryngeal Vestibular Closure
OR	Oral Residue
PC	Pharyngeal Contraction
PES	Pharyngoesophageal Segment Opening
PR	Pharyngeal Residue
PSW	Pharyngeal Stripping Wave
SPE	Soft Palate Elevation
TBR	Tongue Base Retraction
COPD	Chronic Obstructive Pulmonary Disorder
EC	Esophageal Clearance (Upright Position)
HNC	Head And Neck Cancer
ICC	Intraclass Correlation Coefficient
MBSImP	Modified Barium Swallow Impariment Profile
mL	Milliliter
OI	Overall Impression
OT	Oral Total
PAS	Penetration-Aspiration Scale
PD	Parkinson's Disease

PDwDem	Parkinson's Disease With Dementia
Pen/Asp	Penetration/Aspiration
PT	Pharyngeal Total
SDR	Swallowing Data Registry
SLP	Speech Language Pathologist
VFS	Videofluoroscopy

Table of Contents

List of Tables

List of Figures

Dissertation Introduction

Swallowing impairment or dysphagia is caused by an astonishing variety of conditions, including Stroke, Head and Neck Cancer, Pulmonary disorders, Gastric disorders, spinal injuries, etc., and has severe sequelae including increased risk of pneumonia, higher mortality rates, reduced quality of life, and increased cost of care (Altman et al., 2010; Bonilha et al., 2014; Eglseer et al., 2018; Guyomard et al., 2009; Kawashima et al., 2004; Kidambi et al., 2012; Lo et al., 2019; D. A. Patel et al., 2018). Driven by the ubiquity and severe consequences of dysphagia, clinicians and researchers have together developed a science of dysphagia to which this dissertation contributes. This field of dysphagia science is relatively young as can be seen in Figure 1 which shows that before 1990 there were less than 100 papers per year being published on dysphagia or swallowing impairment, and that has grown to nearly 3000 papers per year since 2020 (query from Web of Science: dysphagia OR swallowing impairment). This dissertation is therefore situated in a rapidly evolving and growing science that is attempting to understand how the causes of dysphagia lead to its consequences.

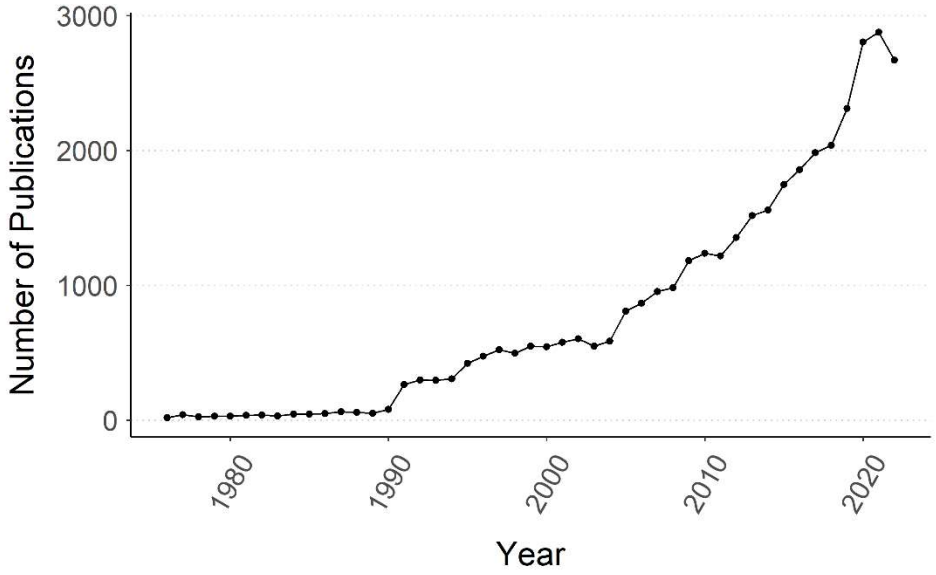


Figure 1 The number of publications per year pertaining to dysphagia or swallowing impairments.

One major potential mediator of this cause-to-sequelae connection is physiologic impairment. The physiology of swallowing and dysphagia is complex, requiring the coordination of multiple organ systems (i.e. neuromuscular, digestive, pulmonary) to generate a cascade of physiologic actions that ideally

moves a bolus from the oral cavity to the esophagus safely (without aspiration) and efficiently (without residue) (Logemann, 1988; Miller, 1986; Walton & Silva, 2018). It is the measurement and characterization of this physiology that is the focus of this dissertation.

The most commonly used method of assessing physiologic impairments of dysphagia is through Modified Barium Swallows Studies (MBSS) using a Videofluoroscopy (VFS) approach. In these assessments, a patient is given a series of barium-infused x-ray opaque boluses to swallow and videofluoroscopic imaging of their swallowing is acquired and recorded. It is then up to the clinician to interpret this videofluoroscopic recording in all of its physiologic complexity and determine which treatments, diet modifications, etc. to recommend. There are thus two sides to this form of assessment: the objective physiologic side and the subjective interpretive side. The objective physiologic side is the actual physical (possibly compromised) biomechanics of the swallowing mechanism. As mentioned above, this involves a cascade of physiologic actions which all contribute to the safe (airway protective) and efficient (bolus clearance) swallow. Therefore to assess “swallowing physiology” inherently means to measure these components of swallowing and their interaction, i.e. it requires a sense of the *profile of physiological impairments*. The subjective element of MBSS primarily comes from three sources: 1) what bolus volumes, consistencies, and methods of administration one *chooses* for the assessment, 2) what physiologic components of the swallow one *chooses* to assess or focus on, and 3) how one *chooses* to rate severity of those physiologic components. These sources of subjectivity have resulted in the proliferation of many non-standardized measurement methods that are often specific to the lab, clinic, or research question (see Swan et al., 2019 for review of methods). This heterogeneity of methods impedes communication across clinicians and comparison across research studies. It was, in part, this heterogeneity of methods and its consequences that led to a need for standardization of the measurement of swallowing physiology.

In response to this need for standardization, the standardized Modified Barium Swallow Impairment Profile (MBSImP) measurement method was developed (Martin-Harris et al., 2008). MBSImP provided a standardized protocol for all three subjective elements of MBS assessments, i.e. the bolus administration and instruction protocol, the physiologic components, and the scoring protocol. The bolus protocol was developed such that the smallest boluses could be safely ingested by even the most severe

patients; the remaining boluses spanned a set of consistencies and methods of administration that each had a known effect on swallowing physiology as later confirmed by their independent predictive value for overall oral and pharyngeal severity (Martin-Harris et al., 2008). A proposed set of physiologic components and scoring rules was chosen using a strong theoretical and empirical basis in the existing literature (Martin-Harris & Jones, 2008). This set of proposed components and scoring rules were discussed and revised using the Delphi method where 10 experts from across disciplines including speech-language pathology, otolaryngology, gastroenterology, and radiology, in the presence of a content-neutral facilitator until 100% consensus was reached. The measurement method underwent rigorous testing of its validity in a sample of 300 patients and was shown to be internally and externally valid on all accounts (Martin-Harris et al., 2008; See Chapter 1 below for details). The result was, and is, a measurement method that consists of: 1) a standardized bolus protocol of 12 swallows with specific volumes, consistencies, and methods of administration, 2) a set of 17 physiologic components each with 3-5 precisely defined levels of impairment that together span the Oral, Pharyngeal, and Esophageal Domains of the swallowing mechanism, and 3) rules for scoring individual swallows as well as for aggregating scores across those 12 swallows into “Overall Impression” (OI) scores. The measurement method was deployed along with an online training that required clinicians to reach 80% agreement to a reference-standard rater to be registered to use the tool. Once registered, clinicians could gain access to an online patient-record system with the option of submitting their de-identified patient records to a centralized database, i.e. the Swallowing Data Registry (SDR) whose data are the foundation that this dissertation is built upon.

MBSImP has had and continues to have widespread uptake. As of April 2023, over 8000 clinicians have been trained and registered to use MBSImP, and the SDR has reached over 85,000 patient records. These records come from clinicians and patients from all 50 states and over 40 countries all over the globe. The SDR was designed to be a clinical tool for clinicians to easily enter, access, compare, and track their patients’ MBSImP scores and therefore all patient records include these standardized physiologic scores. The MBSImP SDR therefore represents the largest ever dataset to date on the physiology of swallowing impairment and presents an opportunity for research that goes far beyond what can be completed in a single dissertation. This dissertation is therefore intended to be the

first steps in a larger arm of research that leverages the uniqueness of the MBSImP SDR to provide real-world characterizations of the measurement and physiology of swallowing impairment. In Chapter 1, we focus on testing the subjective side of MBSImP by examining the degree to which the chosen components and the clinical use of the method, i.e. its psychometric structural validity and psychometric reliability, are maintained in its real-world clinical use following wide-scale dissemination and implementation. In Chapter 2, we focus on the physiologic side of MBSImP to characterize the physiologic swallowing impairment profiles of five of the most commonly dysphagia-associated diagnoses.

Chapter 1 of this dissertation extends the study by Martin-Harris et al. (2008) mentioned above that introduced MBSImP to the field. That study tested the internal and external validity of MBSImP using a sample of 300 patients. Those validity tests included tests of the predictive value of single swallows for Overall Impression (OI) scores, the external validity of MBSImP's relation to swallowing-related outcomes, and the structural validity of the Oral and Pharyngeal Domains. These were all shown to be satisfactory indicating that MBSImP's standardization of the subjective elements of MBS assessment had produced a tool ready for clinical and research implementation. Here we use the SDR to test MBSImP as it is used in real-world clinical practice. Specifically, we reexamine the structural validity of MBSImP in this large clinical dataset, and we extend the validity testing to a test of the internal consistency of these domains. We also use the SDR to guide a smaller-scale study of the reliability of MBSImP so that the results are generalizable to the clinicians who use MBSImP and to the patients who are assessed.

Chapter 2 capitalizes on the unique opportunity afforded by the SDR due to its large scale, physiologic nature, standardized metrics, and in particular, its inclusion of dysphagic patients across many diagnostic populations. Most prior research on dysphagia physiology has been in small studies of single diagnoses using heterogeneous and non-standardized methods (e.g. Barbon et al., 2020; Bingjie et al., 2010; de Deus Chaves et al., 2014; Fattori et al., 2022; Hutcheson et al., 2012; I. S. Kim & Han, 2005; Y. H. Kim et al., 2019; Y. Kim & McCullough, 2010; Langmore et al., 2007; Lee et al., 2015; Mancopes et al., 2020; Miarons et al., 2018; Minagi et al., 2018; Mokhlesi et al., 2002; Namasivayam-MacDonald et al., 2021; Namasivayam-MacDonald & Riquelme, 2019; Park et al., 2010; N. M. Rogus-Pulia et al., 2016; Seo et al., 2016). These limitations make comparison across studies, and therefore

across diagnoses, difficult. The long-term opportunity that the SDR affords is to identify the role that diagnosis plays in the physiology-mediated pathway from cause to sequelae for dysphagia. Chapter 2 of this dissertation takes the first steps toward this goal by characterizing the degree to which different diagnoses have similar or differing impairment profiles.

Chapter 1

Structural Validity, Internal Consistency, and Rater Reliability of the Modified Barium Swallow Impairment Profile (MBSImP): Breaking Ground on a 52,726-Patient, Clinical Dataset

Introduction

The modified barium swallow study (MBSS) using videofluoroscopic (VFS) imaging is the most commonly used diagnostic approach for instrumental assessment of swallowing impairment (Martino et al., 2004; Pettigrew and O'Toole 2007; Rumbach et al. 2018). MBS studies are assessed by Speech Language Pathologists and Radiologists who use visual inspection and their clinical judgement to make informed decisions about the nature of each patient's swallowing impairment and about an appropriate management plan. The use of visual inspection and clinical judgement means that there is a perceptual and subjective aspect to clinician appraisals of MBS studies. Thus, it is of critical importance to verify that MBSS assessments satisfy the requirements of subjective tests, i.e. that they satisfy the required psychometric properties. In particular, it is important to test that MBSS assessments are measuring what they intend to measure, referred to as psychometric validity and that there is agreement across and within clinicians on scoring, referred to as psychometric rater reliability (Lambert et al., 2002, Souza et al., 2017). Testing these psychometric properties becomes especially important when clinicians are using standardized assessment methods. Standardized dysphagia assessments provide several benefits to clinicians and researchers including increased ease of comparing impairment across patients and of tracking impairment trajectories within patients. However, for these benefits to come to fruition, assessment methods must be held to a high standard that includes rigorous testing of their psychometric properties. The present study represents a continuation of the effort to examine the psychometric properties of the standardized Modified Barium Swallow Impairment Profile (MBSImP™), specifically its structural validity, internal consistency, and rater reliability.

MBSImP includes a standardized scale for identifying the nature and severity of physiologic swallowing impairment based on clinician ratings of videofluoroscopic images obtained during MBSS and has had broad, global uptake in the field of dysphagia (Northern Speech Services, 2016). MBSImP includes 17 physiological components rated across 12 swallowing tasks of varying bolus consistencies, bolus volumes, and presentation methods of standardized, customized, commercially prepared, and

stable contrast materials (see Methods for further details). The 17 components are each rated on an ordinal scale, where the scale ranges from no impairment (score of 0) to severe impairment (score of 2, 3, or 4, depending on the component). Both clinically and in research studies, the most common scoring method is referred to as Overall Impression (OI) scoring (e.g., Gullung et al., 2012; Martin-Harris et al., 2015; Wilmskoetter et al., 2018; Xinou et al., 2018; Hutcheson et al., 2012; O'Rourke et al., 2017; Arrese et al., 2017; Arrese et al., 2019; Im et al., 2019; Im and Ko, 2020; Clark et al., 2020). An OI score is determined separately for each MBSImP component by selecting the worst performance (highest score on an ordinal scale) across all the swallowing tasks, which for a single MBSS, results in one OI score for each of the 17 MBSImP components. As OI scoring represents the primary method for using MBSImP both clinically and in research studies, it is these OI scores that will form the basis of all analyses in the present chapter.

Structural Validity

Structural validity refers to the extent to which statistical groupings of items adhere to hypothesized groupings of items (Lambert et al., 2002). For MBSImP, there are 17 components, hypothesized to form three domains: components 1-6 form the Oral domain, components 7-16 form the Pharyngeal domain, and component 17 forms the Esophageal Domain. The structural validity of the multi-component domains, i.e. the Oral and Pharyngeal Domains, was tested and confirmed by factor analysis in a study by Martin-Harris et al. (2008) in a cohort of 300 patients. In the present study, we use a much larger sample (N = 52,726) of clinical patient visits to provide a large-scale test of the structural validity of MBSImP.

Internal Consistency

Internal consistency represents the degree to which sets of items cohere together to measure single constructs. No prior studies have directly investigated the internal consistency of the MBSImP scale. As mentioned above, Martin-Harris et al. (2008) conducted a factor analysis in which MBSImP scores showed a two-factor structure, congruent with the two hypothesized, multi-component domains of the scale: Oral and Pharyngeal. The separation of the components into their hypothesized domains tells us that the domains are valid but does not tell us how well the components of each domain cohere together to form a unified construct. It is common practice in both clinical settings and in research to

assume that each domain forms a unified construct and to then sum up the component scores of each domain into “total” scores, i.e., an Oral Total and a Pharyngeal Total, to characterize the “overall impairment” of each domain (e.g. Hutcheson et al., 2012; Arrese et al., 2017; O’Rourke et al., 2017; Clark et al., 2020; Im and Ko, 2020). This assumption of a unified construct and use of total scores relies on each domain having good internal consistency. Thus, to test whether this assumption holds, here we assess the internal consistency of each of the multi-component domains of MBSImP.

Inter-Rater and Intra-Rater Reliability

Rater reliability, in general, is the degree to which there is agreement between repeated ratings of the same MBS study, where inter-rater reliability representing the agreement between different raters rating the same study and intra-rater reliability represents within-rater agreement when re-rating a previously rated study. In a clinical setting, good inter-rater and intra-rater reliability of an assessment method provides assurance that one can trust both another clinician’s ratings and one’s own previous ratings to be similar to what one would judge. In a research setting, poor reliability of an assessment method results in false reductions in the maximum possible effect size and sensitivity/specificity when examining the relationship between that assessment method and any outcome measure (Lachin 2004). Thus, good rater reliability is of critical importance to clinicians and researchers and should be known for any assessment method. At the time of writing, inter-rater and intra-rater reliability of the MBSImP rating scale is ensured by requiring clinicians to train and reach at least 80% agreement with a gold-standard rater’s scores before they can be certified to use the method. This current method of training to a reliability threshold with a gold-standard guarantees that the clinicians using the method at least agree with the standards set by the MBSImP development team. However, although requiring clinicians to train to reach 80% agreement to a gold standard assures *some* level of agreement across raters, it remains to be examined to what degree this train-to-a-threshold method translates to good inter-rater and intra-rater reliability. We will thus be assessing the inter-rater and intra-rater reliability of MBSImP.

Intent of the Present Study

A recent review of the dysphagia literature showed that, in general, there is a lack of formal reporting of psychometric properties for existing VFS (and fiber endoscopic) swallowing assessment tools (Swan et al., 2019). This general lack of reporting led those authors to conclude that, “there is insufficient

evidence to recommend any individual measure included in the review as valid and reliable to interpret VFSS...". The present work acknowledges the importance of assessing and reporting the psychometric properties of VFS measures of swallowing impairment. The present study therefore represents both an extension of prior work on the psychometric properties of MBSImP (Martin-Harris et al., 2008) and a response to the review by Swan et al. (2019). In particular the aims of this chapter are to 1) provide a further assessment of the structural validity and internal consistency of MBSImP, 2) provide a formal assessment of the inter-rater and intra-rater reliability of MBSImP.

Methods

The present study consists of 1) a retrospective structural validity and internal consistency analysis of patient records, and 2) a prospective study of the inter-rater and intra-rater reliability of MBSImP trained clinicians. Structural validity and internal consistency were assessed based on 52,726 patient records drawn from a centralized repository, the MBSImP Swallowing Data Registry (MBSImP SDR). Inter- and intra-rater reliability were assessed based on the ratings of 4 SLP clinicians on 50 videofluoroscopic (VFS) recordings from MBS studies and re-ratings of 12 such recordings (24% of total cohort).

52,726-Patient Dataset

The dataset for the structural validity and internal consistency analyses was drawn from the MBSImP Swallowing Data Registry (MBSImP SDR). The MBSImP SDR is a centralized electronic health record system that allows MBSImP-registered clinicians to easily store the MBSImP and other de-identified patient data that comes from clinical practice. Thus, the MBSImP SDR consists of data from patient visits entered by clinicians trained and registered to use MBSImP. See Appendix 1 (Supplemental Figures 1-3) for screenshots of the data-entry forms used by clinicians to enter data into the MBSImP SDR. We only included initial visits for each patient to prevent within-subject correlations from affecting the analysis. This left us with a sample size of N = 52,726. Table 1 shows the demographic information for this sample.

N =	52,726						
Sex	Male	27,282	52%	**Ethnicity**	Not Hispanic/Latino	35,317	67%
	Female	23,828	45%		Hispanic/Latino	2,206	4%
	Unknown	1,616	3%		Unknown/Not-Reported	15,203	29%
Age	18-30	1,204	2%	**Race**	White	35,879	68%
	31-40	1,571	3%		Black/African American	6,243	12%
	41-50	3,046	6%		Asian	1,050	2%
	51-60	6,987	13%		American Indian/Alaskan Native	269	0.5%
	61-70	11,243	21%		Native Hawaiian/Pacific Islander	148	0.3%
	71-80	12,605	24%		More than one race	289	0.5%
	81-90	9,827	19%		Unknown/Other/Not Reported	8,848	16.7%
	91 +	4,115	8%				
	Unknown	2,128	4%				

Table 1 Demographic information for the 52,726-patient dataset drawn from the MBSImP Swallowing Data Registry (SDR).

The clinicians who entered data into the MBSImP SDR (N = 6,532) had a wide range of clinical experience. The median number of years treating patients as an SLP was 10 years (5th percentile = 2 years, 95th percentile = 30 years). The median number of years since completing training and becoming registered to use MBSImP was 3 years (5th percentile = <1 year, 95th percentile = 9 years).

The data derive from patients with a wide range of dysphagia associated diagnoses. Patient diagnosis, however, was an optional field for clinicians to enter. Thus, diagnosis information was only available for a subset of patients. Of the patients for whom diagnosis information was available, five of the most common diagnoses were head and neck cancer (N = 2,157), stroke (N = 3,791), chronic obstructive pulmonary disorder (N = 2,033), Parkinson's disease (N = 1,053), and dementia (N = 1,424). The 30 most prevalent diagnoses and their respective sample sizes can be viewed in Supplemental Table 1. The data also come from hospitals and medical centers across the world. These include all 50 states, multiple Canadian provinces, Australia, Norway, the United Kingdom, Singapore, South Korea, among many others. The full list of the geographic locations can be viewed at https://www.mbsimp.com/clinicians.cfm.

The data were collected over 9 years, starting in 2008 with a steady increase in rate of data collections through the present. The number of patient visits input into the dataset from the beginning of 2008 to March of 2020 are shown in Figure 2.

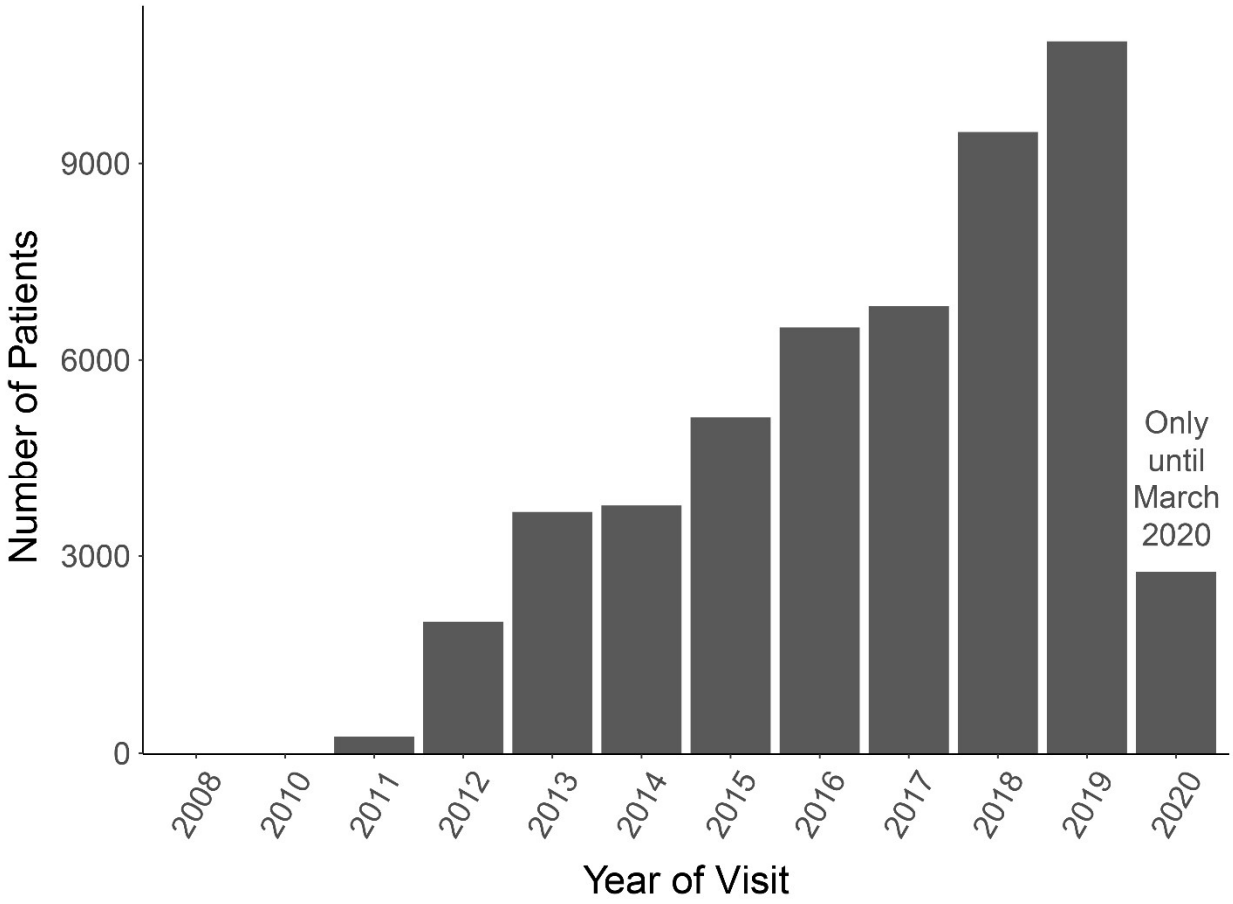


Figure 2. A bar chart of the number of patients entered into the present sample as a function of the year their data was input into the MBSImP Swallowing Data Registry (SDR). Note that the data for 2020 was only that which was collected up until March of that year.

50-Patient Dataset

The second dataset for the present study consists of 50 VFS recordings from modified barium swallow studies. These 50 VFS recordings were randomly drawn from the "high-framerate" (30 fps) recordings of a prior study that was designed to assess possible differences in swallowing assessment between high and low framerate VFSs. The only inclusion criterion for patients was that they were referred for an MBSS and there was no exclusion criterion. The 50 patients fell into the following seven diagnostic categories: General Medicine (N = 24), Head and Neck Cancer (N = 15), Neurology (N = 5), Gastroenterology (N = 4), General Ear-Nose-Throat (N = 2), Pulmonary (N = 2), and Cardiac (N = 1). See Table 2 for the demographic information of the 50-patient sample.

Four SLPs were selected to rate these 50 VFS recordings. In selecting the four SLP raters, we intentionally selected raters with widely varying levels of clinical experience and familiarity with MBSImP. All raters held Master's degrees, were ASHA certified and MBSImP registered. The amount of clinical practice experience ranged from 4-30 years and MBSImP registration ranged between 2-10 years. The number of monthly MBS studies conducted by the raters varied between 0-25 and 0-20 during in clinical practice and clinical research, respectively. These raters also came from differing institutions. Two raters were based in the Medical University of South Carolina, one rater was from Northwestern University, and one rater was based in private practice. See Table 4 for details on each rater.

N =	50						
Sex	Male	26	52%	**Ethnicity**	Not Hispanic/Latino	46	92%
	Female	24	48%		Hispanic/Latino	4	8%
Age	18-30	2	4%		Unknown/Not-Reported	0	0%
	31-40	2	4%	**Race**	White	47	94%
	41-50	5	10%		Black/African American	3	6%
	51-60	7	14%		Asian	0	0%
	61-70	13	26%		American Indian/Alaskan Native	0	0%
	71-80	9	18%		Native Hawaiian/Pacific Islander	0	0%
	81-90	10	20%		More than one race	0	0%
	91 +	2	4%		Unknown/Other/Not Reported	0	0%

Table 2 Demographic information for the 50-patient reliability dataset.

Reliability sample size justification

To choose the number of patients for this study, we focused on what size sample would allow our results to generalize across dysphagic patient populations. In generalizability theory, generalizability is assured when the variability (the variance) of the patient sample accurately estimates the variability of the entire patient population (Webb and Shavelson 2005). We do not have data from the entire dysphagic patient population, we do, however, have a patient database of 52,726 patients that we can use to estimate the generalizability of smaller samples. In order to estimate these smaller sample's generalizability, we have done the following: 1) assumed the variability of the 52,726-patient database represents the true population variance of MBSImP scores of dysphagic patients, 2) calculated the variance of the population MBSImP scores for each component, 3) took 10,000 subsamples (with replacement) of the big dataset with varying sample sizes (10 – 100 patients in steps of 10), 4) calculated the variance of MBSImP component for each subsample, 4) calculated the absolute difference and percent difference between the population variance and each subsample variance. 5) plotted the median difference and median percent difference (error bars represent the 5^{th} and 95^{th} percentile) as a function of sample size (Figure 3A and 3B, respectively). Note that the median difference (Figure 3A) is roughly centered on zero for all sample sizes, showing the subsamples are not systematically over or underestimating the variance of the population. However, also note that in Figure 3A, the range of values (i.e. the size of the error bars) decreases with sample size. This decrease in the range of values can also be seen in the percent difference plot (Figure 3B), where the median percent difference of the subsample vs population variance decreases as a function of sample size.

In Figures 3A and 3B, we observe that sample sizes above 50 patients provide diminishing returns in terms of improving accuracy of variance estimates. Further, with 50 patients, the subsamples had only a median 6% difference in variance relative to the variance of the 52,726-patient dataset, meaning that the variance of a 50-patient sample has a 94% accuracy in estimating the variance of a 52,726-patient sample. We deemed a median of 94% accuracy acceptable for the present study and therefore chose a sample size of 50 patients.

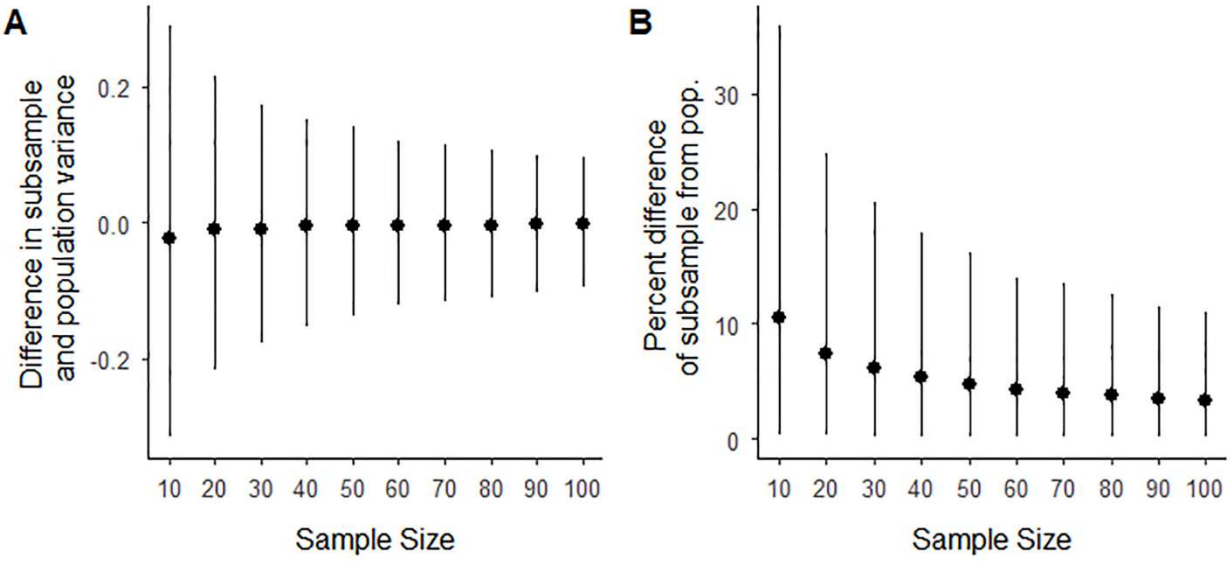


Figure 3. A) The median difference (error bars represent the 95th and 5th percentile) between the variance of the 52,787-patient dataset and the variance of subsamples with sample sizes ranging from 10 to 100. B) The median *percent* difference between the variance of 52,726-patient dataset and the variance of the subsamples, demonstrating that large sample sizes provide better estimates of the between-subjects variance, but have substantially diminishing returns for sample sizes great than ~50 patients.

Similarly to choosing our number of patients, in selecting our raters, we considered how to best estimate the *variance between raters* that occurs in our rater population of interest, i.e. clinicians and clinical researchers. Our total number of raters (N = 4) was limited by logistical constraints, but we attempted to best estimate the variance between raters by selecting raters with the widely varying levels of expertise and familiarity with MBSImP. We have confidence that our raters are representative of the true rater population since the present raters' range of experience (4-30 years practicing clinically; 2-10 years MBSImP registered) mirrors the 5th-95th percentiles of experience of raters in the MBSImP SDR (2-30 years practicing clinically; 0-9 years MBSImP registered).

Protocols

MBSImP Protocol - Clinicians must be trained in the use of the Modified Barium Swallow Impairment Profile (MBSImP) and meet a baseline threshold of scoring reliability (Martin-Harris et al., 2008) before they are permitted to enter data into the MBSImP SDR in order to maintain fidelity of the

data. This training in MBSImP includes both the MBSImP rating scale and the MBSImP bolus administration protocol.

The MBSImP rating scale is a standardized and validated scale of the severity of impairment for 17 physiologic components of swallowing (See Table 3 for list of components). Scores on the rating scale range from 0 to 2, 3, or 4 depending on which physiologic component is being assessed. The number of impairment levels for each component was determined by an expert-panel consensus. For each MBSImP component these experts identified ordered levels of impairment severity that each represented a unique structural movement, bolus flow, or both, related to the physiology of that component. As some components had fewer unique levels than others, this resulted in differing number of impairment levels for different components (Martin-Harris et al., 2008).

The MBSImP bolus administration protocol consists of 12 swallow trials with the following consistencies, volumes, and presentation methods: four thin-liquid (<15 cps) trials (two 5mL via teaspoon, a cup sip [20mL], and sequential swallow from cup [40mL]), four nectar-thick (150-450 cps) trials (two 5mL via teaspoon with one from the typical lateral view and one from an anterior/posterior view, a cup sip, and a sequential swallow), one thin-honey (800-1800 cps) trial (5mL via teaspoon), two pudding (4500-7000 cps) trials (two 5mL via teaspoon with one from the typical lateral view and one from an anterior/posterior view), and one solid trial (a half-portion of a Lorna Doone cookie coated with 3mL of pudding barium). All trials are administered using standardized, “ready-to-use” barium contrast (VARIBAR®, barium sulfate 40% weight/volume; Bracco Diagnostics, Inc., Monroe Township, NJ).

Furthermore, each component is only scored for the swallow trials for which it can be assessed. For example, for the sequential swallow trials, patients are not asked to hold a liquid bolus in the oral cavity, therefore component 2 (tongue control during bolus hold) cannot be assessed and is not scored. Similarly, Component 3 (bolus preparation/mastication) is only scored for the solid bolus trial. For the two swallow trials in the anterior/posterior (A/P) view, the viewing plane provides a perspective ideal for scoring components 13 (pharyngeal contraction) and 17 (esophageal clearance), but this also means no other components are scored from this view.

In clinical practice and typically in research, the MBSImP scores from the individual swallow trials are represented by the “Overall Impression” or OI score for each of the 17 components which is the most

severe score on that component across all swallowing tasks/bolus consistencies. As such, the analyses in the present study are based on OI scores. See Supplemental Figure 4 for the distribution of OI scores per component in the 52,726-patient dataset and Supplemental Figure 5 for the distribution of OI scores per component in the 50-patient dataset.

Inter-Rater and Intra-Rater Protocol

Four SLPs rated ~5 MBS studies per week over a 10-week period. Raters scored each patient's MBSS using the Swallow By Swallow (SbS) method where a score is given for each relevant MBSImP component for each swallowing task given during the protocol. This scoring method results in a total of 127 possible SbS scores for a single MBSS. From these scores an OI score, i.e. the most severe score across swallowing tasks, was computed for each MBSImP component. All analyses in the present study were conducted using OI scores.

The two raters with the least experience with MBSImP (Rater 1 and 4) were allowed to meet with the gold-standard rater once a week to ask questions about scoring for continuing education purposes, but no scores could be changed post-hoc based on these meetings. This question-asking protocol was allowed because in "real-world" settings MBSImP registered clinicians also have the opportunity to send questions to Northern Speech Services and receive guidance on scoring. After all 50 studies were scored, each rater waited 2 weeks and then re-rated 12 studies (~24% of the total) in a like manner to the initial rating for use in the intra-rater analysis.

Statistical Analysis

The data from the 52,726-patient dataset were used to assess the structural validity and internal consistency of MBSImP. Structural validity was assessed by submitting the 17 MBSImP component scores of all patients in the dataset to an exploratory factor analysis. The exploratory factor analysis was computed using polychoric correlations and minimum residuals estimation appropriate to MBSImP components scores as ordinal (Holgado-Tello et al., 2008) and non-normally distributed (Cudek 2000), respectively. Further, an oblimin rotation was used since the hypothesized Oral and Pharyngeal domains were expected to be correlated. The number of factors (2) was chosen based on the hypothesized

number of multi-component domains in the scale, i.e. the Oral and Pharyngeal domains, and also based on a prior study which showed a two-factor solution for MBSImP was adequate (Martin-Harris et al., 2008). An MBSImP component was considered to be a substantive part of a factor if it had a factor loading of >0.4, a criterion set after loadings were computed. To assess the internal consistency, two values of Cronbach's alpha (Cronbach, 1951) were computed: one Cronbach's alpha from the components in the Oral Domain (components 1-6) and one Cronbach's alpha for the Pharyngeal Domain.

The data from the 50-patient dataset were used to assess the inter-rater, and intra-rater reliability of MBSImP. Inter-rater and intra-rater reliability were assessed using the intraclass correlation coefficient (ICC) (Fleiss and Cohen, 1973). ICC was used because it is calculated based on differences in variances across and within raters, and this approach is well suited handling the differing number of severity levels across the 17 components of MBSImP. To compute the ICC, we used a two-way random effects model at the single-rater level, with absolute agreement as the measure of interest. We chose to use a two-way random-effects model as the goal of the model was generalization of the results to clinicians outside of the present sample; this model was chosen to be at the "single-rater" level as our intention was to assess agreement between individual clinicians; and finally the model was chosen to have "absolute agreement" as its basis since our goal is to assess to what extent clinicians have precisely the same score, and not just whether the relative position of scores is consistent across clinicians (Koo and Li, 2016).

All statistical analyses were conducted using the "psych" package (Revelle, 2021) in the R programming language (R Core Team, 2019).

Results

Structural Validity An exploratory factor analysis revealed that MBSImP has a two-factor solution that exactly corresponds to the hypothesized Oral and Pharyngeal domains. Factor 1 had loadings between 0.52 and 0.77 for oral-related components (1-6) and loadings of at most 0.18 for all other components. This factor accounted for 19% of the total variance in MBSImP scores. Factor 2 had loadings between 0.46 and 0.81 for all pharyngeal-related components (7-16) and loadings less than 0.26 for all other components. This factor accounted for 32% of the total variance. Component 17, the only esophageal-related component, had loadings of at most 0.19 for both factors. Exact loadings for each factor and MBSImP component are shown in Table 3. The two factors were correlated with a strength of r = 0.56.

Internal Consistency Cronbach's alpha was 0.81 (95% CI = 0.808-0.812) for the Oral Domain, and 0.87 (95% CI = 0.868-0.872) for the Pharyngeal Domain.

Inter-Rater and Intra-Rater Reliability Inter-rater reliability across all four clinicians as measured by the intraclass correlation coefficient (ICC) was 0.78 (CI = 0.76-0.80). Intra-rater reliability, also measured by ICC for the four clinicians ranged from 0.82 to 0.87 (See Table 4).

Hypothesized Domain	Comp. #	Component Name	Factor 1	Factor 2
Oral	1	Lip closure	**0.77**	-0.08
	2	Tongue control during bolus hold	**0.74**	-0.02
	3	Bolus preparation/mastication	**0.72**	0.07
	4	Bolus transport/lingual motion	**0.76**	-0.01
	5	Oral residue	**0.71**	0.05
	6	Initiation of pharyngeal swallow	**0.52**	0.18
Pharyngeal	7	Soft palate elevation	0.11	**0.46**
	8	Laryngeal elevation	0.08	**0.75**
	9	Anterior hyoid excursion	0.05	**0.76**
	10	Epiglottic movement	-0.08	**0.81**
	11	Laryngeal vestibular closure	0.05	**0.75**
	12	Pharyngeal stripping wave	-0.05	**0.81**
	13	Pharyngeal contraction (A/P view)	-0.01	**0.64**
	14	Pharyngoesophageal segment opening	-0.11	**0.75**
	15	Tongue base retraction	0.26	**0.59**
	16	Pharyngeal residue	0.04	**0.81**
Esophageal	17	Esophageal clearance upright position (A/P view)	-0.05	0.19

Table 3. The hypothesized domains and the factor loadings of each MBSImP component. Factor loadings greater than 0.4 are marked bold and indicate which components can be considered a substantive part of each factor.

				MBSS performed/rated each month			
Rater	Graduation Year of Masters	Years MBSImP Registered	Years Practicing Clinically	For Clinical Purposes	For Research Purposes	Intra-rater Reliability ICC	ICC 95% Confidence Interval
1	2016	2	4	0	10-15	**0.82**	0.77-0.86
2	2014	6	6	20-25	0-10	**0.83**	0.79-0.87
3	2007	10	14	5	10-15	**0.87**	0.83-0.90
4	1990	3	30	0	20	**0.87**	0.83-0.90

Table 4. The clinical, research, and MBSImP experience along with the intra-rater reliability scores and associated confidence intervals of the four SLP raters in the present study.

Discussion

The assessment of the Modified Barium Swallow Study using videofluoroscopic imaging by clinicians requires the use of clinical judgement. Codifying clinical judgement into standardized assessment methods offers a range of benefits including ease of comparison between patients and of tracking of patient trajectories. However in order to reap the benefits of standardization requires that these methods have the necessary psychometric properties (Lambert et al., 2002). A previous investigation of MBSImP showed that multiple aspects of the psychometric validity of MBSImP were quite good, including content validity, hypothesis testing, and structural validity (Martin-Harris et al., 2008; N = 300). The present study complements this prior study by extending the analysis of structural validity and internal consistency to a large-scale, clinical dataset (N = 52,726) and by adding a formal assessment of rater reliability. Both the prior study and the present study use Overall Impression (OI) scores as the basis for analysis due to OI scoring being the main way that MBSImP is used both clinically and in research studies. Thus, the results here show that MBSImP as it is used clinically and in research studies has good structural validity, internal consistency, and inter-rater and intra-rater reliability.

Structural Validity and Internal Consistency

Reports of the structural validity and internal consistency of VFS assessment methods are rare. In part, this may be because many studies of the psychometric properties of VFS assessment methods only examine one or two physiological outcomes (e.g. Rosenbek et al., 1996; Mann et al., 2000; Karnell and Rogus, 2005; Kelly et al., 2006; Hind et al., 2008; Rommel et al., 2015; Hutcheson et al., 2017). Only examining one or two items in a study means that it is either impossible or not very meaningful to examine the correlations between multiple items, which is the basis of structural validity and internal consistency analyses. Even in psychometric studies of VFS assessment methods that examine multiple items (e.g. Gibson et al., 1995; Scott et al., 1998; McCullough et al., 2001; Stoeckli et al., 2003; Frowen et al., 2008, Martin-Harris et al., 2008; Kim et al., 2012; Lee et al., 2017), few examine structural validity and internal consistency (i.e. only Frowen et al., 2008 and Martin-Harris et al., 2008 in the cited examples). Therefore, the present study may serve as a cue to other researchers developing and using VFS assessment methods that it is important to assess structural validity and internal consistency.

The present examination of the structural validity and internal consistency of MBSImP represents a large-scale extension of the work done by Martin-Harris et al. (2008). This prior study showed that in a sample of 300 patients, MBSImP had a two-factor structure, corresponding to the hypothesized Oral and Pharyngeal domains. However, due to missing data and small factor loadings, four components were removed from the analysis (component 1 for inconsistent visualization of lips, component 7 for lack of sufficient variability in sample, and components 13 & 17 for small sample size due to omission of A/P view). The present study extends the analysis of structural validity to a substantially larger sample (N = 52,726), allowing us to include all 17 components in the structural validity analysis and to assess the internal consistency of each of the multi-component domains.

The results of this analysis show that in a large-scale, clinical dataset, MBSImP demonstrates good structural validity and good internal consistency. The present results suggest that MBSImP is structurally valid, as shown by a two-factor structure corresponding to the Oral and Pharyngeal Domain components, respectively, with component 17 (Esophageal clearance, representing the Esophageal Domain) being separable from both of those domains. The good structural validity demonstrated here provides further evidence that the statistical structure of MBSImP is very much aligned with the hypothesized structure of MBSImP.

The present results also show that MBSImP has good internal consistency in that Cronbach's alpha reaches sufficiently high levels for each of the multi-component domains. The good internal consistency of each domain suggests that the components of each domain, as a group, are measuring a latent "overall impairment" variable. This existence of a latent "overall impairment" variable for each domain provides support for the legitimacy of using Oral Total and Pharyngeal Total scores (summed scores of the components of each domain), which are designed to measure this overall oral and pharyngeal impairment, respectively.

The Oral and Pharyngeal domains are structurally valid, but they alone do not fully characterize a patient's impairment. This is evident in the result that the total explained variance of the Oral and Pharyngeal factors accounted for 51% of the variance in the MBSImP assessment method, leaving 49% of the variance to be accounted for by the individual components. This split in the sources of variance in MBSImP means that a patient's impairment is operating at two relatively separable levels: 1) the domain

level characterized by the Oral and Pharyngeal total scores, and 2) the component level characterized by which components are impaired. Each of these levels can potentially be independently influenced by patient-specific factors. This result highlights the importance of conducting assessments and analyses at both the domain level (Oral and Pharyngeal total scores) and at the individual-components level.

Furthermore, although the Oral and Pharyngeal Domains form separable factors, they should be considered separable but related, as the correlation between the factors was r = 0.56. Esophageal function, i.e. component 17 Esophageal Clearance, however, was clearly separable from the Oral and Pharyngeal domains in this data set, since it did not load onto either factor, and showed weak, though not necessarily null, correlations with each oral or pharyngeal component of MBSImP (Table 3 and Supplemental Figure 6). The existence of potentially non-null correlations between the esophageal domain and the oral and pharyngeal domains is consistent with the detection of such correlations in a small case series study by Gullung et al. (2012).

These structural validity and internal consistency results are likely to be highly generalizable due to the dataset's large sample size, clinical nature, wide variety of diagnoses, representative demographics, and diversity of data collection locations. The large sample size of the dataset helps ensure that the correlations that form the basis of these results are not being driven by sampling errors like being overly affected by individual patient scores. The present dataset is also derived from clinical patient visits, which means that the results are likely to be generalizable to our population of interest, i.e. any patient referred to an MBSS conducted by an MBSImP registered clinician. In addition, the 52,726-patient dataset includes scores of patients with any and all diagnoses entered into the MBSImP SDR including but not limited to, head and neck cancer, stroke, chronic obstructive pulmonary disorder, Parkinson's disease, and dementia. This wide variety of diagnoses allows the results to be generalized across populations with differing diagnoses. Furthermore, the demographics of the present data are roughly representative of the demographics of the US population (United States Census Bureau, 2021), admittedly with some underrepresentation of patients who are Hispanic, female, or Asian. This rough representativeness of the present data suggest that the present results are fairly generalizable across US demographics (See Limitations for further discussion). Finally, the present data were collected from a

wide array of locations and institutions across the world. This diversity of data collection locations lends support to the applicability of the present results regardless of geographic location.

Inter-Rater and Intra-Rater Reliability

The literature on the reliability of MBSS assessment scales is quite heterogeneous. Depending on the structures of the scale, studies use a variety of statistics to describe reliability including kappa for dichotomous/binary scales (Gibson et al., 1995; McCullough et al., 2001; Stoeckli et al., 2003; Bryant et al., 2012; Kim et al., 2012; Lee et al., 2017) weighted kappa for ordinal scales with uniform number of levels across items (Bryant et al., 2012, Hutcheson et al., 2017) and ICC for continuous variables, ordinal scales with varying numbers of levels, or in one case a single ordinal item (Frowen et al., 2008; Kim et al., 2012; Rommel et al., 2015). Similarly to the present study, Kim et al. (2012) assessed the inter-rater reliability of the Videofluoroscopic Dysphagia Scale (VDS) using ICC for items with varying numbers of severity levels. That study found that, for VDS, ICC = 0.556, which was deemed a “moderate” level of agreement. Relative to the moderate level of agreement of VDS with ICC = 0.556, the inter-rater ICC of 0.78 for MBSImP found in the present study can be deemed “good”. In addition, the present inter-rater reliability in the present study of ICC = 0.78 (CI = 0.76-0.80) can also be considered “good” as per the guidelines set out by Koo and Li (2016) where an ICC can be deemed “good” if it is >0.75 and <0.9. Along these same lines, the intra-rater ICCs in the present study ranging from 0.82-0.87 can all be considered “good”. The intra-rater ICCs range of 0.82-0.87 is only slightly greater than the inter-rater ICC of 0.78, suggesting that each of the present raters agreed with other raters only slightly less than they agreed with themselves. Furthermore, the two raters with more years of clinical experience (14 and 30 years) showed higher intra-rater ICCs (0.87 and 0.87) than the two raters with fewer years of experience (4 and 6 years), potentially suggesting that SLP’s intra-rater agreement may improve with experience. However, this result should be interpreted with caution because the confidence intervals of all raters are overlapping. A future study could specifically aim to test this hypothesis that intra-rater reliability improves with years of clinical experience.

Together, these results suggest that the standardized MBSImP training that requires clinicians to reach 80% agreement with a gold-standard rater translates to good inter-rater and intra-rater reliability on previously unseen MBSS. Furthermore, these SLPs had widely varying levels of experience across a

range of domains (i.e. years as a SLP, years MBSImP certified, number of studies rated per month, home institution) providing evidence that for these clinicians, it was possible for all experience levels and backgrounds to demonstrate good intra-rater reliability using MBSImP.

In summary, we have shown in a large-scale, clinical dataset that MBSImP has excellent structural validity and internal consistency. In addition, we have provided a formal test of the rater reliability of MBSImP-trained raters and demonstrated that the standardized MBSImP training can result in good inter-rater and intra-rater reliability.

Limitations

1) All results in the present chapter are based on Overall Impression (OI) scores representing the worst performance (highest score on an ordinal scale) across all the swallowing tasks in the MBSImP method. Thus, the results show satisfactory psychometric properties of these OIs, but caution should be applied before generalizing these findings to the psychometric properties of MBSImP ratings for individual swallows. As Swallow By Swallow (SBS) scoring is likely confined to research contexts, the authors recommend that researchers using SBS scoring conduct their own independent rater reliability testing. Furthermore, as the hypothesized structure of MBSImP is at the component level and thus at the level of OI scores, structural validity and internal consistency do not apply to SBS scoring level.
2) The present rater reliability assessment of MBSImP represent the reliability of the *entire* MBSImP scale. A future study with a larger sample of raters and patients would be necessary to investigate the reliability of the individual MBSImP components and would provide substantial value to the field.
3) There is some underrepresentation of Asian, Hispanic, and female populations relative to the US Census (United States Census Bureau, 2021). Interestingly, the representation of Asians in the present sample is quite similar to that seen in the dysphagic population of a study on the prevalence of dysphagia that used a professional survey company to obtain a representative sample of the US population (Adkins et al., 2020). Furthermore, in the present sample 29% of patients did not have their ethnicity input into the MBSImP SDR. It is then possible that the

apparent underrepresentation of Hispanics/Latinos may be due to non-reporting rather than true underrepresentation. In addition, it has been reported that men are more likely to seek care for dysphagia than women (Adkins et al. 2020). This difference in care-seeking behavior may at least partially account for the apparent underrepresentation of women in the present sample. Nonetheless, the apparent underrepresentation of patients who report being Asian, Hispanic, and female could impose a limitation on the generalizability of the results. However, this limitation would only occur if there were systematic differences in MBSImP scores for populations with versus without these demographic identifiers. Future work should compare the MBSImP scores of populations with different demographic identifiers to uncover whether such systematic differences exist.

Chapter 2

Characterizing the Swallow-Physiology Impairment Profiles of Head and Neck Cancer, Stroke, COPD, Dementia, and Parkinson's Disease Using a Large Clinical Database

Introduction

Clinical data registries derived from electronic health records have become a cornerstone of research across the medical sciences. They have been utilized for many purposes including but not limited to the estimation of prevalence and risk factors of disease, the identification patient subpopulations, pragmatic clinical trials, and comparative effectiveness studies (Casey et al., 2016; Cowie et al., 2017). However, clinical data registries are typically not designed to answer specific research questions; rather they are designed for the storage of health records. Therefore, when answering research questions using clinical data registries, the data available in the registry shape the research questions that we can answer.

In the field of dysphagia research, clinical data registries have primarily been used to answer questions about the prevalence, risk factors, and sequalae of dysphagia (Altman et al., 2010; Bonilha et al., 2014; Bray et al., 2017; Cancienne et al., 2016; Eglseer et al., 2018; Guyomard et al., 2009; Henke et al., 2017; Joundi Raed A. et al., 2017; Kawashima et al., 2004; Kidambi et al., 2012; Lo et al., 2019; Patel et al., 2018; Singh et al., 2013) likely because these types of questions only require information on the presence or absence of dysphagia, which can often be found in general-purpose electronic health records or openly available hospital administrative data. These types of large-scale studies provide crucial information that can guide clinical expectation of dysphagia prevalence and risk, and also can be used to assess the severity of dysphagia as a public health issue. The power of large-scale studies, however, has yet to be leveraged toward understanding the swallowing physiology of dysphagia as measured by instrumental assessment. The present study is such a large-scale study of the physiology of dysphagia.

Dysphagia is a disorder that has many possible etiologies whether they be neurologic, oncologic, respiratory, etc. These etiologies each have within them heterogeneous populations and have between them major differences in the nature and severity of their impairments. However, they all have the potential to converge on the swallowing mechanism to cause dysphagia. Any ways in which these

diagnoses differently affect swallowing physiology could reveal important diagnosis-specific issues that could become targets for diagnosis-specific treatments. And conversely, any ways in which these diagnostic categories are similar in how they affect the swallowing mechanism could point towards methods to standardize treatment targets regardless of diagnosis. Furthermore, characterizing these commonalities and differences in swallow-physiology impairment profiles across diagnoses would allow the generation of hypotheses about the varying ways dysphagia manifests. Thus, the research question we sought to answer here was: to what extent do swallow-physiology impairment profiles differ across five of the most commonly dysphagia-associated diagnoses, i.e. Head and Neck Cancer, Stroke, Chronic Obstructive Pulmonary Disorder, Parkinson's Disease, and Dementia? To answer this question we built our analysis around the Modified Barium Swallow Impairment Profile Swallowing Data Registry (MBSImP™© SDR; see below for details), which is a large, de-identified clinical data registry that was expressly built to store physiologic swallowing impairment scores (i.e. MBSImP scores). Our analysis approach consisted of partial proportional odds models, which were used to retain maximal information on all levels of impairment for each diagnosis, along with inverse probability weighting, which aimed to correct for any potential bias introduced by missingness in the dataset. This combined research question and analysis approach was developed based on a clinical and scientific need to understand swallowing physiology that was shaped by the nature of the data that are contained in the MBSImP SDR.

Clinical and Scientific Motivations

Understanding the physiology of swallowing impairment is critical to helping clinicians identify effective targets for treatment. Existing instrumental measures of physiologic swallowing impairment have been shown to be associated with airway protection (swallow safety) and bolus clearance (swallow efficiency) (Barbon et al., 2021), functional swallowing status (Hazelwood et al., 2022), swallowing-specific quality of life (Arrese et al., 2017; Hazelwood et al., 2022; Wishart et al., 2022), and pneumonia risk (Kooi et al., 2019). These associations suggest that treating and improving a patient's physiologic impairments may be a way of broadly improving their outcomes and quality of life. Importantly, however, the "physiology of swallowing impairment" is not a monolith. Swallowing is a complex mechanism that involves multiple organ systems (e.g. neural, muscular, respiratory) and requires the coordinated action of a cascade of physiologic and anatomic movement to control the flow of a bolus safely and efficiently from

mouth to esophagus (Logemann, 1988; Miller, 1986; Walton & Silva, 2018). There are many methods and scales to characterize dysphagia physiology (e.g. temporal measures, kinematics, severity scores), but a common thread between them is that each method includes measures of multiple different physiological actions of the swallow mechanism (Speyer et al., 2021). This consistent inclusion of multiple physiological components of the swallow underlines that swallowing is a multi-faceted and coordinated action of physiology. Thus, when examining the physiology of dysphagia, it is important to characterize a *profile* of impairments, i.e. which components of the swallow mechanism are impaired and to what degree.

Characterizing the commonalities and differences in swallow-physiology impairment profiles *across diagnoses* has the potential to generate hypotheses about the nature and severity of dysphagia in different populations. However, cross-diagnosis comparisons have only been conducted in a handful of studies (Dumican & Watts, 2022; Garand et al., 2018; Mehraban-Far et al., 2021; Tadavarthi et al., 2020). Thus far, in these studies, differences across diagnoses have been detected in penetration/aspiration rates (Dumican & Watts, 2022; Garon et al., 2009; Mehraban-Far et al., 2021; Miles et al., 2019), oral transit times (Dumican & Watts, 2022; Mehraban-Far et al., 2021), pharyngeal swallow reflex impairment (Mehraban-Far et al., 2021), laryngeal vestibular closure temporal measures (Dumican & Watts, 2022), and esophageal transit times (Miles et al., 2019). These studies suggest that comparing across diagnoses may indeed reveal differences in impairment profiles, and therefore may provide a source of characteristic difference for hypothesis generation. The clinical and scientific reasoning for choosing the five diagnoses in the present study, i.e. HNC, Stroke, COPD, PD, and Dementia, was that they are each associated with a high prevalence of dysphagia (García-Peris et al., 2007; González-Fernández et al., 2013; Good-Fratturelli et al., 2000; Ikeda et al., 2002; Kalf et al., 2012; Lindh et al., 2017; Rogus-Pulia et al., 2015).

Data-based Considerations

Theoretically, a comparison of physiologic impairment profiles across different diagnoses could be investigated with meta-analyses of the existing dysphagia literature. There are many studies that investigate the physiology of these diagnoses (e.g. (Curtis et al., 2020; de Deus Chaves et al., 2014b; Fattori et al., 2022; K. L. Garand et al., 2018; Horner et al., 1991; J. S. Kim et al., 2015; Langmore & Krisciunas, 2010; Lin & Shune, 2020; Mancopes et al., 2020; Minagi et al., 2018; Namasivayam-

MacDonald & Riquelme, 2019; N. M. Rogus-Pulia et al., 2014). If these studies could be aggregated, they might readily provide generalizable insights into the physiology of dysphagia. However, there is a substantial amount of heterogeneity in the choice of what physiologic components to measure and how to quantify them (Swan et al., 2019). Review articles that attempt to summarize the literature are limited to reporting lists of typically impaired physiology, which does not allow for direct comparison of the relative severity of diagnoses on those aspects of physiology (Lin & Shune, 2020; B. Patel et al., 2020b; N. Rogus-Pulia et al., 2015; Tjaden, 2008; Wall et al., 2013). The present study is able to side-step this issue of heterogeneity of methods and allow direct comparison of impairment severity by using a single standardized, validated, and reliable physiologic measurement method, i.e. the Modified Barium Swallow Impairment Profile (MBSImP™©). Using this standardized method means data can be easily aggregated and compared. Furthermore, the present study is able to ensure generalizable results by using a large clinical data registry, i.e. the MBSImP Swallowing Data Registry (MBSImP SDR), which has a large sample of patients collected from thousands of clinicians across the world.

MBSImP itself provides a standardized, validated, and reliable method for measuring swallow-physiology impairment profiles (Clain et al., 2022; Martin-Harris et al., 2008). MBSImP has a standardized bolus administration protocol of 12 swallowing tasks and scoring protocol with 17 physiological components, each with 3-5 levels of impairment severity. MBSImP has shown good content validity, external validity, and structural validity, along with good internal consistency (See Chapter 1: Clain et al., 2022; Martin-Harris et al., 2008). MBSImP also has an online training protocol that requires clinicians to reach 80% agreement to a reference-standard rater, a requirement that has been shown to result in good reliability across clinicians (Clain et al., 2022). The standardization and validity ensure that the data are comparable enough across patients, clinicians, clinics, regions, etc. to justify aggregation into a single cohesive dataset. The clinical training and reliability requirements help to ensure the quality and fidelity of data collected from any clinicians who use the MBSImP SDR.

The MBSImP SDR was created for two main purposes: 1) to provide MBSImP-trained clinicians a convenient way to store, access, and track their patients' MBSImP scores, and 2) to generate a large, centralized, de-identified dataset for swallowing research. These purposes together mean that the data contained in the MBSImP SDR consist of clinical patient data from modified barium swallowing studies

entered by clinicians trained and registered to use MBSImP. These data are likely representative of patients referred for MBSS because the data are from patients referred for standard-of-care MBSS without exclusion, the demographics of the dataset roughly match that of the general dysphagic population (Adkins et al., 2020), and the sample size is large (N = 8190 for the five diagnoses). This representativeness and large sample size confer a variety of benefits to the present study. The benefits of these properties are made most clear by examining how the present study extends a proof-of-concept study published by Garand et al. (2018).

Garand et al. (2018) provided evidence that MBSImP provides sufficient sensitivity to detect differences in profiles of impairment of different diagnoses. They found significant differences in MBSImP component impairment rates across patients in five broad medical diagnostic categories, i.e cardiothoracic, head-and-neck cancer, gastroenterologic, neurologic, and pulmonary. Differences were detected mainly when comparing the gastroenterologic category to the other categories. These differences were detected despite the limitations of the study that included its sample size (N = 235), its use of broad diagnostic categories (with small sample sizes within those categories), not controlling for age, sex, and race differences, and dichotomizing MBSImP component scores. The present study addresses each of the limitation of the Garand feasibility project by leveraging the large sample size to allow more specific categories of diagnoses (e.g. COPD vs. Pulmonary), larger sample sizes within diagnoses (>900 per diagnosis), and the use of cumulative logistic regression to allow controlling for age, sex, and race along with the modeling of multiple impairment levels for each MBSImP component.

Using the MBSImP SDR, however, also provides challenges. As mentioned above, MBSImP is a set of 17 physiologic variables each with 3-5 levels of impairment severity. The large sample size of the MBSImP SDR allows us to retain and characterize multiple levels of impairment for each of these physiologic components, rather than dichotomizing each component. However, in order to accomplish this, a statistical approach is required that can account for multiple levels of severity. To this end, we chose to use the proportional odds model (Bender & Grouven, 1997) which, at each possible split of the levels of the scale, estimates the (log) odds of having a score above vs below the split. The proportional odds model, however, requires the assumption that the effect of each covariate is the same regardless of where the cutoff is placed on the ordinal scale. If violated, this assumption can be relaxed for particular

covariates so that the effect of a particular covariate can vary across the cutoffs (Peterson & Harrell Jr., 1990). When this assumption is violated, it can be for a clinically meaningful reason and the relaxing of this assumption can (and in the present study, will) reveal clinically relevant information about the distribution of scores for particular covariates.

Another challenge that arises from using the MBSImP SDR, common in many clinical data registries and electronic health records, is missing data (Kruse et al., 2018; C. H. Lee & Yoon, 2017; Mack et al., 2018a). In the MBSImP SDR, there are two primary sources of missingness. Since the primary purpose of the MBSImP SDR is the storage of MBSImP scores and swallowing-related information for use in clinical practice, non-swallowing-specific medical information was often missing from patient records. This missingness of general medical information informed our decision to investigate across diagnoses simply because diagnosis information was one of the medical fields that had the least missingness. This missingness of medical information means that the MBSImP SDR is best suited for high-level characterization of heterogeneous groups. This type of high-level analysis will inform more controlled smaller-scale studies of the influences on physiologic impairments within individual diagnoses and across diagnoses.

The other source of missingness was within the 17 physiologic MBSImP components. In the MBSImP protocol, two of the MBSImP components (C13: Pharyngeal Contraction & C17: Esophageal Clearance) require the patient to be turned from a lateral viewing plane (wherein 15 of the 17 of components are evaluated) to an anterior/posterior view (Martin-Harris et al., 2008). These two components had high levels of missingness, and based on years of field testing with MBSImP, we hypothesized that this was due to clinicians being less likely to turn patients with higher severity or those that are difficult to reposition. If so, this would mean that patient's swallowing impairment would be systematically biased toward lower severity. Thus, we tested this hypothesis by examining the relationship between patient severity and missingness and corrected for potential bias by using inverse probability weighting (Seaman & White, 2013). Penetration/Aspiration Scale scores and MBSImP component 3 (Mastication) also showed substantial missingness, and so we applied the same inverse probability weighting procedure to them as well.

All taken together, these considerations resulted in the present study being aimed at answering the following question: to what extent do swallow-physiology impairment profiles differ across five of the most commonly dysphagia-associated diagnoses, i.e. Head and Neck Cancer, Stroke, Chronic Obstructive Pulmonary Disorder, Parkinson's Disease and Dementia? This research question and analysis approach are thus the result of using the data available in the large, representative MBSImP SDR to address the clinical and scientific need to understand the physiology of dysphagia, while also addressing the methodological challenges inherent to such a dataset.

Methods

The present study is a cross-sectional analysis of the swallow-physiology impairment profiles of five diagnoses across the 17 components of MBSImP and the Penetration/Aspiration Scale (PAS). The data are derived from the MBSImP Swallowing Data Registry (SDR), which is a real-world clinical database of Modified Barium Swallow Study patient records. The analysis consists of partial proportional odds models with physiologic outcomes as the dependent variables and diagnoses and demographics as the independent variables. Inverse Probability Weighting was used to correct for bias due to missingness for physiologic outcomes with >5% missingness (Seaman & White, 2013). A novel metric related to Stochastic Dominance was used for comparing across diagnoses (Cerchiello et al., 2010).

Dataset

The data for the present analysis were drawn from the MBSImP Swallowing Data Registry (MBSImP SDR) from its inception in 2008 until March of 2020. The data in the SDR come from all 50 states and over 40 countries across the globe. These data are input by MBSImP-trained clinicians who use the MBSImP electronic record system and opt to send a de-identified copy of their patients' records to the MBSImP SDR. Clinicians entering data into the SDR have the option to enter data via an "express form" or a "long form". The express form contains required fields for MBSImP and Penetration-Aspiration Scale (PAS) scores (see below for details), as well as the patient's Primary Diagnosis and demographics. The long form includes additional fields on comorbidities, past medical history, functional swallowing measures, patient reported outcomes, treatment strategies, diagnosis-specific information, etc., however data for these fields were entered much less frequently. Therefore MBSImP scores, PAS scores, Primary

Diagnosis and demographics formed the basis of the present analysis. See Supplemental Figures 1-3 for screenshots of the data-entry forms used by clinicians to enter data into the MBSImP SDR.

Of the many diagnoses in the SDR, five diagnoses were chosen to form the basis of the present analysis due to their high rate of occurrence in the SDR and their known association with swallowing impairment (García-Peris et al., 2007; González-Fernández et al., 2013; Good-Fratturelli et al., 2000; Ikeda et al., 2002; Kalf et al., 2012; Lindh et al., 2017; Rogus-Pulia et al., 2015). These five diagnoses are stroke (N = 3,342), Head and Neck Cancer (HNC; N = 2399), Dementia (N = 1066), Chronic Obstructive Pulmonary Disorder (COPD; N = 995), and Parkinson's disease (PD; N = 923). If there were multiple visits in the SDR for a single patient, only the record of the first visit was included. The distribution of sex, race, and age among the selected diagnoses is displayed in Table 5; also shown are tests of independence that show that there are significant differences across the diagnoses for each of those demographic variables. Hispanic/Latino status had a much larger amount of missingness than the other demographics (26% vs < 1%) as it was often reported as "Unknown/Unreported" and therefore was excluded from the analysis but is included in the table for reference. Distributions of diagnosis-specific information (e.g. subtype of Dementia, the lesion location for Stroke, and cancer staging and treatment for HNC) are included in Supplemental Tables 2-4.

	Stroke (N=3342)	HNC (N=2399)	Dementia (N=1066)	COPD (N=995)	PD (N=923)	Test of Independence
Sex						
Male	1831 (54.8%)	1653 (68.9%)	563 (52.8%)	537 (54.0%)	621 (67.3%)	Chi Sq.
Female	1496 (44.8%)	735 (30.6%)	488 (45.8%)	452 (45.4%)	298 (32.3%)	df = 4 | X^2 = 172
Missing	15 (0.4%)	11 (0.5%)	15 (1.4%)	6 (0.6%)	4 (0.4%)	p < 0.0001
Age (years)						
Mean (SD)	71.7 (14.2)	64.7 (12.1)	82.0 (10.3)	74.3 (11.4)	76.2 (9.93)	ANOVA
Median [Min, Max]	73.0 [19.0, 95.0]	65.0 [18.0, 95.0]	83.0 [36.0, 95.0]	75.0 [30.0, 95.0]	77.0 [28.0, 95.0]	df = 4 | F = 407
Missing	15 (0.4%)	11 (0.5%)	15 (1.4%)	6 (0.6%)	4 (0.4%)	p < 0.0001
Race						
Asian	109 (3.3%)	43 (1.8%)	38 (3.6%)	12 (1.2%)	31 (3.4%)	Chi Sq.
Black/African American	703 (21.0%)	291 (12.1%)	115 (10.8%)	112 (11.3%)	39 (4.2%)	df = 12 | X^2 = 281
White	2159 (64.6%)	1843 (76.8%)	767 (72.0%)	750 (75.4%)	710 (76.9%)	p < 0.0001
Other or Not Reported	356 (10.7%)	211 (8.8%)	131 (12.3%)	115 (11.6%)	139 (15.1%)	
Missing	15 (0.4%)	11 (0.5%)	15 (1.4%)	6 (0.6%)	4 (0.4%)	
EthnicityID						
Non-Hispanic/Non-Latino	2245 (67.2%)	1870 (77.9%)	601 (56.4%)	559 (56.2%)	600 (65.0%)	Chi Sq.
Hispanic/Latino	136 (4.1%)	59 (2.5%)	34 (3.2%)	24 (2.4%)	29 (3.1%)	df = 4 | X^2 = 18.3
Unknown/Not Reported/Missing	961 (28.8%)	470 (19.6%)	431 (40.4%)	412 (41.4%)	294 (31.9%)	p = 0.0011

Table 5. The distributions of demographic variables for each of the five diagnoses. Also included are tests of independence which show that the distributions of each demographic variables significantly differ across diagnoses. Tests of independence were conducted across on all non-missing data.

Protocols

MBSImP Protocol - Clinicians must be trained in the use of the Modified Barium Swallow Impairment Profile (MBSImP) and meet a baseline threshold of scoring reliability (Martin-Harris et al., 2008) before they are permitted to enter data into the MBSImP SDR in order to maintain fidelity of the data. This training in MBSImP includes both the MBSImP rating scale and the MBSImP bolus administration protocol.

The MBSImP rating scale is a standardized and validated scale of the severity of impairment for 17 physiologic components of swallowing (See Table 3 for list of components). Scores on the rating scale range from 0 to 2, 3, or 4 depending on which physiologic component is being assessed (Martin-Harris et al., 2008). The Penetration-Aspiration Scale (PAS) is also typically scored as part of the MBSImP scoring protocol. PAS is an 8-point scale developed separately from MBSImP that describes the degree to which the bolus entered the airway and whether it was subsequently ejected.

The MBSImP bolus administration protocol consists of 12 swallow trials with the following consistencies, volumes, and presentation methods: four thin-liquid (<15 cps) trials (two 5mL via teaspoon, a cup sip [20mL], and sequential swallow from cup [40mL]), four nectar-thick (150-450 cps) trials (two 5mL via teaspoon with one from the typical lateral view and one from an anterior/posterior view, a cup sip, and a sequential swallow), one thin-honey (800-1800 cps) trial (5mL via teaspoon), two pudding (4500-7000 cps) trials (two 5mL via teaspoon with one from the typical lateral view and one from an anterior/posterior view), and one solid trial (a half-portion of a Lorna Doone cookie coated with 3mL of pudding barium). All trials are administered using standardized, “ready-to-use” barium contrast (VARIBAR®, barium sulfate 40% weight/volume; Bracco Diagnostics, Inc., Monroe Township, NJ).

Furthermore, each component is only scored for the swallow trials for which it can be assessed. For example, for the sequential swallow trials, patients are not asked to hold a liquid bolus in the oral cavity, therefore component 2 (tongue control during bolus hold) cannot be assessed and is not scored. Similarly, Component 3 (bolus preparation/mastication) is only scored for the solid bolus trial. For the two swallow trials in the anterior/posterior (A/P) view, the viewing plane provides a perspective ideal for scoring components 13 (pharyngeal contraction) and 17 (esophageal clearance), but this also means no

other components are scored from this view. When no A/P view is captured, Component 17 can be scored in the lateral view.

In clinical practice and typically in research, the MBSImP scores from the individual swallow trials are represented by the "Overall Impression" or OI score for each of the 17 components and PAS, which is the most severe score on that component across all swallowing tasks/bolus consistencies. As such, the analyses in the present study are based on OI scores.

Model-Specification

For each component we followed the same procedure for determining the final model. We first determined that the appropriate model for the present data and research question was 1) a cumulative logistic regression model where 2) the outcome variable is an MBSImP component and 3) the independent variables are the diagnoses, age, and demographic variables (i.e. race and sex).

We used a cumulative logistic regression to retain the maximal amount of information from the component scores, allowing all of the potentially clinically meaningful variability in severity of impairment to be available for differentiating between diagnoses. Oftentimes MBSImP scores are either summed into Oral Total and Pharyngeal Total scores (e.g. Clark et al., 2020; Garand et al., 2018; Im et al., 2019; Kooi et al., 2019; Wilmskoetter et al., 2019) or dichotomized into unimpaired vs impaired (e.g. Garand et al., 2018; Vose et al., 2019; Wilmskoetter et al., 2019). Using Oral or Pharyngeal total scores prevents identification of the specific physiology driving difference in severity and dichotomizing scores means that, among impaired patients, any differences in severity are ignored. Therefore, we opted to use the cumulative logistic regression for each component to allow for analysis of each specific physiologic component and for to allow for differences in severity amongst impaired patients to influence the impairment estimates of each diagnosis.

We included age, race, and sex as independent variables as their distributions differed across diagnoses (as mentioned above) and there is some evidence that each of these variables may have an association with physiologic impairment (Ahn et al., 2020; Daniels et al., 2017a; Dozier et al., 2006; Gall et al., 2010; Gonz et al., 2011; Hiss et al., 2004; Mehraban-Far et al., 2021; Mogensen et al., 2013), though the evidence for sex differences is mixed.

Co-occurrence of these five diagnoses was generally rare in the present dataset. Co-occurrence of any two (of the five) diagnoses was defined as when a patient had both of those diagnoses in their record, regardless of whether they were listed as Primary Diagnosis (of which there could be multiple), Comorbidities, or were in their past medical history. Co-occurrence between each pair of diagnoses was less than 6% except for between Dementia and Parkinson's, where there was 13% co-occurrence (N = 267 patients with both). To account for this 13% co-occurrence, we added an interaction term between Parkinson's Disease and Dementia, which allows for the estimation of the impairment levels of patients specifically with Parkinson's with Dementia (PDwDem).

We also included age x diagnosis interactions in our models. Across the five diagnoses, there were substantial differences in the distribution of ages (See Table 5). In addition, age has been shown to be associated with increased impairment of swallowing physiology within healthy populations (Feng et al., 2013; Hiramatsu et al., 2015; Jardine et al., 2020; Mancopes, Gandhi, et al., 2021; Wang et al., 2015) and in diagnostic populations (Ahn et al., 2020; Mehraban-Far et al., 2021). We therefore decided to test whether allowing the effect of age to differ for each diagnosis (i.e. including an interaction term between age and each diagnosis) would significantly improve the model fit compared to a model where the effect of age is consistent across all diagnoses (i.e. no interaction term). We compared model fit using likelihood ratio tests and found that for all components, the model with the age interaction had higher likelihood than the model without age interactions (using $\alpha < 0.1$; See Table 6).

Component	Df	Chisq	Pr(Chisq)
C1 - Lip Closure	5	32.3	5.3×10^{-6}
C2 - Bolus Hold	5	16	0.007
C3 - Mastication	5	26.7	6.5×10^{-5}
C4 - Lingual Motion	5	13.2	0.022
C5 - Oral Residue	5	16.5	0.006
C6 - Ph. Swall. Init.	5	51.9	5.5×10^{-10}
C7 - Soft Palate Elev.	5	10.5	0.063
C8 - Laryn. Elev.	5	43.2	3.4×10^{-8}
C9 - Hyoid Excursion	5	24.6	1.7×10^{-4}
C10 - Epiglot. Move.	5	18.6	0.002
C11 - Laryn. Vest. Clos.	5	46.4	7.6×10^{-9}
C12 - Phary. Strip	5	34.2	2.2×10^{-6}
C13 AP - Phary. Contract.	5	15	0.010
C14 - PES open.	5	12.3	0.031
C15 - Tongue Base Ret.	5	16.1	0.006
C16 - Phary. Residue	5	23.5	2.7×10^{-4}
C17 AP - Esoph. Clear	5	19.4	0.002
Penetration / Aspiration	5	44.2	2.1×10^{-8}

Table 6 Likelihood ratio tests for each outcome variable comparing models with versus without age x diagnosis interactions.

For all components, we first fit cumulative logistic regressions using a proportional odds assumption across all covariates. However, when HANC patients were included in the models for the other components, there were substantial deviations from the proportional odds assumption in the binary score residuals (Harrell, 2015). When the HANC patients and covariates were removed from the model, the deviations in the binary score residuals either disappeared or were reduced to acceptable levels for all other variables. Therefore, we relaxed the proportional odds assumption for HANC and left the proportional odds assumption intact for all other covariates. The only components where the we left the proportional odds assumption intact for all covariates (including HNC) were Oral Residue (C5), Pharyngeal Contraction (C13), Tongue Base Retraction (C15), Esophageal Clearance (C17), and Penetration/Aspiration.

For nine of the seventeen components, in order to allow model fit to converge, some scores needed to be combined. If a component's severity score for any diagnosis contained less than ~30 patients, that severity score was combined with the next less severe score. Due to the most severe scores invariably being the least prevalent, when this score-collapsing procedure was necessary, it typically resulted in the two most severe ordinal levels being collapsed into a single score. Specifically the two most severe scores were collapsed into a single severity score for Lingual Motion (C4), Oral Residue (C5), Pharyngeal Swallow Initiation (C6), Laryngeal Elevation (C8), Pharyngeal Stripping Wave (C14), Tongue Base Retraction (C15), Pharyngeal Residue (C16), and Esophageal Clearance (C17); the three most severe scores were collapsed into a single score for Soft Palate Elevation (C7). These collapsed scores were treated as the most-severe ordinal value in the regression analyses. For computing the Normalized Mean OI Score (see below for details), however, the score of the collapsed category was taken as the average of the uncollapsed scores across the full dataset. For example, if scores of 3 and 4 were collapsed and there were three times as many 3s as 4s in the dataset, the collapsed score would be 3.25. In addition, Penetration Aspiration Scale Scores were collapsed into three levels to ensure ordinality: 1-2 for healthy-normal, 3-6 for penetration, and 7-8 for aspiration.

All analyses were conducted using the R programming language (R Core Team, 2022). Initial proportional odds models were fit using the 'rms' package (Harrell Jr., 2023); partial proportional odds models were fit using the 'VGAM' package (Yee & Moler, 2023).

Missing Data in Outcome Variables

Four of the seventeen MBSImP components had greater than 5% missing data. Furthermore, these rates of missingness varied across diagnoses and across patient severity as quantified by Oral Total and Pharyngeal Total scores (omitting the five components above). This association between diagnosis/severity and missingness suggests that the pattern of missing data in the MBSImP SDR is not completely random. Rather it could be considered "Missing at Random" where the variables in the MBSImP SDR can potentially predict which patients are likely to have missingness for which components (Mack et al., 2018b).

We leveraged this potential association to use Inverse Probability Weighting (IPW) to attempt to correct for any bias introduced in the non-random pattern of missingness (Seaman & White, 2013). For each of the four components with >5% missingness (C3, C13, C17, and PAS), we derived inverse probability weights from a logistic regression where the outcome variable was whether the data were complete (0 vs 1) and the explanatory variables were diagnosis, an oral sum score, and a pharyngeal sum score derived from the remaining 13 MBSImP components (oral sum = sum[C1, C2, C4, C5, C6], pharyngeal sum = sum[C7, C8, C9, C10, C11, C12, C14, C15]). These explanatory variables were chosen because we expected that more severe patients (as determined by sum scores) would have higher missingness. This was found to be the case, especially for c3, c13, and c17, as can be seen in Figure 4. The weights derived from these logistic regressions, i.e. the inverse of the probability of having complete data, were truncated if any of the weights were >100 such that weights larger than 100 were truncated to the 99th percentile of the weights (Seaman & White, 2013). This was necessary for component 3 (Bolus Preparation / Mastication) and PAS such that their max weights of 103 and 154 were truncated to 18 and 33, respectively. These weights were then applied to the data of each of their respective MBSImP/PAS models to produce final estimates.

Logistic regressions for the IPW models were conducted using 'stats' package in the R programming language (R Core Team, 2022).

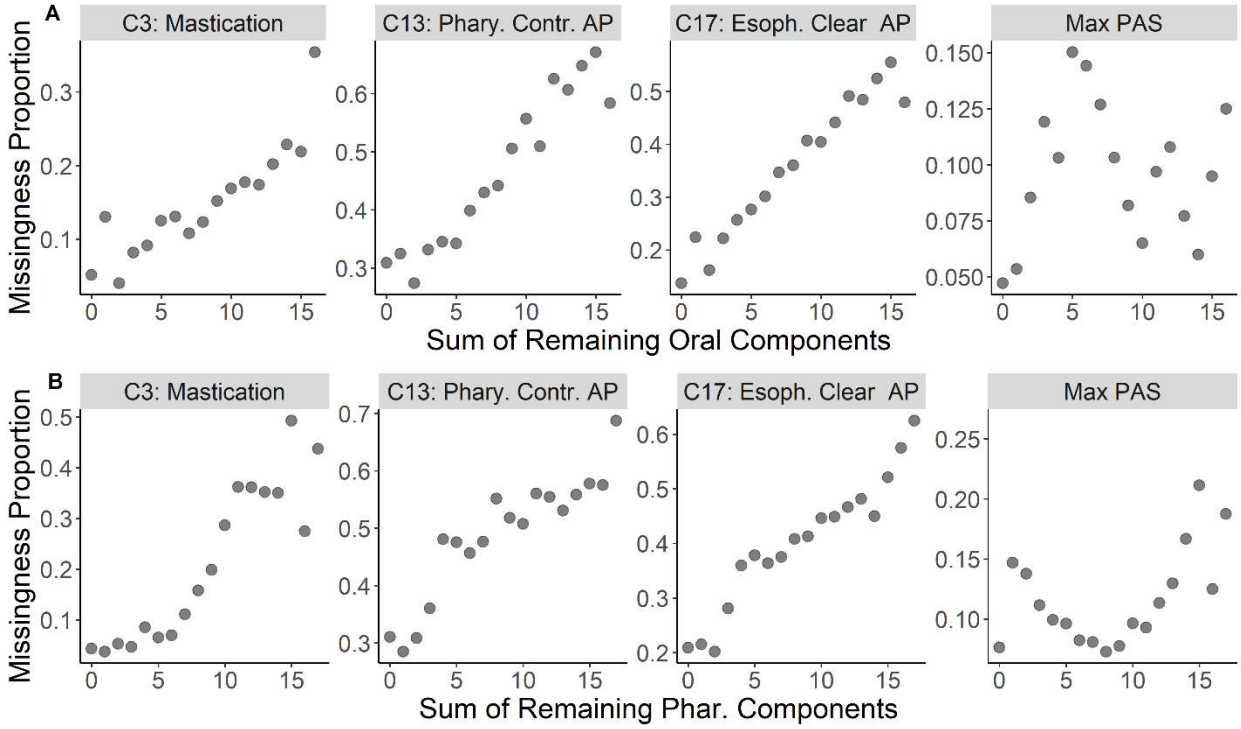


Figure 4 For the four components with >5% missingness, here is plotted the proportion of missing data versus the sum of the remaining components scores (a proxy for patient severity). A) The plot where the severity sum is for Oral components. B) The plot where the severity sum is for the Pharyngeal Components.

Model Outputs and Plots

Due to the violation of the proportional odds assumption by the HNC group, we could not use odds ratios as the overall statistic for comparing diagnoses as the odds ratio would differ depending on the impairment level of the components. Instead, for each component, we computed a statistic that we will refer to as the Normalized Mean Overall Impression (OI) Score. To compute the Normalized Mean OI Score for a particular MBSImP component, we first acquire the predicted probabilities for each impairment level unconditional on race and sex (see below for details). We then multiply those predicted probabilities by their respective MBSImP scores and normalized to (i.e. divide by) the max score for that component. This provides a statistic that is bounded between 0 and 1 where 0 means every patient is predicted to have no impairment, i.e. a score of 0, and 1 means that every patient is predicted to have the max score for that component. This metric can also be thought of as a normalized version of the stochastic dominance index presented by Cerchiello et al., (2010) that is altered so it is bounded between 0 and 1

and so that higher values mean worse impairment. The Normalized Mean OI score can therefore be equivalently defined as follows: Normalized Mean OI Score = $\frac{\sum_{i=0}^{K} 1-F_i}{K}$, where i is the impairment score, K is the max impairment score, and F_i is the cumulative distribution function of the predicted probabilities at impairment level i.

The predicted probabilities required to calculate this normalized score were generated from the fitted cumulative logistic regression models. To ensure that any comparisons of the Normalized Mean OI Scores were unconditional on age, race, and sex, we computed these predicted probabilities as weighted averages across race and sex for each diagnosis (Harrell, 2015). In practice, this meant that for each diagnosis we first acquired the predicted probabilities of each impairment level for each race and sex combination (e.g. Asian-Male, Black-Male, etc). For each impairment level, we then computed a weighted average of the probabilities across all race-sex combinations, where the weight was the frequency of the race-sex combination in the entire dataset. Normalized Mean OI Scores were then calculated based on the resulting weighted averages of probabilities.

Age, sex, and race variables did not violate the proportional odds assumption and thus their differences could be compared directly using odds ratios. As there were significant improvements in model fit when age x diagnosis interactions were included in the models, we extracted an age-related odds ratio for each diagnosis for each component. As race and sex were independent of diagnosis and age in the present model, we computed a single odds ratio for each of these variables for each component. Using odds ratio for the categorical variables of race and sex requires an arbitrary choice of a "baseline" group for each; "Asian" was chosen as the baseline for race and "female" was chosen as the baseline for sex.

Confidence Interval & P-value Estimation

Confidence intervals and p-values for all variables were computed using custom bootstrapping to estimate the variability of model estimates (Efron & Hastie, 2016). Custom bootstrapping conferred two benefits: 1) confidence intervals could be estimated for the Normalized Mean OI Scores which would otherwise be difficult to obtain given that the proportional odds model more naturally produces estimates of odds ratios; and 2) for the components with substantial missingness, the uncertainty from the model-

based IPW weights could be incorporated in the width of the confidence intervals and the resulting p-values. We incorporated the IPW-based uncertainty into our confidence intervals by re-calculating IPW weights for each bootstrap sample and using those weights to compute the estimates for each bootstrap, thus allowing sampling error in the IPW estimation to impact the overall sampling error represented in the final confidence intervals.

Ten thousand bootstrapped samples were used to calculate both the confidence intervals and the p-values. Final confidence intervals were calculated from these bootstraps using the bias-corrected and accelerated method (DiCiccio & Efron, 1996). P-values were calculated in two steps: first, subtracting the model estimate from each bootstrap estimate to produce a distribution of bootstrapped estimates centered on the null; second p-values were calculated as the percentage of null-centered bootstrap estimates whose absolute value was greater than or equal to the absolute value of the main model estimate, i.e. as extreme or more extreme than the observed value. For comparing Normalized Mean OI Scores, the threshold for significance was set at $\alpha = 0.01$ to balance the exploratory nature of this analysis with the potential false positives due to the substantial number of paired comparisons between diagnoses [270 comparisons = (17 components + PAS) x (all possible comparisons between 5 diagnoses + PDwDem)]. As age, sex, and race were primarily just control variables, the threshold for significance for these variables was at $\alpha = 0.05$.

Results

MBSImP Impairment Rates and Score Distributions

Figure 5 presents the model-predicted distributions of scores for each component and diagnosis. This figure serves as a reference for the underlying distributions that produce the Normalized Mean OI score, and also illustrates the reason for the deviation of HNC from the proportional odds assumption. In addition, Table 7 presents the impairment rates of each component as calculated from the raw data of each diagnostic group.

In Figure 5, the order of the diagnoses on the x-axis is a ranking from least to most severe impairment, as measured by their Normalized Mean OI score. The size of each colored bar within each diagnosis represents the expected proportion of patients with that impairment level. The violation of the proportional odds assumption can be seen most clearly in Laryngeal Elevation (C8). For Laryngeal

Elevation, HNC's impairment rate (and by definition, its odds of impairment) is lower than Stroke and PDwDem, but HNC is ranked higher in severity. The reason for this is that HNC has higher rates (and thus odds) of the most-severe score, despite having lower impairment rates. HNC's higher rates of the most severe score can also be seen clearly for Mastication (C3).

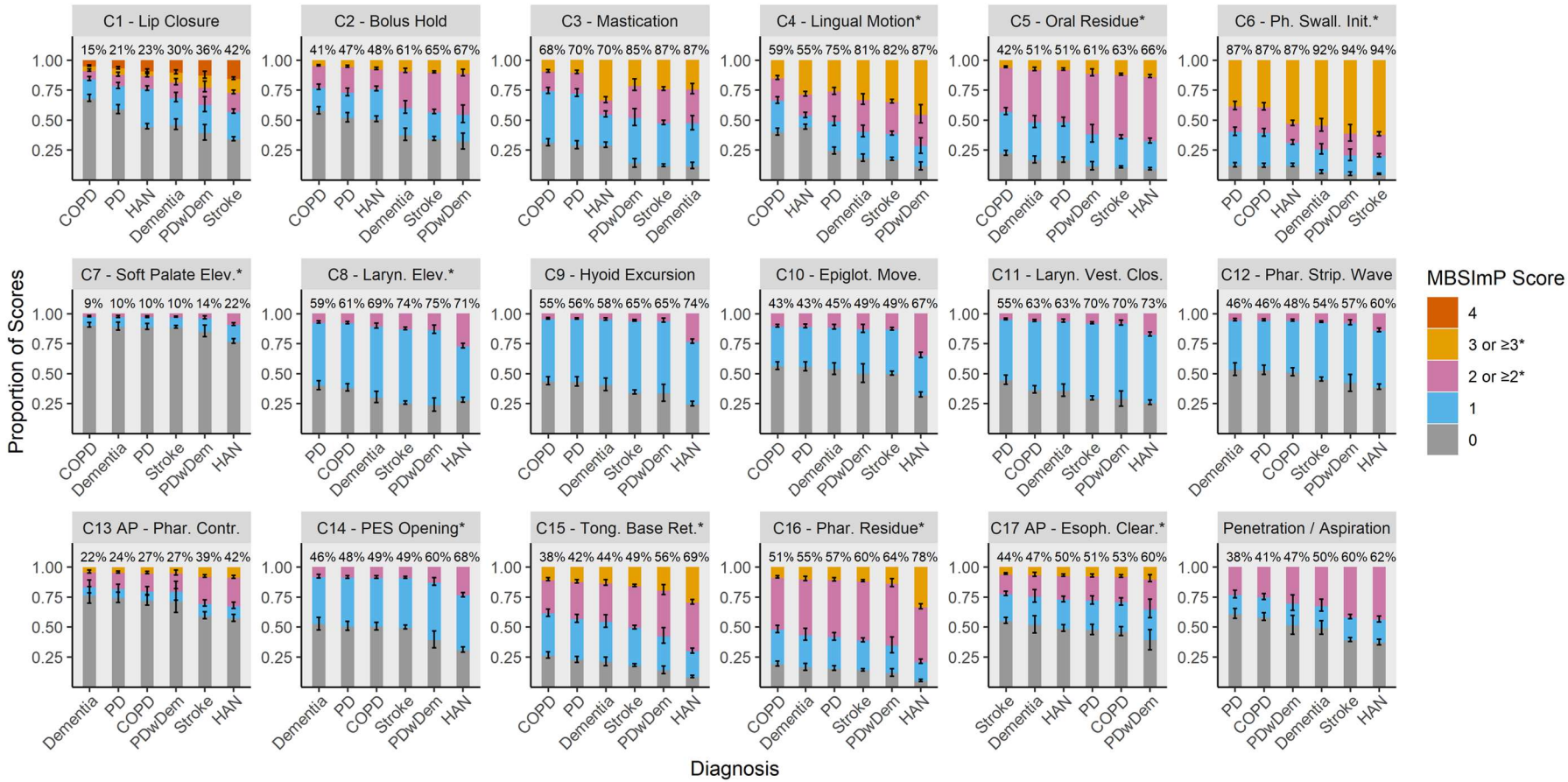


Figure 5 The model-based expected proportion of scores for each impairment level at median age of the dataset (72) across the diagnoses for each MBSImP component and Penetration/Aspiration. Diagnoses are ordered from least to most impaired as measured by Normalized Mean OI Score. The percentage of patients with any amount of impairment is displayed above the bar plot for each diagnosis. Components for which the most severe scores were collapsed are marked with an *. Error bars represent 95% confidence intervals.

	Oral Components					
Diagnosis	1–LC	2–BH	3–BP/M	4–BT	5–OR	6–IPS
COPD	12.9%	42.0%	68.2%	60.8%	41.4%	85.5%
Dementia	26.2%	60.6%	87.5%	83.8%	48.0%	91.8%
HNC	21.4%	47.5%	64.9%	53.9%	65.5%	86.4%
PD	22.5%	54.1%	75.6%	78.7%	52.4%	89.9%
PDwDem	32.7%	68.0%	89.5%	88.0%	56.9%	93.6%
Stroke	43.3%	64.7%	86.8%	82.4%	63.8%	94.7%

	Pharyngeal Components											
Diagnosis	7–SPE	8–LE	9–AHE	10–EM	11–LVC	12–PSW	13–PC	14–PES	15–TBR	16–PR	17–EC	Pen/Asp
COPD	9.6%	61.4%	54.3%	43.3%	63.0%	47.0%	24.1%	49.6%	38.4%	51.6%	54.0%	42.8%
Dementia	10.7%	72.1%	57.0%	44.3%	65.5%	50.6%	21.9%	50.0%	44.7%	56.9%	50.6%	52.5%
HNC	21.8%	69.2%	73.6%	66.1%	71.3%	58.3%	37.6%	67.0%	67.0%	77.4%	45.3%	63.5%
PD	11.9%	66.5%	60.8%	47.0%	63.1%	53.5%	25.3%	54.6%	47.1%	61.2%	54.0%	46.5%
PDwDem	15.0%	76.8%	66.9%	52.1%	72.3%	63.5%	26.9%	63.9%	58.0%	68.6%	57.5%	53.0%
Stroke	10.3%	73.5%	64.5%	48.1%	69.1%	53.0%	32.4%	48.4%	49.2%	58.9%	44.0%	59.8%

Table 7. The impairment rates for each MBSImP component and penetration/aspiration as calculated from the raw data.

Diagnosis Impairment Profiles

In the Oral Domain, the five diagnoses (and PDwDem) appear to be separable into the three following groups: 1) Stroke, Dementia, and PDwDem, 2) PD and COPD, and 3) HNC. Stroke, Dementia, and PDwDem patients consistently had significantly worse impairment than COPD or PD without Dementia [all $p < 0.007$]. The exception to this was for Oral Residue (C5), where Dementia patients and PDwDem patients were not significantly different from PD [$p = 0.96$ and $p = 0.02$ respectively]. Among the Stroke, Dementia, and PDwDem patients, Stroke patients had significantly worse impairment than Dementia patients for Lip Closure (C1) and Oral Residue (C5) [both $p < 0.0001$], otherwise they were not significantly different [all $p > 0.01$]. On the other hand, patients with PDwDem had worse impairment than those with either Stroke or Dementia for Lingual Motion (C4) [all $p < 0.005$], and otherwise were not significantly different from patients with Stroke or Dementia [all $p > 0.02$]. Between PD and COPD, the PD group had significantly worse impairment than the COPD group for Lip Closure (C1), Lingual Motion (C4), and Oral Residue (C5) [all $p < 0.0003$]; and were not significantly different otherwise [all $p > 0.02$]. HANC appeared to show its own characteristic impairment profile, sometimes having high impairment not significantly different from the Stroke and PDwDem groups as with Mastication (C3) and Oral Residue (C5) [all $p > 0.1$], while other times having lower impairment either not significantly different from PD as with Bolus Hold (C2) [$p = 0.99$] or significantly lower than PD as with Lingual Motion (C4) [$p = 0.0002$], and in the cases of Lip Closure (C1) and Pharyngeal Swallow Initiation (C6) having impairment significantly below Stroke and PDwDem [all $p < 0.004$], but still above COPD and PD [all $p < 0.0003$].

In the Pharyngeal Domain, the five diagnoses (plus PDwDem) appear to separate into the four following groups: 1) HNC, 2) Stroke and PDwDem, 3) PD, COPD, and Dementia,. The HNC group had consistently worse impairment than every other diagnosis for nearly all components [all $p < 0.006$]; the exceptions were Laryngeal Elevation (C8) and Pharyngeal Stripping Wave (C12) where HNC was not significantly different from PDwDem [$p = 0.03$ & $p = 0.04$, respectively], along with Pharyngeal Contraction (C13) where HNC was not significantly different from Stroke [$p = 0.31$]. Stroke showed consistently worse impairment than both COPD and PD [all $p < 0.006$]. The exceptions to this were Soft Palate Elevation (C7) and Pharyngoesophageal Segment Opening (C14), where Stroke was not significantly different from either COPD or PD, and Pharyngeal Residue (C16) where it was not

significantly different from PD [all $p > 0.1$]. PDwDem had a similar profile to Stroke, though its confidence intervals often overlapped with other diagnoses where Stroke did not. Similar to Stroke, it had significantly worse impairment than PD and COPD for Laryngeal Elevation (C8) and Tongue Base Retraction (C15) [all $p < 0.0001$]; in other cases, it was only significantly worse than COPD as with Pharyngeal Residue (C16) [$p = 0.0003$] or only significantly worse than PD as with Laryngeal Vestibular Closure (C11) [$p < 0.0001$]. One uniqueness PDwDem had was that it was significantly worse than all other diagnosis except HNC for Pharyngoesophageal Opening (C14) [all $p < 0.007$]. In the remaining components, it was not significantly different from Stroke, PD, COPD, or Dementia [all $p > 0.01$]. Dementia, PD, and COPD were generally not significantly different from each other with the exceptions of Laryngeal Elevation where Dementia had significantly worse impairment than COPD and PD [$p = 0.0042$ & $p = 0.0009$, respectively], Laryngeal Vestibular Closure (C11) where COPD and Dementia had significantly worse impairment than PD [$p = 0.0016$ & $p = 0.0089$, respectively], and Pharyngeal Residue (C16) where PD had significantly worse impairment than COPD [$p = 0.0051$].

In the Esophageal Domain, i.e. Esophageal Clearance (C17), Stroke had significantly lower impairment than all other diagnoses [all $p < 0.004$], except Dementia [$p = 0.42$]. Otherwise there were no significant differences between the remaining diagnoses [all $p > 0.02$].

For Penetration / Aspiration, Stroke and HNC patients showed the worst impairment, significantly above the other diagnoses [all $p < 0.003$], and not significantly different from each other [$p = 0.18$]. Dementia was the next most impaired with significantly worse impairment than PD or COPD [$p = 0.0073$ and $p = 0.001$ respectively], but not significantly different from PDwDem [$p = 0.63$]. PDwDem, COPD, and PD were all not significantly different from each other [all $p > 0.04$].

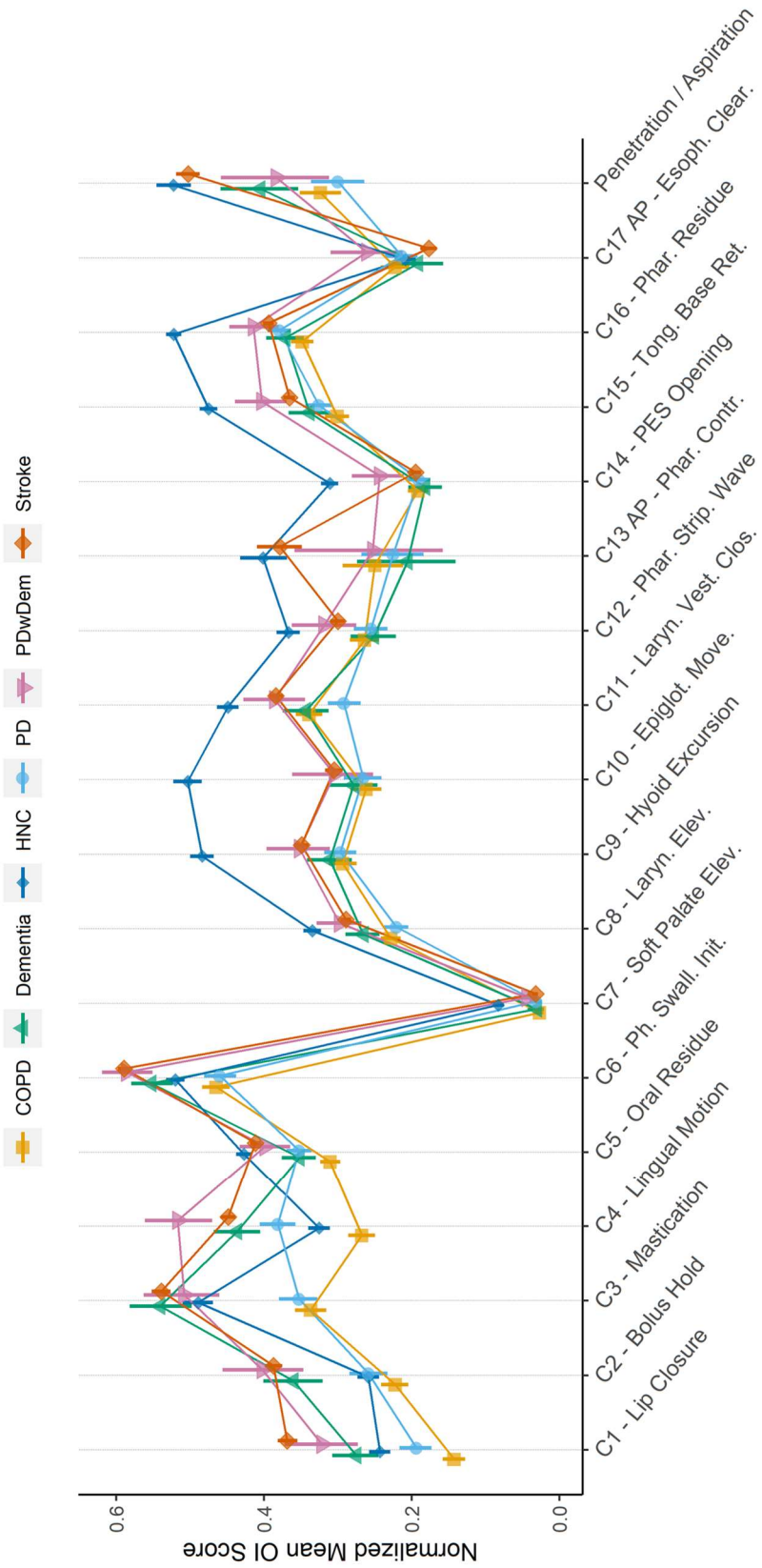


Figure 6 The impairment profiles of diagnoses at median age of the data (72) as Normalized Mean OI Scores across 17 MBSImP Components and Penetration/Aspiration. Error bars represent 95% confidence intervals.

		Oral Components					
Diag. 1	Diag. 2	1–LC	2–BH	3–BP/M	4–BT	5–OR	6–IPS
COPD	Dementia	***<0.0001***	***<0.0001***	***<0.0001***	***<0.0001***	***0.002***	***<0.0001***
COPD	HAN	***<0.0001***	***0.0025***	***<0.0001***	***<0.0001***	***<0.0001***	***<0.0001***
COPD	PD	***0.0002***	0.0228	0.3551	***<0.0001***	***0.0001***	0.723
COPD	PDwDem	***<0.0001***	***<0.0001***	***<0.0001***	***<0.0001***	***<0.0001***	***<0.0001***
COPD	Stroke	***<0.0001***	***<0.0001***	***<0.0001***	***<0.0001***	***<0.0001***	***<0.0001***
Dementia	HAN	0.0657	***<0.0001***	0.0331	***<0.0001***	***<0.0001***	0.037
Dementia	PD	***<0.0001***	***<0.0001***	***<0.0001***	***0.0066***	0.9644	***<0.0001***
Dementia	PDwDem	0.114	0.2212	0.3358	***0.0048***	0.0297	0.1349
Dementia	Stroke	***<0.0001***	0.1973	0.949	0.4744	***<0.0001***	0.0123
HAN	PD	***0.0002***	0.9887	***<0.0001***	***0.0002***	***<0.0001***	***<0.0001***
HAN	PDwDem	***0.0035***	***<0.0001***	0.4999	***<0.0001***	0.1164	***0.0004***
HAN	Stroke	***<0.0001***	***<0.0001***	***0.0001***	***<0.0001***	0.0157	***<0.0001***
PD	PDwDem	***<0.0001***	***<0.0001***	***<0.0001***	***<0.0001***	0.0202	***<0.0001***
PD	Stroke	***<0.0001***	***<0.0001***	***<0.0001***	***<0.0001***	***<0.0001***	***<0.0001***
PDwDem	Stroke	0.072	0.6023	0.2564	***0.0038***	0.4577	0.8471

Table 8 The p-values for the pairwise comparisons between each pair of diagnoses for each MBSImP component and Penetration/Aspiration. Bold/italics indication significant differences between diagnoses at for an $\alpha = 0.01$ significance threshold.

		Pharyngeal Components									
Diag. 1	Diag. 2	7–SPE	8–LE	9–AHE	10–EM	11–LVC	12–PSW	13–PC	14–PES	15–TBR	16–PR
COPD	Dementia	0.5367	***0.0042***	0.3292	0.3844	0.811	0.4912	0.2682	0.4767	0.0224	0.1264
COPD	HAN	***<0.0001***	***<0.0001***	***<0.0001***	***<0.0001***	***<0.0001***	***<0.0001***	***<0.0001***	***<0.0001***	***<0.0001***	***<0.0001***
COPD	PD	0.3412	0.4832	0.82	0.815	***0.0016***	0.5577	0.4054	0.9288	0.0381	***0.0051***
COPD	PDwDem	0.0507	***<0.0001***	0.0147	0.1487	0.0409	0.0261	0.9507	***0.0061***	***<0.0001***	***0.0003***
COPD	Stroke	0.1344	***<0.0001***	***<0.0001***	***0.0003***	***0.0001***	***0.0019***	***<0.0001***	0.674	***<0.0001***	***<0.0001***
Dementia	HAN	***<0.0001***	***<0.0001***	***<0.0001***	***<0.0001***	***<0.0001***	***<0.0001***	***<0.0001***	***<0.0001***	***<0.0001***	***<0.0001***
Dementia	PD	0.872	***0.0009***	0.4475	0.5241	***0.0089***	0.8564	0.6312	0.5286	0.4611	0.5681
Dementia	PDwDem	0.1631	0.102	0.1246	0.4217	0.1014	0.0132	0.4259	***0.003***	***0.007***	0.0383
Dementia	Stroke	0.7828	0.054	0.0176	0.1467	0.0156	***0.0027***	***<0.0001***	0.2762	0.0668	0.0823
HAN	PD	***<0.0001***	***<0.0001***	***<0.0001***	***<0.0001***	***<0.0001***	***<0.0001***	***<0.0001***	***<0.0001***	***<0.0001***	***<0.0001***
HAN	PDwDem	***0.0001***	0.0296	***<0.0001***	***<0.0001***	***0.0052***	0.0444	***0.0053***	***0.0002***	***0.0003***	***<0.0001***
HAN	Stroke	***<0.0001***	***<0.0001***	***<0.0001***	***<0.0001***	***<0.0001***	***<0.0001***	0.3067	***<0.0001***	***<0.0001***	***<0.0001***
PD	PDwDem	0.1718	***<0.0001***	0.0264	0.2026	***<0.0001***	0.0123	0.5998	***0.0061***	***<0.0001***	0.0618
PD	Stroke	0.9061	***<0.0001***	***<0.0001***	***0.0055***	***<0.0001***	***0.0006***	***<0.0001***	0.6358	***<0.0001***	0.1148
PDwDem	Stroke	0.1542	0.6033	0.9002	0.9978	0.9331	0.4308	0.0186	***0.0067***	0.0416	0.24

Table 8 continued. The p-values for the pairwise comparisons between each pair of diagnoses for each MBSImP component and Penetration/Aspiration. Bold/italics indication significant differences between diagnoses at for an $\alpha = 0.01$ significance threshold.

Diag. 1	Diag. 2	17–EC	Pen/Asp
COPD	Dementia	0.1336	***0.0073***
COPD	HAN	0.2614	***<0.0001***
COPD	PD	0.5629	0.3018
COPD	PDwDem	0.1485	0.129
COPD	Stroke	***0.0001***	***<0.0001***
Dementia	HAN	0.3931	***0.0001***
Dementia	PD	0.2964	***0.001***
Dementia	PDwDem	0.0257	0.6298
Dementia	Stroke	0.4259	***0.0007***
HAN	PD	0.696	***<0.0001***
HAN	PDwDem	0.0445	***0.0005***
HAN	Stroke	***0.0006***	0.1776
PD	PDwDem	0.0817	0.0442
PD	Stroke	***0.0039***	***<0.0001***
PDwDem	Stroke	***0.0011***	***0.0022***

Table 8 continued. The p-values for the pairwise comparisons between each pair of diagnoses for each MBSImP component and Penetration/Aspiration. Bold/italics indication significant differences between diagnoses at for an $\alpha = 0.01$ significance threshold.

Age-Associated Changes in Impairment Profiles

As age did not violate the proportional odds assumption, for each diagnosis & component we represent the association of age and impairment using an odds ratio that represents the increase or decrease in the odds of having worse impairment per 1-year increase in age, as shown in Figure 7; the associated p-values are shown in Table 9.

COPD and HNC were the only two diagnoses for which a majority of components showed significant age-associated increases in the odds of having worse impairment. For COPD, three of the six Oral components, six of the ten Pharyngeal components, as well as Esophageal Clearance (C17) and Penetration/Aspiration all showed significant age-associated increases in odds of having worse impairment [all $p < 0.05$]. For HNC, four of six Oral components, nine of ten Pharyngeal Components, the Esophageal Domain (C17) and Penetration-Aspiration all had significant increases in odds with increasing age [all $p < 0.05$].

PD and PDwDem both primarily showed age-associated increases in the odds of having worse impairment in the Pharyngeal Domain, with Mastication (C3) having the only odds ratio significantly greater than 1 in the Oral Domain for PD [$p = 0.019$], and no other Oral components being significant across the two diagnoses [all $p > 0.15$]. In the Pharyngeal Domain, Soft Palate Elevation (C7), Epiglottic Movement (C10), Pharyngeal Stripping Wave (C12), and Pharyngeal Residue (C16) all had significant age-associated increases in odds of having worse impairment across both PD and PDwDem [all $p < 0.05$]. For PD, but not PDwDem odds ratios for Anterior Hyoid Excursion (C9) and Penetration/Aspiration were significantly greater than 1 [$p = 0.037$ & $p = 0.007$, respectively]. For PDwDem, but not PD, the odds ratio for PES Opening (C14) was significantly greater than 1 [$p = 0.037$].

For Stroke, age was significantly associated with increased odds of having worse impairment for Mastication (C3), Lingual Motion (C4), and Esophageal Clearance (C17) [all $p < 0.05$], and significantly associated with decreased odds of having worse impairment for Lip Closure (C1), Oral Residue (C5), and Pharyngeal Swallow Initiation (C6) [all $p < 0.04$]; no other components for Stroke showed significant associations with age [all $p > 0.05$]. For Dementia, age was significantly associated with decreased odds of having worse impairment for Lip Closure (C1) [$p < 0.0001$], but was otherwise not significant [all $p > 0.06$].

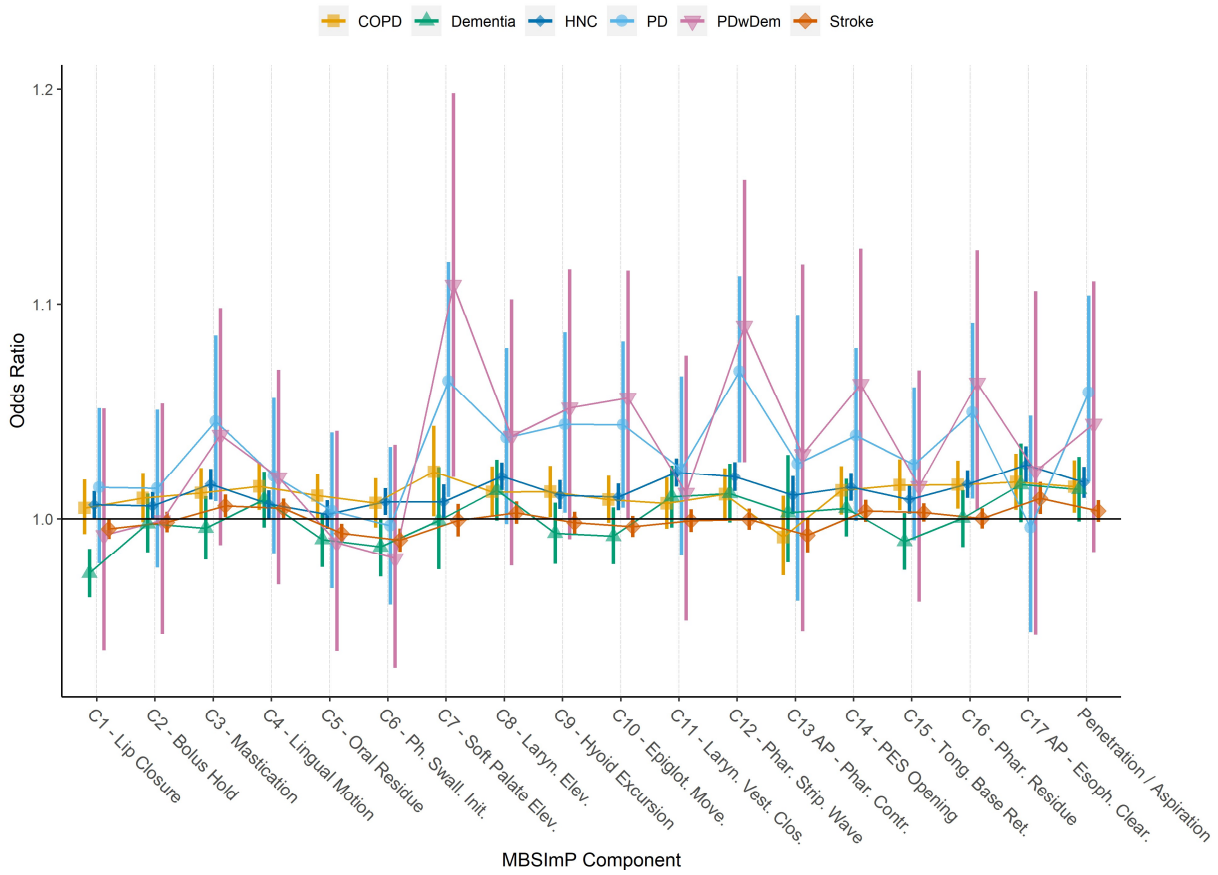


Figure 7 shows the odds ratios associated with a 1-year increase in age for each diagnosis across the 17 MBSImP Components and Penetration/Aspiration. The horizontal black line represents no age-associated change in odds of having worse impairment. Error bars represent 95% confidence intervals.

	Oral Components					
Diagnosis	1–LC	2–BH	3–BP/M	4–BT	5–OR	6–IPS
COPD	0.4253	0.0942	***0.0444***	***0.0071***	***0.0441***	0.2057
Dementia	***<0.0001***	0.7528	0.5609	0.1826	0.1301	0.0613
HNC	***0.0452***	0.0647	***<0.0001***	***0.0376***	0.5647	***0.0125***
PD	0.4261	0.4494	***0.0199***	0.2822	0.8425	0.8609
PDwDem	0.7944	0.9777	0.1527	0.4461	0.6734	0.4953
Stroke	***0.0492***	0.5142	***0.0238***	***0.0385***	***0.003***	***0.0003***

	Pharyngeal Components											
Diagnosis	7–SPE	8–LE	9–AHE	10–EM	11–LVC	12–PSW	13–PC	14–PESO	15–TBR	16–PR	17–EC	Pen/Asp
COPD	***0.0401***	***0.0393***	***0.0387***	0.1146	0.2329	0.0625	0.3578	***0.024***	***0.0109***	***0.0058***	***0.011***	***0.0181***
Dementia	0.9459	0.0664	0.352	0.233	0.1586	0.0931	0.8194	0.4887	0.1168	0.9605	0.0893	0.0736
HNC	0.0572	***<0.0001***	***0.0022***	***0.0014***	***<0.0001***	***<0.0001***	***0.015***	***<0.0001***	***0.0079***	***<0.0001***	***<0.0001***	***<0.0001***
PD	***0.0188***	0.0633	***0.0366***	***0.0257***	0.2678	***0.0013***	0.442	0.0545	0.162	***0.0132***	0.8734	***0.0069***
PDwDem	***0.0136***	0.2114	0.0956	***0.0497***	0.7022	***0.0051***	0.4878	***0.0372***	0.5745	***0.0329***	0.5966	0.1543
Stroke	0.8474	0.2896	0.4743	0.1437	0.73	0.9071	0.0658	0.16	0.1831	0.9566	***0.0116***	0.1662

Table 9 The p-values for the age-associated odds ratio of each diagnosis for each MBSImP component and Penetration/Aspiration. Bold/italics indicate significant effects of age at an $\alpha = 0.05$ significance threshold.

Sex- and Race-Associated Differences in Impairment Profiles

Sex and Race were primarily included as control variables in the present study. Nonetheless there were significant differences across sexes and races for many of the MBSImP components. Figure 8A shows the odds ratios for Male verus Female patients, and Figure 8B shows the odds ratios for each race versus Asian patients. Table 10 shows the p-values of each sex- and race-associated odds ratio for each MBSImP component. For sex, Male patients had significantly worse odds of having worse impairment than Female patients for Lip Closure (C1) and Bolus Hold (C2) in the Oral Domain, for all Pharyngeal components (with the exception of Soft Palate Elevation – C7), and for Esophageal Clearance (C17) and Penetration/Aspiration [all $p < 0.02$]. For race, one consistent finding was that in the Oral Domain, Black/African American patients had significantly higher odds of having worse impairment than Asian patients for all components except Mastication (C3) [all $p < 0.05$]. In addition, in the Pharyngeal Domain Black/African American and White patients, as well as patients all other patients (i.e. “Other” and Not Reported) had significantly higher odds of having worse impairment than Asian patients for Soft Palate Elevation (C7), Tongue Base Retraction (C15), and Pharyngeal Residue (C16) [all $p < 0.02$].

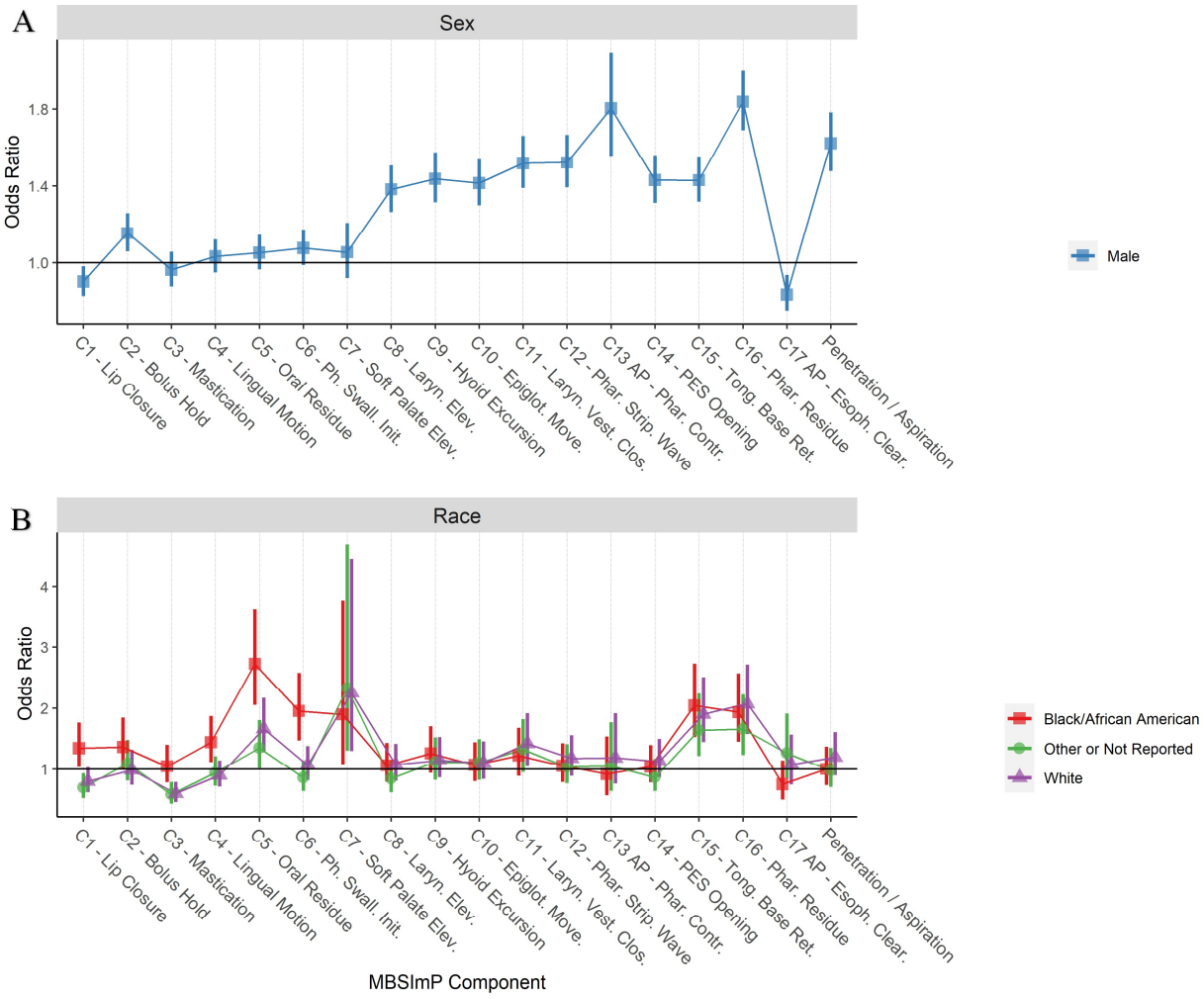


Figure 8 The odds ratios for sex and race across MBSImP Components and Penetration/Aspiration. The horizontal black line represents no difference from the baseline comparison category. A) Odds ratios for male patients versus female patients. B) Odds ratio for who reported their race as Black/African American, White, and Unknown/Not Reported versus Asian.

		Oral Components					
Variable	Group	1–LC	2–BH	3–BP/M	4–BT	5–OR	6–IPS
Sex	Female	-	-	-	-	-	-
	Male	***0.0177***	***0.0014***	0.4011	0.455	0.262	0.1021
Race	Asian	-	-	-	-	-	-
	Black/AA	***0.0295***	***0.0425***	0.7576	***0.0044***	***<0.0001***	***<0.0001***
	Other / NR	***0.0125***	0.5587	***0.0002***	0.6182	***0.042***	0.2892
	White	0.0765	0.9051	***0.0001***	0.3762	***0.0002***	0.6407

		Pharyngeal Components											
	Group	7–SPE	8–LE	9–AHE	10–EM	11–LVC	12–PSW	13–PC	14–PES	15–TBR	16–PR	17–EC	Pen/Asp
Sex	Female	-	-	-	-	-	-	-	-	-	-	-	-
	Male	0.4634	***<0.0001***	***<0.0001***	***<0.0001***	***<0.0001***	***<0.0001***	***<0.0001***	***<0.0001***	***<0.0001***	***<0.0001***	***0.0022***	***<0.0001***
Race	Asian	-	-	-	-	-	-	-	-	-	-	-	-
	B/AA	***0.0468***	0.7	0.1275	0.6594	0.2113	0.7236	0.7351	0.7908	***<0.0001***	***<0.0001***	0.1701	0.9934
	O/NR	***0.0155***	0.2825	0.4722	0.5125	0.0946	0.8101	0.8564	0.3411	***0.0017***	***0.0011***	0.2605	0.8665
	White	***0.0152***	0.6252	0.3431	0.5001	***0.0237***	0.2547	0.464	0.3838	***<0.0001***	***<0.0001***	0.7002	0.227

Table 9 The p-values for the sex- and race-associated odds ratio of each diagnosis for each MBSImP component and Penetration/Aspiration. Bold/italics indicate significant effects of age at an α = 0.05 significance threshold. B/AA = Black/African-American; O/NR = Other/Not Reported.

MBSImP^{lot} for Comparing Populations and Individuals

Using the models we developed to compare the diagnoses, we can also compare the impairment profiles of individual patients to the model-expected impairment profile for a patient with the same diagnosis and age. Figure 9 shows an example of such a plot for a 79-year-old Dementia patient. We can see that the patient has higher than average impairment for Lip Closure (C1), Mastication (C3), Lingual Motion (C4), Oral Residue (C5), Hyoid Excursion (C9), Laryngeal Vestibular Closure (C11), and Pharyngeal Stripping Wave (C12); they had approximately average impairment for Bolus Hold (C2), Soft Palate Elevation (C7), Laryngeal Elevation and (C8); and they had below average impairment for Pharyngeal Swallow Initiation (C6), Epiglottic Movement (C10), Pharyngeal Contraction (C13), PES Opening (C14), Tongue Base Retraction (C15), Pharyngeal Residue (C16), Esophageal Clearance (C17), and Penetration/Aspiration.

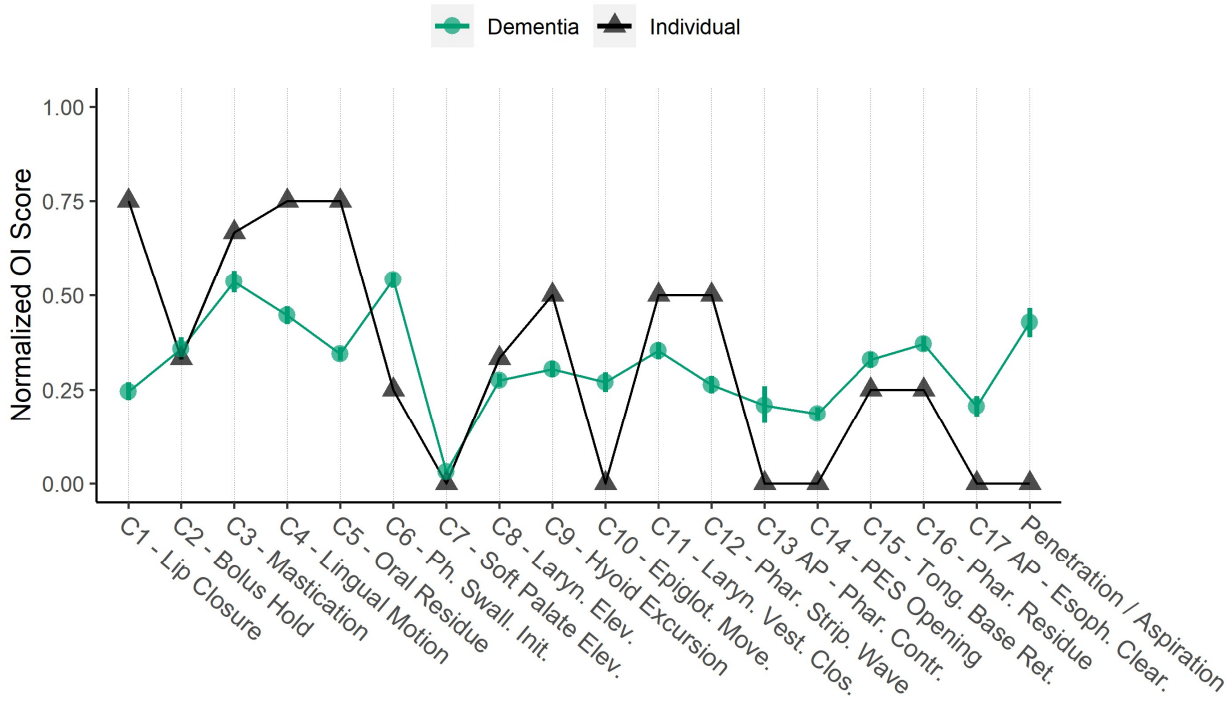


Figure 9 The MBSImP^{lot} showing the measured impairment levels and the model-expected impairment levels for a 79-year-old patient with Dementia across the 17 MBSImP components and Penetration/Aspiration.

Discussion

The present study characterized swallow-physiology impairment profiles of five dysphagia-associated diagnoses using a standardized, validated, and widely used clinical method. The characterizations of swallow-physiology impairment profiles in the present study can be used to generate hypotheses and data-driven predictions about the nature and severity of dysphagia in clinical patient populations. The general findings of the present study include 1) COPD and PD having similar impairment profiles, 2) Stroke, Dementia and PDwDem having significantly worse Oral Impairment than PD or COPD, 3) Stroke having worse Pharyngeal Impairment than COPD or PD, 4) HNC have a unique Oral Impairment profile as well as having the worst Pharyngeal Impairment on nearly every component, 5) HNC and Stroke having the worst Penetration/Aspiration scores, and 6) Dementia having worse Penetration/Aspiration scores than COPD or PD.

In the following discussion, we will show how the present results relate to the general literature on swallowing-physiology impairments in our five diagnoses, as well as prior studies that have directly compared those diagnoses. We will also discuss evidence for the clinical significance of our findings and then highlight key commonalities and differences found between diagnoses, along with hypotheses generated by those results. In addition, we will highlight results that support the importance of acquiring the AP view. We will conclude with cautions in the interpretation of the age/race/sex results, a discussion of the potential uses of the MBSImP^{lot}, and limitations of present work.

Prior Studies on Swallowing Physiology

Much of the prior research on the physiology of swallowing in the five diagnoses examined here has focused on examining the diagnoses alone or comparing the diagnoses (or subpopulations thereof) to controls (e.g. Barbon et al., 2020; Bingjie et al., 2010; de Deus Chaves et al., 2014; Fattori et al., 2022; Hutcheson et al., 2012; I. S. Kim & Han, 2005; Y. H. Kim et al., 2019; Y. Kim & McCullough, 2010; Langmore et al., 2007; Lee et al., 2015; Mancopes et al., 2020; Miarons et al., 2018; Minagi et al., 2018; Mokhlesi et al., 2002; Namasivayam-MacDonald et al., 2021; Namasivayam-MacDonald & Riquelme, 2019; Park et al., 2010; N. M. Rogus-Pulia et al., 2016; Seo et al., 2016). This research approach is quite valuable because it reveals which physiologic components are impaired in which populations, and thus

which physiologic components have the potential to be targets for treatment. However, due to heterogeneity in which physiologic components are measured and how impairment is defined, direct comparison of diagnoses is difficult. Reviews of this swallow-physiology literature are typically limited to lists of impairments that are characteristic of a particular diagnosis (e.g. Lin & Shune, 2020; Patel et al., 2020; Rogus-Pulia et al., 2015; Tjaden, 2008). The issue is that these lists of impairments tend to be highly overlapping, which on the one hand could mean diagnoses have similar impairment profiles, or on the other hand that all diagnoses have *some* impairment for the components in the lists, but particular diagnoses have worse impairment than others.

A few prior studies have directly compared diagnoses and found results that are either fully consistent with or can be reconciled with the findings of the present study. Mehraban-Far et al. (2021) found that Dementia and Stroke had longer Oral Transit Times (OTT) and more frequent Premature Pharyngeal Entry than HNC; similarly, Garand et al. (2018) found a Neurologic group had worse Bolus Hold (C2) than HNC. Both of these studies are thus consistent with our finding that Dementia and Stroke have generally worse Oral impairments than HNC. In addition, Dumican & Watts (2022) found Stroke had longer OTT than PD, which is consistent with our finding that Stroke patients have worse Oral impairments than PD patients. There are nonetheless some apparent discrepancies between prior studies and the present study. These discrepancies, however, can be potentially explained by differences in study samples and by what variables were or were not controlled. For example, Dumican & Watts (2022) found that PD had worse pharyngeal impairments than Stroke, including worse vallecular residue, laryngeal elevation, and laryngeal vestibular closure, contrary to Stroke being worse or not significantly different than PD for these physiologic components in the present study. This apparent discrepancy is likely explained by 1) the separation of PD from PDwDem in the present study with no such separation in the prior study, which is relevant given that PDwDem was significantly more impaired than PD for these components, and 2) differences in the way these were physiologic components were measured since the prior study did not provide operational definitions for how physiologic impairments were judged. In addition, Garand et al., (2018) found that for Bolus Hold (C2) and Mastication (C3), a Pulmonary category had worse rates of impairment than the HNC category. Although this may appear inconsistent with our finding that COPD (the only pulmonary diagnosis in our study) was significantly less impaired than HNC

for both of these components, it is likely the Pulmonary category in the prior study contained non-COPD pulmonary diagnoses (e.g. Pneumonia) that would skew the results to much higher rates of impairment.

Clinical Significance of Present Study

Comparing the impairment rates in the present study to impairment rates in healthy adults by Garand et al. (2022) provides evidence that each of the five diagnoses in the present study show substantial and clinically significant impairment. That prior study showed that for 11 of the 17 MBSImP components, 0% of patients (N = 195) were impaired (i.e. all patients had scores of zero). In contrast, our results showed that regardless of diagnoses, for 15 of the 17 components at least 25% of patients were impaired, and for 8 of the 17 components at least 50% of patients were impaired (See Figure 5). Although the interpretation of this comparison is limited by the adults in that prior study being younger than those in the present study (mean age: 47 vs 73), the contrast still suggests that all diagnoses in the present study have substantial swallowing deficits.

Furthermore, the *differences* in impairment between diagnoses seen in the present study are also clinically significant. The main metric of comparison across diagnoses we use is the Normalized Mean OI Score, which accounts for differences between diagnoses at all levels of impairment. This metric, however, is novel, and thus an examination of the more familiar metric of impairment rates helps to make the effect size of the differences between diagnoses clearer. In particular, we can take two diagnoses and calculate the absolute difference in impairment rates (i.e. risk of impairment; Ranganathan et al., 2016) between those diagnoses. Take, for example, Laryngeal Vestibular Closure (C11), where 55% of PD patients are impaired and 73% of HNC patients are impaired. This means that there is an 18% absolute risk difference between PD and HNC for LVC, which suggests that, in equivalent sets of 10 patients, a clinician seeing HNC patients could expect to see ~2 more patients with LVC impairment than would a clinician seeing PD patients. This comparison serves to illustrate that the differences in impairment rates seen in the present study represent meaningful differences in clinical expectation across diagnoses. Although this impairment-rate comparison is useful for aiding interpretability and understanding clinical significance, these impairment rates do not account for the severity of impairment or differences in impairment distributions (like those seen in HNC), and therefore we used Normalized Mean OI Scores as the primary basis for our cross-diagnosis comparisons.

Commonalities, Differences, and Hypotheses

Using MBSImP and Normalized Mean OI Scores as the basis for comparisons in the present study provided us with a standardized measurement tool that allowed characterization of the commonalities and differences in impairment profiles across diagnoses accounting for differences in severity of impairment. Here we will highlight key findings of commonalities and differences across diagnoses and the hypotheses that were generated by these findings.

Across all 17 MBSImP components and PAS, there were 13 measures for which COPD and PD did not significantly differ. This apparent similarity in the impairment profiles of COPD and PD is interesting given that COPD and PD have differing etiologies (i.e. Pulmonary and Neurological, respectively). In addition, within the Pharyngeal and Esophageal Domains, Dementia did not significantly differ from COPD for 10 of 11 components nor from PD for 9 of 11 components. From these findings, we generated the hypothesis that the shared impairment profile of COPD and PD (and of Dementia in the pharyngeal domain) may represent a kind of baseline diagnosis-independent disruption to the swallowing mechanism. If this is the case, then this hypothesis could be tested by future studies examining the impairment profiles of other diagnoses to see if this baseline impairment profile appears in those as well. The impairment profiles of COPD and PD are, however, not identical. In the Oral domain the largest difference between COPD and PD is for Lingual motion (C4), with PD showing worse impairment than COPD. This result is likely due to the characteristic “tongue pumping” or repetitive rocking motion of the tongue associated with PD (Suttrup & Warnecke, 2016; Tjaden, 2008). In the Pharyngeal Domain, the largest difference is for Laryngeal Vestibular Closure (LVC - C11) where COPD and Dementia both have worse impairment than PD. The causes of an LVC-specific deficit are unclear in COPD, but are consistent with reports from Mancopes et al. (2021) that COPD may have particularly impaired laryngeal vestibular closure. In Dementia, it may be that a combination of poor Bolus Hold (i.e. posterior escape of the bolus) combined with delayed Pharyngeal Swallow Initiation may allow the bolus to enter the laryngeal vestibule, which is then not perfectly cleared leading to worse LVC scores. These hypotheses and questions could potentially be the focus of future research specifically on whether there are characteristic relationships between different physiologic components for particular diagnoses.

Between the Oral and Pharyngeal Domains, there was more variation in the relative position of diagnoses, i.e. in the "shape" of impairment profiles for the Oral domain than for the Pharyngeal domain. This can be seen in that the relative impairment of the diagnoses stays fairly constant across the Pharyngeal domain with the exception of a few components (like Pharyngeal Contraction and PES Opening). In contrast, in the Oral domain, the position of the diagnoses relative to one another changes much more frequently. For example, Dementia is typically more severe than HNC, but not for Oral Residue. This finding is consistent with the generally held hypothesis that the Oral phase of swallowing is more volitional, while the Pharyngeal is more hard-wired (Goyal & Mashimo, 2006; Panara et al., 2023). The present finding suggests that the more volitional nature of the Oral domain may result in more freedom for individual components of the swallow to be impaired independently. This result is also consistent with the finding from Chapter 1 of this dissertation that the Pharyngeal Domain of MBSImP has higher internal consistency than the Oral Domain (α_{oral} = 0.81; α_{phar} = 0.87 | Clain et al., 2022).

Within the Oral Domain, we found that Stroke, Dementia, and PDwDem consistently had worse impairment than PD and COPD. The worse impairment in PDwDem compared to PD alone potentially indicates that patients with PD may develop substantially worse Oral impairments once they start to develop Dementia. The worse impairment in Stroke and Dementia in comparison to PD and COPD may be related to the progressive nature of PD, in that patients may be referred before swallowing impairments have gotten severe, and the non-neurologic nature of COPD, in that COPD may not directly affect the voluntary control of the Oral Domain of swallowing as much as a neurologic diagnosis. Even though Stroke, Dementia, and PDwDem had worse Oral impairment than PD or COPD, there were still differences still among them for Lip Closure (C1), Lingual Motion (C2), and Oral Residue (C5). We therefore hypothesize that the more severe Oral impairments seen in these diagnoses are likely not due solely to some common cause, but that these impairments are likely due to a combination of diagnosis-independent and diagnosis-specific factors.

Beyond these three diagnoses, we would also like to highlight the unique behavior of the Oral Residue component (C5). One instance of this uniqueness is that Dementia does not significantly differ from PD in Oral Residue (C5) even though Dementia has significantly worse impairment than PD for all other Oral components. Additionally, HNC had among the highest levels of Oral Residue (not significantly

different from Stroke and PDwDem) even though it had lower impairment than those diagnoses for all other Oral components other than Mastication (C3). The high Mastication impairment in HNC was most likely driven by the MBSImP protocol requiring that a max impairment score be given if the bolus cannot be given due to safety concerns, and solid boluses (the only contributor to C3) often not being given to HNC patients post-operatively due to safety concerns. Consistent with this is the especially high proportion of max scores seen for Mastication for HNC (See Figure 5). Thus, the unique behavior of Oral Residue relative to other oral impairments is not easily explainable given the present analysis. This result could suggest that the degree to which Oral impairments produce Oral Residue may differ in differing diagnostic populations. Future studies could test this hypothesis by examining the relationships between physiologic components of the swallow and Oral Residue in different diagnostic populations.

In the Pharyngeal Domain, HNC consistently had the worst impairment among the diagnoses (though sometimes tied for worst). With this result, we would like to highlight that the impairment profiles seen for each diagnosis are likely averages of subpopulations that exist within each diagnosis. In the case of HNC, there is substantial heterogeneity in HNC treatments in terms of surgery location, resection size, (chemo)radiation dose, etc., and differences in these factors have often been shown to produce differences in the risk and severity of dysphagia (Christopherson et al., 2019; Giannitto et al., 2017; Jackson et al., 2010). Part of the reason that HNC patients had the worst severity in the Pharyngeal domain was that HNC patients had disproportionately higher rates of having the most severe impairments (discovered based on violation of the proportional odds assumption.) One possible explanation for these results is that there were individual HNC patients that had the most severe scores across all components. However, given the heterogeneity of the underlying population mentioned above, it is also possible that particular subpopulations had particularly severe impairments in the physiology most related to the nature and location of surgery and radiation, and radiation dose, which after averaging across populations resulted in high rates of the most severe impairment levels across pharyngeal domain. To test this hypothesis, future studies could extend the present study's approach to examining impairment profiles within the HNC population to test whether treatment-specific factors produce characteristic swallow-physiology impairment profiles. Research along these lines in other diagnoses has already found evidence of characteristic impairment profiles based on stroke lesion location (Steinhagen et al., 2009;

Wilmskoetter et al., 2019, 2020) and dementia subtypes (N. Rogus-Pulia et al., 2015; Suh et al., 2009). Our results emphasize the promise and importance of continuing to characterize population-specific impairment profiles.

Also in the Pharyngeal Domain, Stroke consistently had worse impairment than COPD, PD, and Dementia. One hypothesis to explain this difference is that the Stroke patients in the present dataset (for whom Stroke onset date was available; see Supplemental Table 2) are on average 5-years post-stroke, and so may represent more persistent and thus more severe issues than PD and COPD. One component where Stroke did not significantly differ from COPD and PD was PES Opening (C14). One possible reason not detecting a difference here may be that impairments in PES opening are associated with brainstem (specifically lateral medullary) strokes (Steinhagen et al., 2009) and subcortical stroke in general appear to be less common in the present dataset (see Supplemental Table 2).

In the Esophageal domain, similar reasoning may explain why Stroke was significantly less impaired than COPD and PD for Esophageal Clearance (C17), especially considering that PD and COPD are both associated with esophageal dysfunction (Gadel et al., 2012; J. S. Kim et al., 2015; Suttrup & Warnecke, 2016). It may be that cortical Stroke patients tend to have less Esophageal Impairment. This lower impairment of Stroke patients is likely not an artifact of the most severe Stroke patients not being switched to the AP view as 1) that bias due to missingness was corrected for with inverse probability weighting, and 2) Stroke had among the highest impairment for the other AP-view component, i.e. Pharyngeal Contraction (C13).

For penetration and aspiration, we found that Stroke and HNC did not significantly differ, and yet they significantly differed on their impairment levels for nearly every MBSImP Component. From this finding, we generated the hypothesis that there may be multiple different combinations of physiologic impairments that can generate risk of penetration and aspiration. Future studies could test this hypothesis by examining whether the strength of association between particular MBSImP components and PAS can differ across diagnoses. In addition, Dementia had worse PAS than COPD or PD, even though they had similar pharyngeal impairment profiles. It is thus possible that the worse PAS seen in Dementia is related to the additional effect of its Oral impairments. For instance, it could be that the same reason mentioned above for Dementia's worse LVC than PD, i.e. especially impaired Bolus Hold (C2) leading to posterior

escape of the bolus, combined with impaired Initiation of the Pharyngeal Swallow (C6) potentially leading to bolus entering the laryngeal vestibule, could also be a hypothesis to explain Dementia's higher risk of penetration and aspiration. Finally, we observed that PD and COPD had similar impairment levels for most MBSImP components and for PAS. From this result, we hypothesized that although it may be that multiple different impairment profiles can result in similar risk of penetration and aspiration (like for HNC and Stroke), it may also be true that a particular impairment profile may consistently result in a particular risk of penetration and aspiration.

AP View and Missingness

Pharyngeal Contraction (C13 – AP View) and Pharyngeal Stripping Wave (C12 – Lateral View) are both related to the action of the pharyngeal constrictor muscles. However, the results from the present study suggest that these components are not redundant. The key results here are that Stroke patients were less severe than HNC patients for Pharyngeal Stripping Wave (C12), but not for Pharyngeal Contraction (C13). Our hypothesis to explain this difference between C12 and C13 is 1) that there may be a higher prevalence of unilateral impairments in Stroke patients than HNC patients and 2) the AP view allows for the detection of these unilateral impairments (i.e. via unilateral bulging in C13), which are not detectable in the lateral view. If this hypothesis is true, this difference in the relative impairment of Stroke and HNC between Pharyngeal Stripping Wave (C12) and Pharyngeal Contraction (C13) would emphasize the importance of acquiring the AP View to fully characterize patient impairment. This finding is especially important given that it is often the more severe patients that do not have the AP view assessed (Figure 4). Therefore, without the AP view, clinicians may be missing a potential target for treatment.

Age-Associated effects

Age by diagnosis interactions were included in the present analysis because they were shown to improve model fit. As a consequence, separate age-associated odds ratios were obtained for each diagnosis for each component. However, "age" here must be interpreted in light of the idea that it is likely a proxy of and confounded with typical age-related changes in physiology, disease progression, and other factors. Given that much of the research on age-related changes in swallow physiology has been conducted in healthy populations (Jardine et al., 2020; Garand et al., 2020; Wang et al., 2015; Feng et al.,

2013; Hiramatsu et al., 2015; Wirth et al., 2016; Rofes et al., 2010; Mancopes, Gandhi, et al., 2021), the present results provide support to the importance of understanding the nature of age-related changes in diagnostic populations as well.

Sex- and Race-Associated Effects

Sex and race were primarily included in the present study as covariates to control for differences across diagnoses. However, the present analysis revealed sex- and race-associated differences in impairment. In particular, male patients had consistently more severe impairment than female patients, especially in the pharyngeal domain, and Black patients had more severe impairment than Asian patients in the Oral, but not Pharyngeal domains. These findings are consistent with studies that have found worse pharyngeal impairment in male patients (In PD: Dumican et al., 2023; In general dysphagic population: Kassem et al., 2022), and found worse oral but not pharyngeal impairment in Black patients with Stroke (Daniels et al., 2017). Considering that the causes of sex- and race-associated disparities are often complex and related to social determinants of health and health policy, the disparities seen in the present study underline the need for future studies to be conducted that more directly attempt to understand the causes of race- and sex-associated disparities in the physiology of dysphagia.

MBSImP^{lot}

The MBSImP^{lot} is a novel tool for visualizing patient impairments across the 17 MBSImP components and PAS. In the present study, it is being used to compare an individual's impairment profile to the SDR-based prediction for a patient with the same diagnosis and age. The differences seen between the impairment profile of a particular patient compared to their population-based expectation might help to identify the components where a patient’s swallow is particularly severe or healthy, which may help clinicians make inferences about and target treatment to the nature of that patient’s impairment. In the example presented here, a 79-year-old patient with Dementia is compared to the SDR-based prediction and has a normalized Penetration/Aspiration score of zero, which is low relative to their expected value despite many of their MBSImP scores being worse than expected. This context is valuable as it suggests that the patient may be effectively compensating for their swallowing other physiologic impairments. In this way clinicians could use MBSImP^{lot} and the SDR-based predictions to help contextualize and guide their understanding of their patients’ impairments.

Limitations

1. As mentioned in the discussion of clinical significance, no healthy-normal control group was included due to all patients in the MBSImP SDR being referred for MBSS and therefore all suspected of having dysphagia. Therefore it is not entirely clear the extent to which the diagnosis groups are more severe than controls, though as mentioned above, current evidence from Garand et al. (2022) suggests that most MBSImP components have very low impairment rates in healthy adult patients (0% for 11/17 components with N = 195).
2. All of the diagnoses included in the present study are heterogeneous groups composed of many subpopulations. The impairment profiles presented here are therefore averages across these subpopulations, which each likely have their own particular variations on the impairment profiles seen here.
3. For the MBSImP^{lot}, the SDR-based predictions currently are only based on the diagnosis and age of the patient. There are many factors beyond diagnosis and age that would likely affect patient impairment, so ideally, future studies would incorporate more variables into the model to allow better matching between the parameters of the predictions with the specific characteristics of individual patients. However, it is important to note that not all variables are equally valid to include in predictions. The predictions in the present were computed unconditional on race and sex because including those variables would create different impairment expectations which can lead to unequal treatment of individual patients on the basis of their race and sex (Paulus & Kent, 2017). Thus, future studies aimed at developing prediction models for physiologic impairment of swallowing should focus on including variables that can specifically be tied to physiology and disease processes.

Dissertation Discussion

Prior to this dissertation, the standardized MBSImP measurement method was developed and widely implemented. MBSImP was developed in response to heterogeneity in the existing measurement methods for swallowing physiology, a heterogeneity driven in large part by the inherent subjectivity in choosing what aspects of physiology to measure and how to operationalize them. This subjectivity stands in contrast to the objective and physical nature of swallowing physiology itself. MBSImP attempted to more closely align the subjective and objective sides of swallowing-physiology measurement through standardization of the subjective elements based on a foundation of rigorous theoretical and empirical work (Martin-Harris & Jones, 2008) refined through consensus of a group of experts and testing of the validity of the resulting tool (Martin-Harris et al., 2008). The success of this approach can be seen through the widespread uptake of MBSImP. The result of this widespread uptake was the development of a unique Swallowing Data Registry (SDR) that is the first large-scale dataset of clinical physiologic swallowing impairment data.

This dissertation uses the physiologic data contained in the SDR to further test and characterize the subjective and objective sides of MBSImP. On the subjective side, Chapter 1 showed that in this large-scale clinical dataset, the standardized physiologic components had excellent validity in terms of how well their correlation structure aligned with the underlying physiologic domains (though these domains only accounted for 50% of the total variance in scores). Chapter 1 also showed that the MBSImP standardized training adequately minimized the subjectivity of scoring across and within clinicians, i.e. it produced good inter- and intra-rater reliability. On the physiologic side, Chapter 2 revealed a mix of common and unique impairment profiles across the five diagnoses examined. The existence of common impairment profiles indicates that although dysphagia is physiologically complex and multi-etiological, it is possible for disparate etiologies to converge to similar impairment profiles. The existence of unique impairment profiles suggests that for some diagnoses, their etiologies can exert specific and unique effects on the swallowing mechanism.

Each of these chapters therefore has its own importance and contribution to the understanding and characterizing of swallowing impairment and its measurement. However, beyond their independent conclusions, both of these chapters appear to be pointing to a need for an integrated approach to

swallowing-physiology measurement that simultaneously considers overall domain severity, the behavior of individual components, and the effect of patient-level factors. The structural validity and internal consistency of the domains of MBSImP from Chapter 1 suggest that the severity of each of the domains could be represented by a latent "domain-level severity" variable. Furthermore, it is possible that such domain-level latent severity variables would be sufficient to capture the hypothesized "typical profile of impairments" suggested by the common impairment profiles seen across diagnoses in Chapter 2. At the same time, it appears that domain-level latent variables alone would not be entirely sufficient to describe impairment given that individual components and patient-level variables can have unique contributions to impairment. As mentioned above, in Chapter 1, we found that the Oral and Pharyngeal domain factors of MBSImP accounted for about 50% of the variance in patients scores, leaving 50% of the variance to be explained by variation in individual components and patient-level variables. In addition, in Chapter 2 we found unique behavior of the components within diagnoses as with Lingual Motion (C4) for PD, Oral Residue (C5) for Dementia and HNC, and Laryngeal Vestibular Closure (C11) for COPD and Dementia. Thus, from these results we can see there is need for an analysis that can simultaneously address overall severity of domains, the uniqueness of individual components, and patient-level factors.

Looking forward, I propose that the optimal method for adequately addressing all of these levels of analysis will be to use extended item response theory (EMEIRT; Chalmers, 2015). An item response theory approach would allow each physiologic component to be situated on a valid domain-severity scale and would allow a domain-severity score to be derived from any patient's component scores. The extended item response theory approach would then allow for the incorporation of component and patient-level factors to explain potential uniquenesses of impairment within particular patients and populations. The benefits of this proposed approach can be seen in how it would enhance the proposed MBSImP^{lot} ©.

Using EMEIRT would allow the MBSImP^{lot} to be integrated with the item-person or Wright Map visualization (Torres Irribarra & Freund, 2014). In the current MBSImP^{lot}, the scores of each component are normalized to their max score and distributed equally between 0 and 1. Using the EMEIRT approach would allow for the scores of each component to be distributed according to how indicative that component and particular score are for the overall severity of impairment. Each patient's overall severity

level (for each domain) could be given a percentile value in the overall distribution of patients or for a particular diagnostic population. A patient's particular impairment profile could still be visualized (as is done in the current MBSImP^{lot}) with the added benefit that any components that are particularly impaired or unimpaired could be identified based on their impairment level relative to the patient's overall severity. This extension of the MBSImP^{lot} demonstrates that the EMEIRT approach to analysis and visualization would allow a simultaneous examination of overall severity (as normed to a population of interest), how each physiologic component relates to that overall severity, and whether there are uniquenesses in the impairments of particular patients or populations.

The findings of this dissertation therefore set the stage for the development of a standardized and validated method of analysis for the already standardized and validated method of measurement of MBSImP. This dissertation would not have been possible without the dedicated and visionary work that led to the development and implementation of MBSImP. This dissertation shows the accomplishments and power of MBSImP, and also points a way to the path forward. It has been an honor to have the opportunity to do this work and it is humbly appreciated.

References

Adkins, C., Takakura, W., Spiegel, B. M. R., Lu, M., Vera-Llonch, M., Williams, J., & Almario, C. V. (2020). Prevalence and Characteristics of Dysphagia Based on a Population-Based Survey. *Clinical Gastroenterology and Hepatology*, *18*(9), 1970-1979.e2. https://doi.org/10.1016/j.cgh.2019.10.029

Ahn, D. H., Yang, H. E., Kang, H. J., Do, K. H., Han, S. C., Jang, S. W., & Lee, J. H. (2020). Changes in etiology and severity of dysphagia with aging. *European Geriatric Medicine*, *11*(1), 139–145. https://doi.org/10.1007/s41999-019-00259-0

Altman, K. W., Yu, G.-P., & Schaefer, S. D. (2010). Consequence of Dysphagia in the Hospitalized Patient: Impact on Prognosis and Hospital Resources. *Archives of Otolaryngology–Head & Neck Surgery*, *136*(8), 784–789. https://doi.org/10.1001/archoto.2010.129

Barbon, C. E. A., Chepeha, D. B., Hope, A. J., Peladeau, -Pigeon Melanie, Waito, A. A., & Steele, C. M. (2020). Mechanisms of Impaired Swallowing on Thin Liquids Following Radiation Treatment for Oropharyngeal Cancer. *Journal of Speech, Language, and Hearing Research*, *63*(9), 2870–2879. https://doi.org/10.1044/2020_JSLHR-19-00220

Bender, R., & Grouven, U. (1997). Ordinal Logistic Regression in Medical Research. *Journal of the Royal College of Physicians of London*, *31*(5), 546–551.

Bingjie, L., Tong, Z., Xinting, S., Jianmin, X., & Guijun, J. (2010). Quantitative videofluoroscopic analysis of penetration-aspiration in post-stroke patients. *Neurology India*, *58*(1), 42. https://doi.org/10.4103/0028-3886.60395

Bonilha, H. S., Simpson, A. N., Ellis, C., Mauldin, P., Martin-Harris, B., & Simpson, K. (2014). The one-year attributable cost of post-stroke dysphagia. *Dysphagia*, *29*(5), 545–552. https://doi.org/10.1007/s00455-014-9543-8

Casey, J. A., Schwartz, B. S., Stewart, W. F., & Adler, N. E. (2016). Using Electronic Health Records for Population Health Research: A Review of Methods and Applications. *Annual Review of Public Health*, *37*(1), 61–81. https://doi.org/10.1146/annurev-publhealth-032315-021353

Cerchiello, P., Dequarti, E., Giudici, P., & Magni, C. (2010). Scorecard models to evaluate perceived quality of academic teaching. *Statistica & Applicazioni*, *2*, 145–156.

Chalmers, R. P. (2015). Extended Mixed-Effects Item Response Models With the MH-RM Algorithm: Extended Mixed-Effects Item Response Models. *Journal of Educational Measurement*, *52*(2), 200–222. https://doi.org/10.1111/jedm.12072

Christopherson, K. M., Ghosh, A., Mohamed, A. S. R., Kamal, M., Gunn, G. B., Dale, T., Kalpathy-Cramer, J., Messer, J., Garden, A. S., Elhalawani, H., Frank, S. J., Lewin, J., Morrison, W. H., Phan, J., Gross, N., Ferrarotto, R., Weber, R. S., Rosenthal, D. I., Lai, S. Y., … Fuller, D. (2019). Chronic radiation-associated dysphagia in oropharyngeal cancer survivors: Towards age-adjusted dose constraints for deglutitive muscles. *Clinical and Translational Radiation Oncology*, *18*, 16–22. https://doi.org/10.1016/j.ctro.2019.06.005

Clain, A. E., Alkhuwaiter, M., Davidson, K., & Martin-Harris, B. (2022). Structural Validity, Internal Consistency, and Rater Reliability of the Modified Barium Swallow Impairment Profile: Breaking Ground on a 52,726-Patient, Clinical Data Set. *Journal of Speech,*

Language, and Hearing Research: JSLHR, *65*(5), 1659–1670. https://doi.org/10.1044/2022_JSLHR-21-00554

Clark, H. M., Stierwalt, J. A. G., Tosakulwong, N., Botha, H., Ali, F., Whitwell, J. L., & Josephs, K. A. (2020). Dysphagia in Progressive Supranuclear Palsy. *Dysphagia*, *35*(4), 667–676. https://doi.org/10.1007/s00455-019-10073-2

Cowie, M. R., Blomster, J. I., Curtis, L. H., Duclaux, S., Ford, I., Fritz, F., Goldman, S., Janmohamed, S., Kreuzer, J., Leenay, M., Michel, A., Ong, S., Pell, J. P., Southworth, M. R., Stough, W. G., Thoenes, M., Zannad, F., & Zalewski, A. (2017). Electronic health records to facilitate clinical research. *Clinical Research in Cardiology*, *106*(1), 1–9. https://doi.org/10.1007/s00392-016-1025-6

Curtis, J. A., Molfenter, S. M., & Troche, M. S. (2020). Pharyngeal Area Changes in Parkinson's Disease and Its Effect on Swallowing Safety, Efficiency, and Kinematics. *Dysphagia*, *35*(2), 389–398. https://doi.org/10.1007/s00455-019-10052-7

Daniels, S. K., Pathak, S., Mukhi, S. V., Stach, C. B., Morgan, R. O., & Anderson, J. A. (2017a). The Relationship Between Lesion Localization and Dysphagia in Acute Stroke. *Dysphagia*, *32*(6), 777–784. https://doi.org/10.1007/s00455-017-9824-0

Daniels, S. K., Pathak, S., Mukhi, S. V., Stach, C. B., Morgan, R. O., & Anderson, J. A. (2017b). The Relationship Between Lesion Localization and Dysphagia in Acute Stroke. *Dysphagia*, *32*(6), 777–784. https://doi.org/10.1007/s00455-017-9824-0

de Deus Chaves, R., Chiarion Sassi, F., Davison Mangilli, L., Jayanthi, S. K., Cukier, A., Zilberstein, B., & Furquim de Andrade, C. R. (2014a). Swallowing transit times and valleculae residue in stable chronic obstructive pulmonary disease. *BMC Pulmonary Medicine*, *14*(1), 62. https://doi.org/10.1186/1471-2466-14-62

de Deus Chaves, R., Chiarion Sassi, F., Davison Mangilli, L., Jayanthi, S. K., Cukier, A., Zilberstein, B., & Furquim de Andrade, C. R. (2014b). Swallowing transit times and valleculae residue in stable chronic obstructive pulmonary disease. *BMC Pulm Med*, *14*, 62. https://doi.org/10.1186/1471-2466-14-62

DiCiccio, T. J., & Efron, B. (1996). Bootstrap confidence intervals. *Statistical Science*, *11*(3), 189–228. https://doi.org/10.1214/ss/1032280214

Dozier, T. S., Brodsky, M. B., Michel, Y., Walters Jr, B. C., & Martin-Harris, B. (2006). Coordination of Swallowing and Respiration in Normal Sequential Cup Swallows. *The Laryngoscope*, *116*(8), 1489–1493. https://doi.org/10.1097/01.mlg.0000227724.61801.b4

Dumican, M., & Watts, C. (2022). Swallow Safety and Laryngeal Kinematics: A Comparison of Dysphagia Between Parkinson's Disease and Cerebrovascular Accident. *Journal of Parkinson's Disease*. https://doi.org/10.3233/JPD-222372

Dumican, M., Watts, C., Drulia, T., & Zhang, Y. (2023). Dysphagia Presentation, Airway Invasion, and Gender Differences in a Clinically Based Sample of People with Parkinson's Disease. *Dysphagia*, *38*(1), 353–366. https://doi.org/10.1007/s00455-022-10472-y

Efron, B., & Hastie, T. (2016). *Computer Age Statistical Inference*. Cambridge University Press.

Eglseer, D., Halfens, R. J. G., Schols, J. M. G. A., & Lohrmann, C. (2018). Dysphagia in Hospitalized Older Patients: Associated Factors and Nutritional Interventions. *The Journal of Nutrition, Health & Aging*, *22*(1), 103–110. https://doi.org/10.1007/s12603-017-0928-x

Fattori, B., Nacci, A., Farneti, D., Ceravolo, R., Santoro, A., Bastiani, L., Simoni, F., Pagani, R., & De Bortoli, N. (2022). Dysphagia in Parkinson's disease: Pharyngeal manometry and

fiberoptic endoscopic evaluation. *Auris Nasus Larynx*. https://doi.org/10.1016/j.anl.2022.03.016

Feng, X., Todd, T., Lintzenich, C. R., Ding, J., Carr, J. J., Ge, Y., Browne, J. D., Kritchevsky, S. B., & Butler, S. G. (2013). Aging-Related Geniohyoid Muscle Atrophy Is Related to Aspiration Status in Healthy Older Adults. *The Journals of Gerontology: Series A*, *68*(7), 853–860. https://doi.org/10.1093/gerona/gls225

Gadel, A. A., Mostafa, M., Younis, A., & Haleem, M. (2012). Esophageal motility pattern and gastro-esophageal reflux in chronic obstructive pulmonary disease. *Hepato-Gastroenterology*, *59*(120), 2498–2502. https://doi.org/10.5754/hge10433

Gall, S. L., Donnan, G., Dewey, H. M., Macdonell, R., Sturm, J., Gilligan, A., Srikanth, V., & Thrift, A. G. (2010). Sex differences in presentation, severity, and management of stroke in a population-based study. *Neurology*, *74*(12), 975–981. https://doi.org/10.1212/WNL.0b013e3181d5a48f

Garand, K. L., Beall, J., Hill, E. G., Davidson, K., Blair, J., Pearson, W., & Martin-Harris, B. (2022). Effects of Presbyphagia on Oropharyngeal Swallowing Observed during Modified Barium Swallow Studies. *The Journal of Nutrition, Health & Aging*, *26*(11), 973–980. https://doi.org/10.1007/s12603-022-1854-0

Garand, K. L. (Focht), Armeson, K. E., Hill, E. G., & Martin-Harris, B. (2018a). Identification of Phenotypic Patterns of Dysphagia: A Proof of Concept Study. *American Journal of Speech-Language Pathology*, *27*(3), 988–995. https://doi.org/10.1044/2018_AJSLP-17-0173

Garand, K. L. (Focht), Armeson, K. E., Hill, E. G., & Martin-Harris, B. (2018b). Identification of Phenotypic Patterns of Dysphagia: A Proof of Concept Study. *American Journal of*

Speech-Language Pathology, *27*(3), 988–995. https://doi.org/10.1044/2018_AJSLP-17-0173

Garand, K. L. (Focht), Culp, L., Wang, B., Davidson, K., & Martin-Harris, B. (2020). Aging Effects on Esophageal Transit Time in the Upright Position During Videofluoroscopy. *Annals of Otology, Rhinology & Laryngology*, *129*(6), 618–624. https://doi.org/10.1177/0003489420903332

Garand, K. L., Strange, C., Paoletti, L., Hopkins-Rossabi, T., & Martin-Harris, B. (2018). Oropharyngeal swallow physiology and swallowing-related quality of life in underweight patients with concomitant advanced chronic obstructive pulmonary disease. *International Journal of Chronic Obstructive Pulmonary Disease*, *Volume 13*, 2663–2671. https://doi.org/10.2147/COPD.S165657

Giannitto, C., Preda, L., Zurlo, V., Funicelli, L., Ansarin, M., Di Pietro, S., & Bellomi, M. (2017). Swallowing Disorders after Oral Cavity and Pharyngolaryngeal Surgery and Role of Imaging. *Gastroenterology Research and Practice*, *2017*, e7592034. https://doi.org/10.1155/2017/7592034

Gonz, ález-F. M., Sein, M. T., & Palmer, J. B. (2011). Clinical Experience Using the Mann Assessment of Swallowing Ability for Identification of Patients at Risk for Aspiration in a Mixed-Disease Population. *American Journal of Speech-Language Pathology*, *20*(4), 331–336. https://doi.org/10.1044/1058-0360(2011/10-0082)

Goyal, R. K., & Mashimo, H. (2006). Physiology of oral, pharyngeal, and esophageal motility. *GI Motility Online*. https://doi.org/10.1038/gimo1

Guyomard, V., Fulcher, R. A., Redmayne, O., Metcalf, A. K., Potter, J. F., & Myint, P. K. (2009). Effect of Dysphasia and Dysphagia on Inpatient Mortality and Hospital Length of

Stay: A Database Study. *Journal of the American Geriatrics Society*, *57*(11), 2101–2106. https://doi.org/10.1111/j.1532-5415.2009.02526.x

Harrell, F. (2015). *Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis*. Springer International Publishing. https://doi.org/10.1007/978-3-319-19425-7

Harrell Jr., F. E. (2023). *rms: Regression Modeling Strategies* (6.6-0). https://cran.r-project.org/web/packages/rms/index.html

Hiramatsu, T., Kataoka, H., Osaki, M., & Hagino, H. (2015). Effect of aging on oral and swallowing function after meal consumption. *Clinical Interventions in Aging*, *10*, 229–235. https://doi.org/10.2147/CIA.S75211

Hiss, S. G., Strauss, M., Treole, K., Stuart, A., & Boutilier, S. (2004). Effects of Age, Gender, Bolus Volume, Bolus Viscosity, and Gustation on Swallowing Apnea Onset Relative to Lingual Bolus Propulsion Onset in Normal Adults. *Journal of Speech, Language, and Hearing Research*, *47*(3), 572–583. https://doi.org/10.1044/1092-4388(2004/044)

Horner, J., Buoyer, F. G., Alberts, M. J., & Helms, M. J. (1991). Dysphagia Following Brain-Stem Stroke: Clinical Correlates and Outcome. *Archives of Neurology*, *48*(11), 1170–1173. https://doi.org/10.1001/archneur.1991.00530230078026

Hutcheson, K. A., Lewin, J. S., Barringer, D. A., Lisec, A., Gunn, G. B., Moore, M. W. S., & Holsinger, F. C. (2012). Late dysphagia after radiotherapy-based treatment of head and neck cancer. *Cancer*, *118*(23), 5793–5799. https://doi.org/10.1002/cncr.27631

Im, I., Jun, J.-P., Crary, M. A., Carnaby, G. D., & Hong, K. H. (2019). Longitudinal Kinematic Evaluation of Pharyngeal Swallowing Impairment in Thyroidectomy Patients. *Dysphagia*, *34*(2), 161–169. https://doi.org/10.1007/s00455-018-9949-9

Jackson, A., Caria, N., Setton, J., Romanyshyn, J., McNamara, S. A., Wolden, S. L., & Lee, N. (2010). Dose Response of Dysphagia in Oropharynx Cancer Patients. *International Journal of Radiation Oncology, Biology, Physics*, *78*(3), S151. https://doi.org/10.1016/j.ijrobp.2010.07.374

Jardine, M., Miles, A., & Allen, J. (2020). A Systematic Review of Physiological Changes in Swallowing in the Oldest Old. *Dysphagia*, *35*(3), 509–532. https://doi.org/10.1007/s00455-019-10056-3

Kassem, F., Masalha, M., Biadsee, A., Nageris, B., Kagan, R., & Nachmani, A. (2022). Analysis of Sex-related Differences in Patients with Dysphagia: Using a Videofluoroscopy. *The Israel Medical Association Journal: IMAJ*, *24*(7), 464–469.

Kawashima, K., Motohashi, Y., & Fujishima, I. (2004). Prevalence of dysphagia among community-dwelling elderly individuals as estimated using a questionnaire for dysphagia screening. *Dysphagia*, *19*(4), 266–271. https://doi.org/10.1007/BF02638594

Kidambi, T., Toto, E., Ho, N., Taft, T., & Hirano, I. (2012). Temporal trends in the relative prevalence of dysphagia etiologies from 1999-2009. *World Journal of Gastroenterology : WJG*, *18*(32), 4335–4341. https://doi.org/10.3748/wjg.v18.i32.4335

Kim, I. S., & Han, T. R. (2005). Influence of Mastication and Salivation on Swallowing in Stroke Patients. *Archives of Physical Medicine and Rehabilitation*, *86*(10), 1986–1990. https://doi.org/10.1016/j.apmr.2005.05.004

Kim, J. S., Youn, J., Suh, M. K., Kim, T.-E., Chin, J., Park, S., & Cho, J. W. (2015). Cognitive and Motor Aspects of Parkinson's Disease Associated with Dysphagia. *Canadian Journal of Neurological Sciences*, *42*(6), 395–400. https://doi.org/10.1017/cjn.2015.304

Kim, Y. H., Han, T. R., Nam, H. S., Seo, H. G., & Oh, B.-M. (2019). Temporal characteristics of laryngeal penetration and aspiration in stroke patients. *NeuroRehabilitation*, *44*(2), 231–238. https://doi.org/10.3233/NRE-182569

Kim, Y., & McCullough, G. H. (2010). Maximal hyoid excursion in poststroke patients. *Dysphagia*, *25*(1), 20–25. https://doi.org/10.1007/s00455-009-9224-1

Kooi, A. H.-J., Boo, J. P.-L., Ng, S. Y.-E., Acharyya, S., Goh, K.-H., Tay, K.-Y., Au, W.-L., & Tan, L. C.-S. (2019). The Modified Barium Swallow Impairment Profile as a Predictor of Clinical Outcomes of Admission for Pneumonia or Choking in Dysphagic Patients with Parkinson's Disease. *Dysphagia*, *34*(6), 896–903. https://doi.org/10.1007/s00455-019-09986-9

Kruse, C. S., Stein, A., Thomas, H., & Kaur, H. (2018). The use of Electronic Health Records to Support Population Health: A Systematic Review of the Literature. *Journal of Medical Systems*, *42*(11), 214. https://doi.org/10.1007/s10916-018-1075-6

Langmore, S. E., & Krisciunas, G. P. (2010). Dysphagia After Radiotherapy for Head and Neck Cancer: Etiology, Clinical Presentation, and Efficacy of Current Treatments. *Perspectives on Swallowing and Swallowing Disorders (Dysphagia)*, *19*(2), 32. https://doi.org/10.1044/sasd19.2.32

Langmore, S. E., Olney, R. K., Lomen-Hoerth, C., & Miller, B. L. (2007). Dysphagia in Patients With Frontotemporal Lobar Dementia. *Archives of Neurology*, *64*(1), 58–62. https://doi.org/10.1001/archneur.64.1.58

Lee, C. H., & Yoon, H.-J. (2017). Medical big data: Promise and challenges. *Kidney Research and Clinical Practice*, *36*(1), 3–11. https://doi.org/10.23876/j.krcp.2017.36.1.3

Lee, S.-Y., Kim, B. H., & Park, Y. H. (2015). Analysis of Dysphagia Patterns Using a Modified Barium Swallowing Test Following Treatment of Head and Neck Cancer. *Yonsei Medical Journal*, *56*(5), 1221. https://doi.org/10.3349/ymj.2015.56.5.1221

Lin, T., & Shune, S. (2020). Chronic Obstructive Pulmonary Disease and Dysphagia: A Synergistic Review. *Geriatrics*, *5*(3), Article 3. https://doi.org/10.3390/geriatrics5030045

Lo, W.-L., Leu, H.-B., Yang, M.-C., Wang, D.-H., & Hsu, M.-L. (2019). Dysphagia and risk of aspiration pneumonia: A nonrandomized, pair-matched cohort study. *Journal of Dental Sciences*, *14*(3), 241–247. https://doi.org/10.1016/j.jds.2019.01.005

Logemann, J. A. (1988). Swallowing physiology and pathophysiology. *Otolaryngologic Clinics of North America*, *21*(4), 613–623.

Mack, C., Su, Z., & Westreich, D. (2018a). *Managing Missing Data in Patient Registries: Addendum to Registries for Evaluating Patient Outcomes: A User's Guide, Third Editio*. https://www.ncbi.nlm.nih.gov/books/NBK493611

Mack, C., Su, Z., & Westreich, D. (2018b). Types of Missing Data. In *Managing Missing Data in Patient Registries: Addendum to Registries for Evaluating Patient Outcomes: A User's Guide, Third Edition [Internet]*. Agency for Healthcare Research and Quality (US). https://www.ncbi.nlm.nih.gov/books/NBK493614/

Mancopes, R., Borowsky, da R. F., Tomasi, L. L., Pasqualoto, A. S., & Steele, C. M. (2021). Chronic Obstructive Pulmonary Disease and Dysphagia: What Have We Learned So Far and What Do We Still Need to Investigate? *Perspectives of the ASHA Special Interest Groups*, *6*(5), 1212–1221. https://doi.org/10.1044/2021_PERSP-20-00288

Mancopes, R., Gandhi, P., Smaoui, S., & Steele, C. M. (2021). Which Physiological Swallowing Parameters Change with Healthy Aging? *OBM Geriatrics*, *5*(1). https://doi.org/10.21926/obm.geriatr.2101153

Mancopes, R., Peladeau, -Pigeon Melanie, Barrett, E., Guran, A., Smaoui, S., Pasqualoto, A. S., & Steele, C. M. (2020). Quantitative Videofluoroscopic Analysis of Swallowing Physiology and Function in Individuals With Chronic Obstructive Pulmonary Disease. *Journal of Speech, Language, and Hearing Research*, *63*(11), 3643–3658. https://doi.org/10.1044/2020_JSLHR-20-00154

Martin-Harris, B., Brodsky, M. B., Michel, Y., Castell, D. O., Schleicher, M., Sandidge, J., Maxwell, R., & Blair, J. (2008). MBS Measurement Tool for Swallow Impairment—MBSImp: Establishing a Standard. *Dysphagia*, *23*(4), 392–405. https://doi.org/10.1007/s00455-008-9185-9

Martin-Harris, B., & Jones, B. (2008). The Videofluorographic Swallowing Study. *Physical Medicine and Rehabilitation Clinics of North America*, *19*(4), 769–785. https://doi.org/10.1016/j.pmr.2008.06.004

Mehraban-Far, S., Alrassi, J., Patel, R., Ahmad, V., Browne, N., Lam, W., Jiang, Y., Barber, N., & Mortensen, M. (2021). Dysphagia in the elderly population: A Videofluoroscopic study. *American Journal of Otolaryngology*, *42*(2), 102854. https://doi.org/10.1016/j.amjoto.2020.102854

Miarons, M., Clavé, P., Wijngaard, R., Ortega, O., Arreola, V., Nascimento, W., & Rofes, L. (2018). Pathophysiology of Oropharyngeal Dysphagia Assessed by Videofluoroscopy in Patients with Dementia Taking Antipsychotics. *Journal of the American Medical*

Directors Association, *19*(9), 812.e1-812.e10. https://doi.org/10.1016/j.jamda.2018.04.016

Miller, A. J. (1986). Neurophysiological basis of swallowing. *Dysphagia*, *1*(2), 91–100. https://doi.org/10.1007/BF02407121

Minagi, Y., Ono, T., Hori, K., Fujiwara, S., Tokuda, Y., Murakami, K., Maeda, Y., Sakoda, S., Yokoe, M., Mihara, M., & Mochizuki, H. (2018). Relationships between dysphagia and tongue pressure during swallowing in Parkinson's disease patients. *Journal of Oral Rehabilitation*, *45*(6), 459–466. https://doi.org/10.1111/joor.12626

Mogensen, U. B., Olsen, T. S., Andersen, K. K., & Gerds, T. A. (2013). Cause-specific Mortality after Stroke: Relation to Age, Sex, Stroke Severity, and Risk Factors in a 10-Year Follow-up Study. *Journal of Stroke and Cerebrovascular Diseases*, *22*(7), e59–e65. https://doi.org/10.1016/j.jstrokecerebrovasdis.2012.04.006

Mokhlesi, B., Logemann, J. A., Rademaker, A. W., Stangl, C. A., & Corbridge, T. C. (2002). Oropharyngeal Deglutition in Stable COPD. *Chest*, *121*(2), 361–369. https://doi.org/10.1378/chest.121.2.361

Namasivayam-MacDonald, A. M., Alomari, N., Attner, L., Benjamin, R. D., Chill, A., Doka, S., Guastella, R., Marchese, J., Oppedisano, S., Ressa, K., Rider, B. E., Sandoval, G. K., Soyfer, A., Thompson, R., Walshe, C. M., & Riquelme, L. F. (2021). A Retrospective Analysis of Swallowing Function and Physiology in Patients Living with Dementia. *Dysphagia*. https://doi.org/10.1007/s00455-021-10350-z

Namasivayam-MacDonald, A. M., & Riquelme, L. F. (2019). Quantifying Airway Invasion and Pharyngeal Residue in Patients with Dementia. *Geriatrics (Basel, Switzerland)*, *4*(1), E13. https://doi.org/10.3390/geriatrics4010013

Panara, K., Ramezanpour Ahangar, E., & Padalia, D. (2023). Physiology, Swallowing. In *StatPearls*. StatPearls Publishing. http://www.ncbi.nlm.nih.gov/books/NBK541071/

Park, T., Kim, Y., Ko, D.-H., & McCullough, G. (2010). Initiation and Duration of Laryngeal Closure During the Pharyngeal Swallow in Post-Stroke Patients. *Dysphagia*, *25*(3), 177–182. https://doi.org/10.1007/s00455-009-9237-9

Patel, B., Legacy, J., Hegland, K. W., Okun, M. S., & Herndon, N. E. (2020a). A comprehensive review of the diagnosis and treatment of Parkinson's disease dysphagia and aspiration. *Expert Review of Gastroenterology & Hepatology*, *14*(6), 411–424. https://doi.org/10.1080/17474124.2020.1769475

Patel, B., Legacy, J., Hegland, K. W., Okun, M. S., & Herndon, N. E. (2020b). A comprehensive review of the diagnosis and treatment of Parkinson's disease dysphagia and aspiration. *Expert Review of Gastroenterology & Hepatology*, *14*(6), 411–424. https://doi.org/10.1080/17474124.2020.1769475

Patel, D. A., Krishnaswami, S., Steger, E., Conover, E., Vaezi, M. F., Ciucci, M. R., & Francis, D. O. (2018). Economic and survival burden of dysphagia among inpatients in the United States. *Diseases of the Esophagus*, *31*(1). https://doi.org/10.1093/dote/dox131

Paulus, J. K., & Kent, D. M. (2017). Race and ethnicity – a part of the equation for personalized clinical decision making? *Circulation. Cardiovascular Quality and Outcomes*, *10*(7), e003823. https://doi.org/10.1161/CIRCOUTCOMES.117.003823

R Core Team. (2022). *R: A language and environment for statistical computing.* (4.2.0). R Foundation for Statistical Computing. https://www.r-project.org/

Ranganathan, P., Pramesh, C. S., & Aggarwal, R. (2016). Common pitfalls in statistical analysis: Absolute risk reduction, relative risk reduction, and number needed to treat. *Perspectives in Clinical Research*, *7*(1), 51–53. https://doi.org/10.4103/2229-3485.173773

Rofes, L., Arreola, V., Romea, M., Palomera, E., Almirall, J., Cabré, M., Serra-prat, M., & Clavé, P. (2010). Pathophysiology of oropharyngeal dysphagia in the frail elderly. *Neurogastroenterology & Motility*, *22*(8), 851-e230. https://doi.org/10.1111/j.1365-2982.2010.01521.x

Rogus-Pulia, N. M., Larson, C., Mittal, B. B., Pierce, M., Zecker, S., Kennelty, K., Kind, A., & Connor, N. P. (2016). Effects of Change in Tongue Pressure and Salivary Flow Rate on Swallow Efficiency Following Chemoradiation Treatment for Head and Neck Cancer. *Dysphagia*, *31*(5), 687–696. https://doi.org/10.1007/s00455-016-9733-7

Rogus-Pulia, N. M., Pierce, M. C., Mittal, B. B., Zecker, S. G., & Logemann, J. A. (2014). Changes in Swallowing Physiology and Patient Perception of Swallowing Function Following Chemoradiation for Head and Neck Cancer. *Dysphagia*, *29*(2), 223–233. https://doi.org/10.1007/s00455-013-9500-y

Rogus-Pulia, N., Malandraki, G. A., Johnson, S., & Robbins, J. (2015). Understanding Dysphagia in Dementia: The Present and the Future. *Current Physical Medicine and Rehabilitation Reports*, *3*(1), 86–97. https://doi.org/10.1007/s40141-015-0078-1

Seaman, S. R., & White, I. R. (2013). Review of inverse probability weighting for dealing with missing data. *Statistical Methods in Medical Research*, *22*(3), 278–295. https://doi.org/10.1177/0962280210395740

Seo, H. G., Oh, B.-M., & Han, T. R. (2016). Swallowing Kinematics and Factors Associated with Laryngeal Penetration and Aspiration in Stroke Survivors with Dysphagia. *Dysphagia*, *31*(2), 160–168. https://doi.org/10.1007/s00455-015-9670-x

Speyer, R., Cordier, R., Bouix, C., Gallois, Y., & Woisard, V. (2021). Using Classical Test Theory to Determine the Psychometric Properties of the Deglutition Handicap Index. *Dysphagia*. https://doi.org/10.1007/s00455-021-10250-2

Steinhagen, V., Grossmann, A., Benecke, R., & Walter, U. (2009). Swallowing Disturbance Pattern Relates to Brain Lesion Location in Acute Stroke Patients. *Stroke*, *40*(5), 1903–1906. https://doi.org/10.1161/STROKEAHA.108.535468

Suh, M. K., Kim, H., & Na, D. L. (2009). Dysphagia in Patients With Dementia: Alzheimer Versus Vascular. *Alzheimer Disease & Associated Disorders*, *23*(2), 178–184. https://doi.org/10.1097/WAD.0b013e318192a539

Suttrup, I., & Warnecke, T. (2016). Dysphagia in Parkinson's Disease. *Dysphagia*, *31*(1), 24–32. https://doi.org/10.1007/s00455-015-9671-9

Swan, K., Cordier, R., Brown, T., & Speyer, R. (2019). Psychometric Properties of Visuoperceptual Measures of Videofluoroscopic and Fibre-Endoscopic Evaluations of Swallowing: A Systematic Review. *Dysphagia*, *34*(1), 2–33. https://doi.org/10.1007/s00455-018-9918-3

Tjaden, K. (2008). Speech and Swallowing in Parkinson's Disease. *Topics in Geriatric Rehabilitation*, *24*(2), 115–126. https://doi.org/10.1097/01.TGR.0000318899.87690.44

Torres Irribarra, D., & Freund, R. (2014). *Wright Map: IRT item-person map with ConQuest integration*. Recuperado de http://github. com/david-ti/wrightmap.

Vose, A. K., Marcus, A., & Humbert, I. (2019). Kinematic Visual Biofeedback Improves Accuracy of Swallowing Maneuver Training and Accuracy of Clinician Cues During Training in Stroke Patients with Dysphagia. *PM&R*, *11*(11), 1159–1169. https://doi.org/10.1002/pmrj.12093

Wall, L. R., Ward, E. C., Cartmill, B., & Hill, A. J. (2013). Physiological changes to the swallowing mechanism following (chemo)radiotherapy for head and neck cancer: A systematic review. *Dysphagia*, *28*(4), 481–493. https://doi.org/10.1007/s00455-013-9491-8

Walton, J., & Silva, P. (2018). Physiology of swallowing. *Surgery (Oxford)*, *36*(10), 529–534. https://doi.org/10.1016/j.mpsur.2018.08.010

Wang, C.-M., Chen, J.-Y., Chuang, C.-C., Tseng, W.-C., Wong, A. M., & Pei, Y.-C. (2015). Aging-related changes in swallowing, and in the coordination of swallowing and respiration determined by novel non-invasive measurement techniques. *Geriatrics & Gerontology International*, *15*(6), 736–744. https://doi.org/10.1111/ggi.12343

Wilmskoetter, J., Bonilha, L., Martin-Harris, B., Elm, J. J., Horn, J., & Bonilha, H. S. (2019). Mapping acute lesion locations to physiological swallow impairments after stroke. *NeuroImage: Clinical*, *22*, 101685. https://doi.org/10.1016/j.nicl.2019.101685

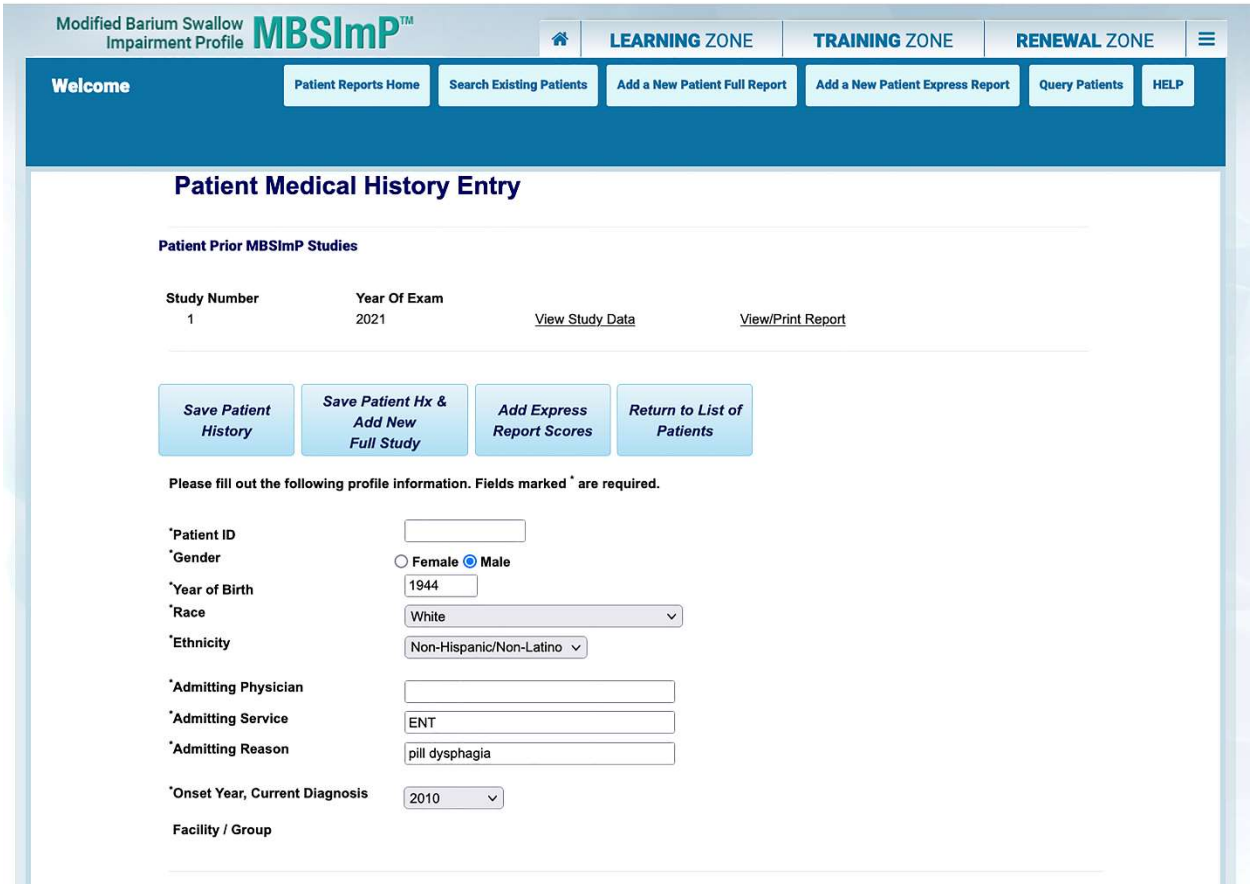
Wilmskoetter, J., Daniels, S. K., & Miller, A. J. (2020). Cortical and Subcortical Control of Swallowing—Can We Use Information From Lesion Locations to Improve Diagnosis and Treatment for Patients With Stroke? *American Journal of Speech-Language Pathology*, *29*(2 Suppl), 1030–1043. https://doi.org/10.1044/2019_AJSLP-19-00068

Wirth, R., Dziewas, R., Beck, A. M., Clavé, P., Hamdy, S., Heppner, H. J., Langmore, S., Leischker, A., Martino, R., Pluschinski, P., Rösler, A., Shaker, R., Warnecke, T., Sieber,

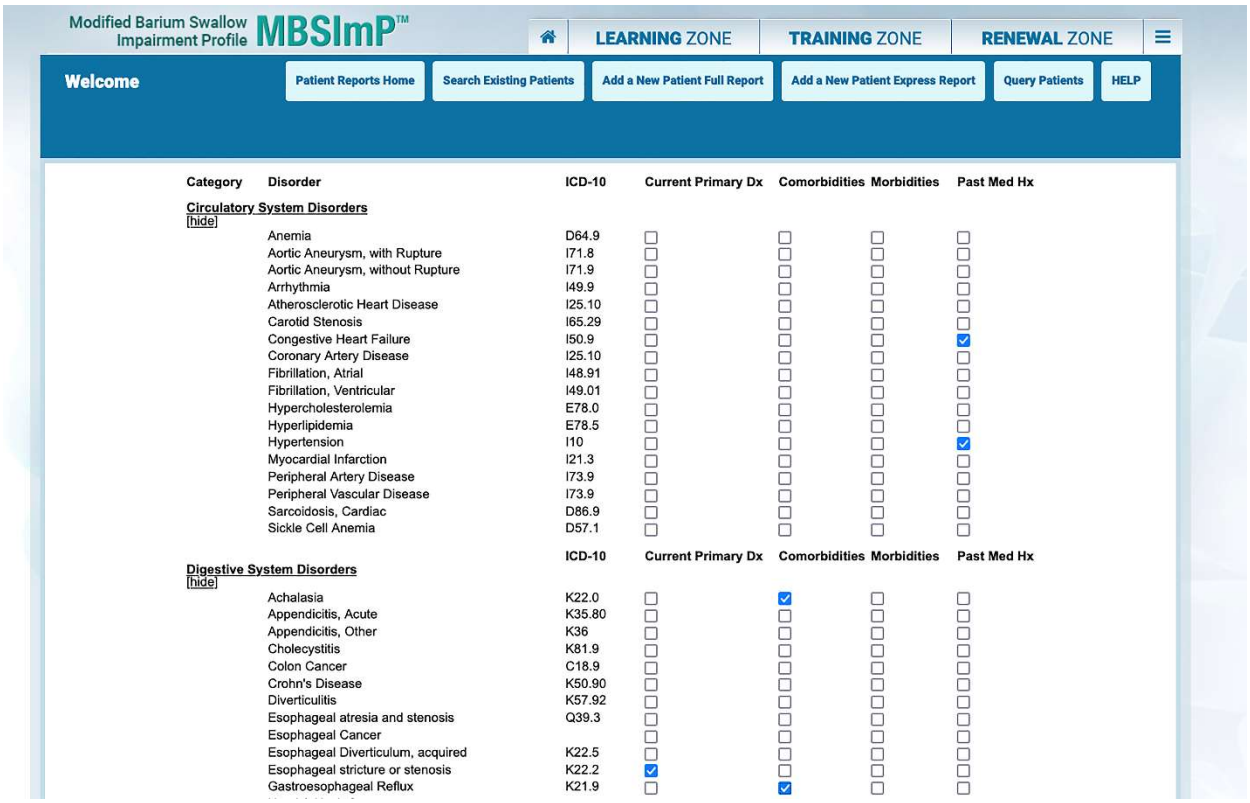
C. C., & Volkert, D. (2016). Oropharyngeal dysphagia in older persons and from pathophysiology to adequate intervention: A review and summary of an international expert meeting. *Clinical Interventions in Aging*, *11*, 189–208. https://doi.org/10.2147/CIA.S97481

Yee, T., & Moler, C. (2023). *VGAM: Vector Generalized Linear and Additive Models* (1.1-8). https://cran.r-project.org/web/packages/VGAM/index.html

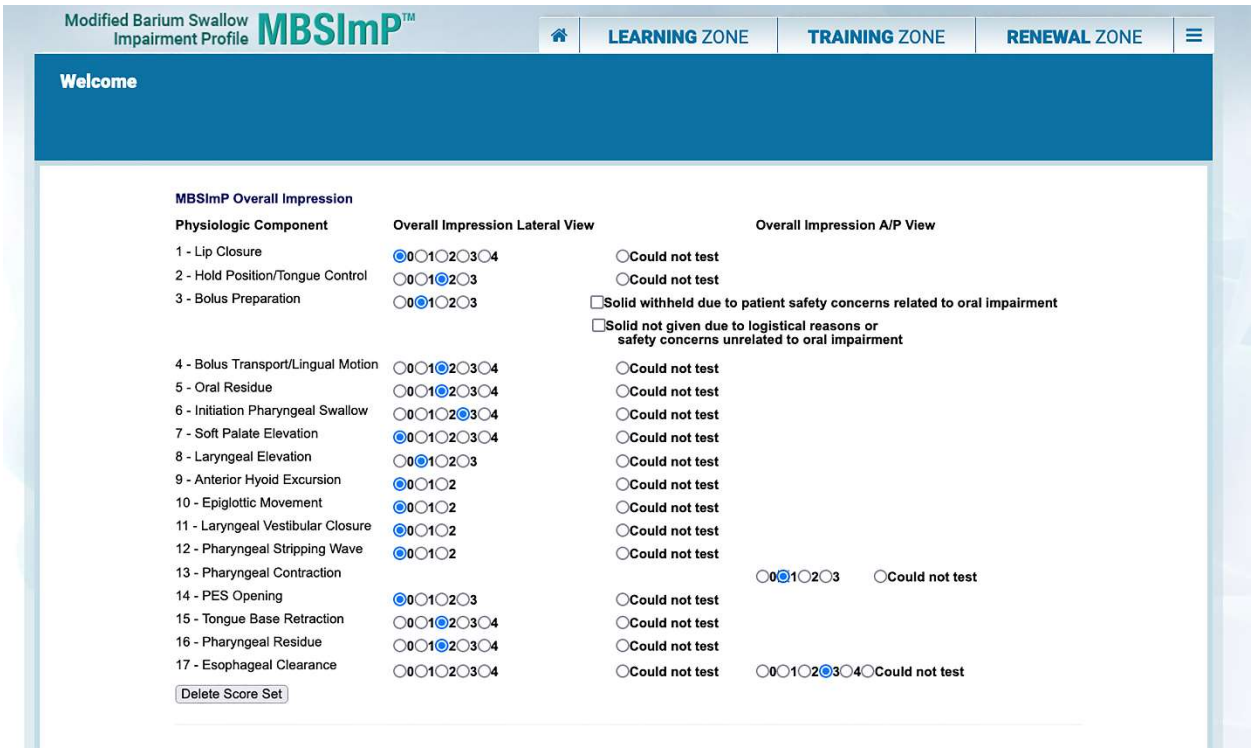
Appendix 1



Supplemental Figure 1 MBSImP SDR patient information entry form 1.



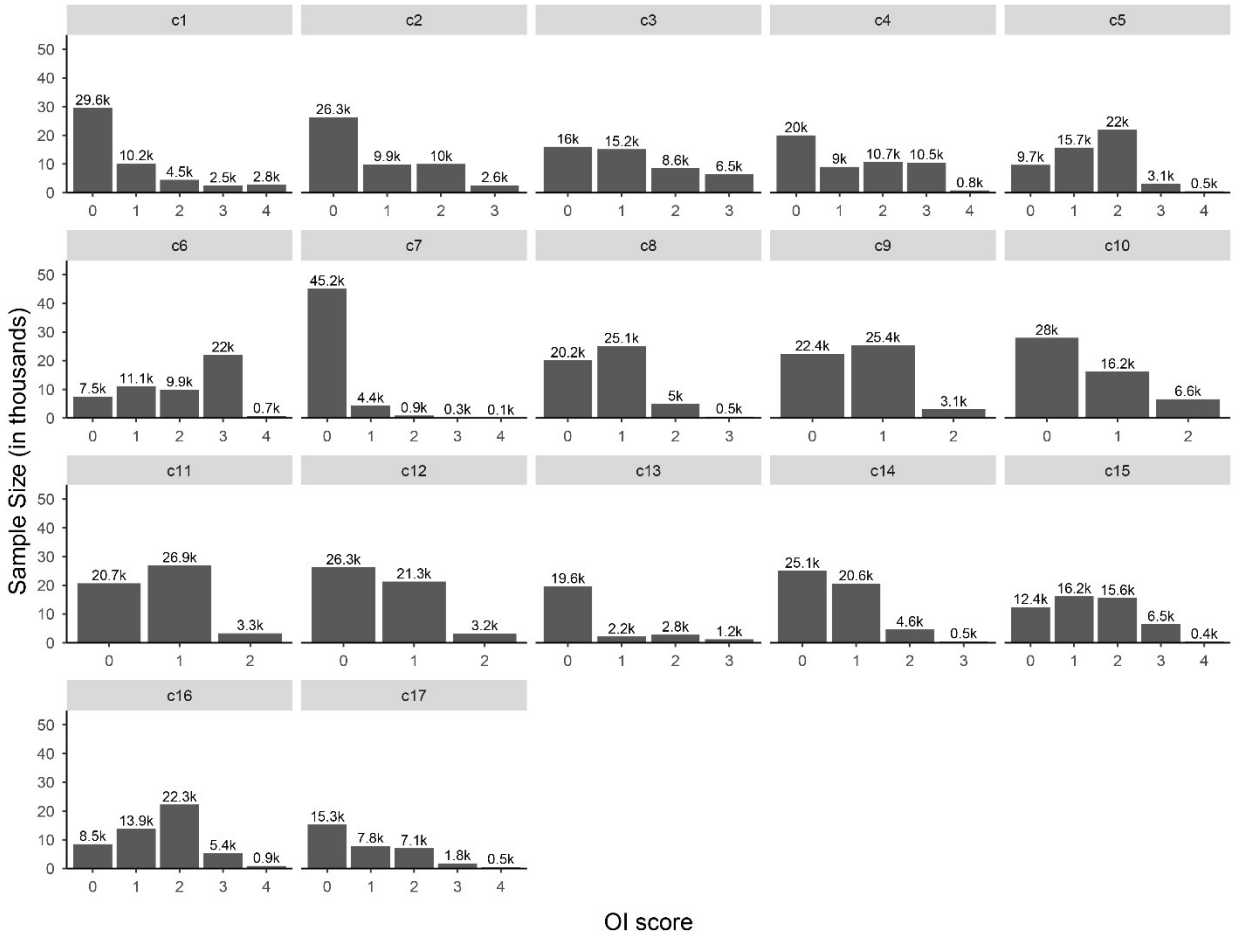
Supplemental Figure 2 MBSImP SDR patient information entry form 2.



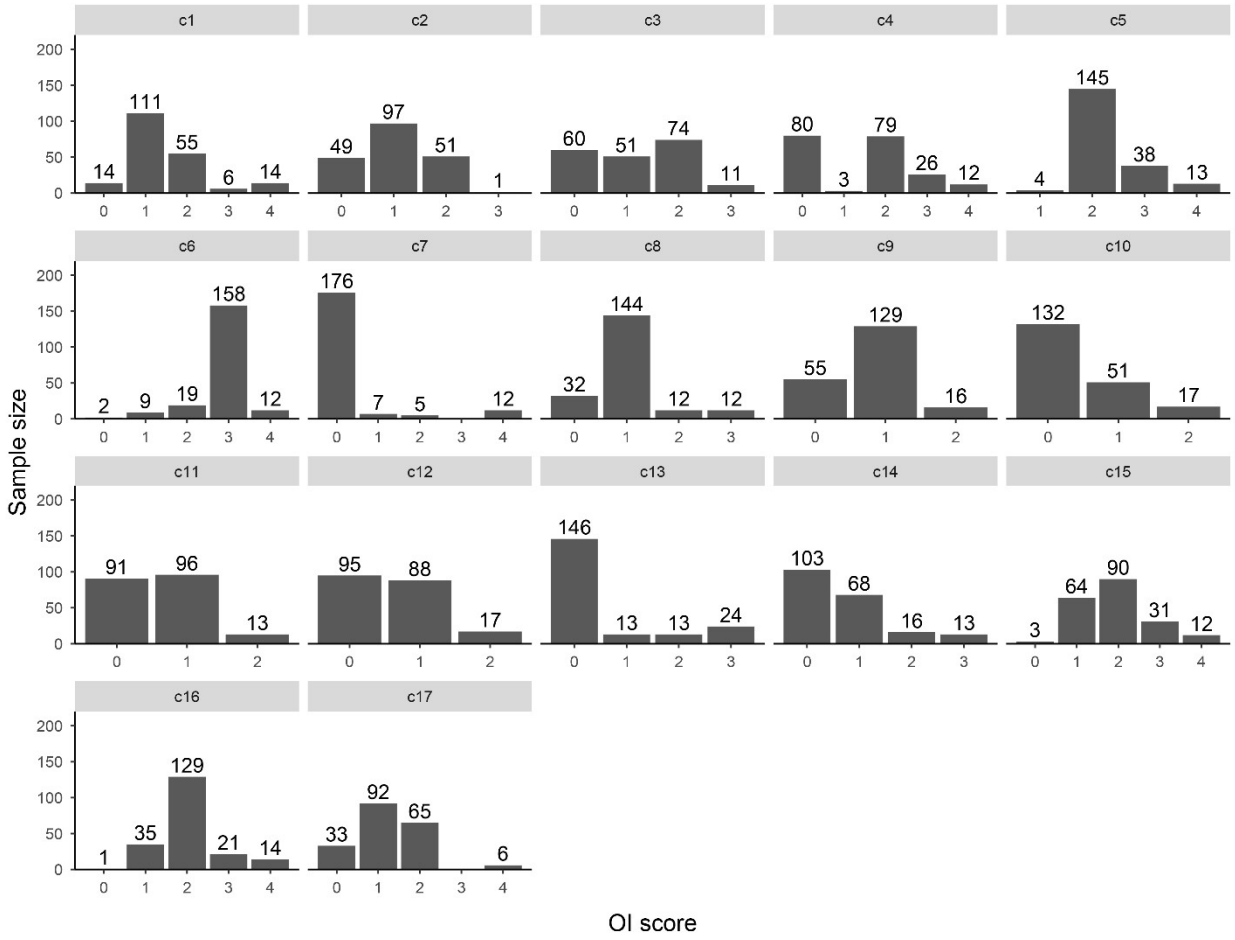
Supplemental Figure 3 MBSImP SDR patient information entry form 3.

Rank	Diagnosis	N
1	Gastroesophageal Reflux	4236
2	Hypertension	3977
3	Cerebrovascular Accident	3747
4	Chronic Obstructive Pulmonary Disease	2072
5	Head & Neck Cancer	1981
6	Hyperlipidemia	1543
7	Pneumonia-Acquired	1217
8	Coronary Artery Disease	1197
9	Respiratory Failure	1188
10	Fibrillation, Atrial	1010
11	Diabetes, Insulin Dependent	995
12	Congestive Heart Failure	984
13	Dementia, general	842
14	Diabetes, Non-Insulin Dependent	815
15	Parkinson's Disease, without Dementia	712
16	Shortness Of Breath	676
17	Cough	660
18	Pneumonia-Aspiration, Right, Lower Lobe	554
19	Anemia	497
20	Pneumonia-Aspiration, Left, Lower Lobe	467
21	Epilepsy/Seizure Disorder	368
22	Lung Cancer	364
23	Traumatic Brain Injury	348
24	Parkinson's Disease, with Dementia	334
25	Sepsis	334
26	Encephalopathy	308
27	Dementia, other	283
28	Alzheimer's Disease	277
29	Other specified disorders of thyroid	271
30	Multiple Sclerosis	262
NA	Not Reported	31,380

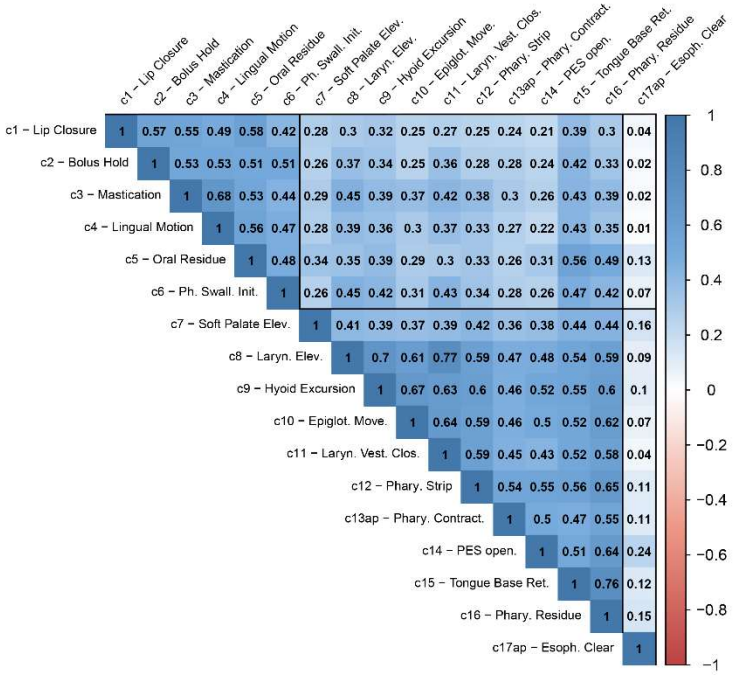
Supplemental Table 1 The 30 most common diagnoses in the MBSImP SDR and their respective sample sizes.



Supplemental Figure 4 Distribution of MBSImP Scores in the 52,726-patient sample of the SDR.



Supplemental Figure 5 Distribution of MBSImP Scores in the 50-patient sample.



Supplemental Figure 6 Correlation matrix of all MBSImP components for the 52,726-patient sample of the SDR.

Appendix 2

	Stroke Patients (N=3342)
Laterality	
Bilateral	54 (1.6%)
Left	389 (11.6%)
Right	375 (11.2%)
Missing	2524 (75.5%)
Subcortical	
No	209 (6.3%)
Yes	70 (2.1%)
Unknown	1525 (45.6%)
Missing	1538 (46.0%)
Years Post-Stroke (Years)	
Mean (SD)	5.02 (2.49)
Median [Min, Max]	5.00 [0, 26.0]
Missing	2380 (71.2%)

Supplemental Table 2 Stroke-specific characteristics.

HNC (N=2399)		Tumor Properties		Treatments	
Stage T		**Pathology**		**Surgery**	
0	5 (0.2%)	Other	52 (2.2%)	No	1105 (46.1%)
1	35 (1.5%)	Squamous Cell	303 (12.6%)	Yes	246 (10.3%)
2	78 (3.3%)	Missing	2044 (85.2%)	Missing	1048 (43.7%)
3	60 (2.5%)				
4	107 (4.5%)			**Radiation Therapy**	
X	2 (0.1%)	**Recurrent**		No	736 (30.7%)
Missing	2112 (88.0%)	No	1258 (52.4%)	Yes	615 (25.6%)
Stage N		Yes	93 (3.9%)	Missing	1048 (43.7%)
0	83 (3.5%)	Missing	1048 (43.7%)		
1	41 (1.7%)				
2	119 (5.0%)			**Chemotherapy**	
3	10 (0.4%)	**Metastatic**		No	961 (40.1%)
X	5 (0.2%)	Yes	1351 (56.3%)	Yes	390 (16.3%)
Missing	2141 (89.2%)	Missing	1048 (43.7%)	Missing	1048 (43.7%)

Supplemental Table 3 HNC-specific tumor properties and treatments.

Dementia Subtypes	**(N = 1066)**
General Dementia	478
PDwDem	267
Alzheimer's Disease	193
Lewy Body	31
Other	146

Supplemental Table 4 Dementia Subtypes.