

12-10-2013

Nursing Administration of the 3-Ounce Water Swallow Screen

Heather L. Warner Dr.
Heather.Lisitano@yahoo.com

Follow this and additional works at: <https://opencommons.uconn.edu/dissertations>

Recommended Citation

Warner, Heather L. Dr., "Nursing Administration of the 3-Ounce Water Swallow Screen" (2013). *Doctoral Dissertations*. 271.
<https://opencommons.uconn.edu/dissertations/271>

Nursing Administration of the 3-Ounce Water Swallow Screen

Heather Lisitano Warner, Ph.D., CCC/SLP

University of Connecticut, 2013

Abstract

Introduction

Accurate and timely evaluation of swallowing is necessary to determine how to safely administer medications, maintain adequate nutrition and hydration, and avoid deleterious sequelae of prandial aspiration pneumonia. Use of a validated and reliable screening tool for determination of aspiration risk is a critical component of dysphagia management. The 3-ounce water swallow challenge (Suiter & Leder, 2008) is a validated and reliable screening tool that is well supported in the literature.

Statement of the Problem

While use of the 3-ounce water swallow challenge (Suiter & Leder 2008) administered by speech-language pathologists (SLPs) is supported by current research, who else should administer and interpret the challenge is not addressed. Therefore, health care professionals other than SLPs, i.e., registered nurses, should be involved in screening for aspiration risk (Bours et al., 2008). Deficiencies in current nurse administered screens are a barrier to this practice change.

Background

There is a paucity of literature supporting nursing administration of validated screening tools for determining aspiration risk in hospitalized patients. Current practice involving administration of swallow screens by nurses is comprised of investigations that utilize largely non-evidenced based variables.

Research Purpose

This study investigated accuracy of a registered nurse administered 3-ounce water swallow challenge with hospitalized patients deemed at-risk for prandial aspiration compared with blinded ratings from speech-language pathology.

Methods

Patients were administered the 3-ounce water swallow challenge protocol by a SLP. The nurse then administered the screen to the same patient within 1 hour and independently recorded results and diet recommendations. Simultaneous with the nurse administered screen, a SLP re-rated the patient's 3-ounce challenge for comparison with initial results as well as determined accuracy of the nurse administered screen.

Results

Intra- and inter-rater agreements for the two speech-language pathologists were 100% (Cohen's kappa of 1.0). Inter-rater agreement between registered nurses and speech-language pathologists was 98.01% (Cohen's kappa of 0.95).

Conclusion

Results confirm the reliability and accuracy of a registered nurse administered swallow screen. The finding of 98% agreement (Cohen's kappa 0.95) combined with previously reported 96.5% sensitivity, 97.9% negative predictive value, and $\leq 2\%$ false negative rate support adoption of the 3-ounce challenge for the general hospital population.

**Nursing Administration of the 3-ounce Water Swallow
Screening Protocol**

Heather Lisitano Warner

B.A. University of Connecticut (1998)

M.A. University of Connecticut (2000)

A Dissertation

Submitted in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

at the

University of Connecticut

2013

Copyright by
Heather Lisitano Warner

2013

APPROVAL PAGE

Doctor of Philosophy Dissertation

Nursing Administration of the 3-ounce Water Swallow Screen

Presented by

Heather Lisitano Warner, M.A.

Major Advisor _____

Carl Coelho, Ph.D.

Associate Advisor _____

Rachel Theodore, Ph.D.

Associate Advisor _____

Steven Leder, Ph.D.

University of Connecticut

2013

Acknowledgements

I would like to express gratitude the members of my committee for facilitating my navigation of this educational journey. To Carl Coelho, who, instilled both a passion for the study of speech pathology and understanding of clinical excellence that inspired me to be the clinician I am today. Who, in the now distant past, spent much time assisting me with the intellectually challenging transition from research in the hard sciences to clinical research and who spent equally as much time encouraging me to return to my doctoral work. I thank him also for always recognizing the fact that life exists outside of academia and for remembering what it is like to go through this program with a spouse and children.

I would like to thank Rachel Theodore for her willingness to sit on my committee. Though our academic relationship has been brief, her thoughtful input and thorough review of my submissions have been much appreciated. I only regret that I hadn't made her acquaintance sooner.

I must express my deep gratitude to Steve Leder for all he has done, past and present, to contribute to my growth as a clinician researcher. First and foremost, for demonstrating day in and day out what it means to be a successful clinician researcher, specifically, how to incorporate research into clinical work - his performance epitomizes clinical research in such a way that it makes it difficult to determine where one ends and the other begins. I thank him for fostering my ability to read critically, years before I returned to school – providing me with training that

was invaluable to my growth as a discerning consumer of evidence based publications, a doctoral student, and a clinical researcher. I would also like to thank him for giving me the unique opportunity to be involved in so many aspects of clinical research prior to my acceptance into this program. Most recently, I would like to express my sincere gratitude for his willingness to join my committee. His unwavering support, encouragement, and availability for assist and consultation was vital to my success in this final requirement of my doctoral program. I will always be grateful for all he has done to shape my identity as a clinical researcher.

There are others, too numerous to single out, who have walked with me on this journey, who have supported my need for flexibility, emotional support, and clinical assistance whose efforts, friendship, and support have not gone unnoticed and were equally as critical to my success in this program. I thank my parents, who have always believed in me and who have helped me to believe I can accomplish anything I put my mind to. To all of you, I express my sincere gratitude.

Finally and perhaps most importantly I would like thank my husband John Warner for his patience and support during this journey. While academically and professionally satisfying, this process can only be described as trying with respect to home and family. Without his support both emotionally and logistically, I would not have been able to accomplish what I have.

A sincere thank you to all who have contributed to this amazing journey.

Table of Contents

Chapter I: Introduction

Introduction.....	page 1
Statement of Problem.....	page 4
Purpose of Study.....	page 5

Chapter II: Review of Literature

Purpose of Screening.....	page 6
Terminology of Swallow Screening.....	page 8
Evidence Based Screening.....	page 11
Definition of Effective Screen.....	page 14
Critical Caveat: Silent Aspiration.....	page 17
Swallowing Screen Stimuli.....	page 19
Individual Efforts Associated with Swallow.....	page 20
Individual Efforts Not Associated with Swallow.....	page 21
Water Swallow Tests.....	page 22
Bolus Swallow Tests.....	page 46
Other Swallow Screens.....	page 57
Nursing Administration of Swallow Screening.....	page 63
Conclusion.....	page 67

Chapter III: Methods and Procedures

Participants.....	page 69
Procedures.....	page 69
Data Analysis.....	page 71

Chapter IV: Results

Results.....page 73

Chapter V: Discussion

Discussion.....page 76

Limitations.....page 79

Conclusions.....page 81

Appendices.....page 82

Tables.....page 86

References.....page 93

CHAPTER I

Introduction

In the acute care population, evaluation and management of oropharyngeal dysphagia and aspiration risk is of critical importance. Complications from dysphagia have a direct impact on patient health, management, and cost of care. Accurate and timely evaluation of swallowing to determine aspiration risk is necessary to determine how to safely administer medications, maintain adequate nutrition and hydration for healing, and avoid deleterious sequelae of dysphagia, specifically prandial aspiration pneumonia. In addition to primary concerns for the health of the patient, complications associated with dysphagia, i.e., malnutrition, dehydration, and pneumonia, have a direct and negative financial impact on the healthcare system at large. Increased costs are directly associated with increased length of stay resulting from delays in oral intake, need for tube feeding, and complications from pneumonia (Altman et al., 2010).

The criterion (gold) standards for objective dysphagia evaluation are Videofluoroscopic Swallow Study (VFSS) and Fiberoptic Endoscopic Evaluation of Swallowing (FEES). It is not feasible, however, to objectively evaluate every patient that is at-risk for aspiration. Although these tests have become more widely available, not all facilities have access to equipment and staffing needed to carry them out. Objective testing, while often necessary, is both time consuming and costly and as a result, it is not temporally or financially prudent to expect that each patient with suspected dysphagia undergo an objective assessment. Therefore, non-instrumental swallowing screening for potential aspiration risk is a critical component of dysphagia management.

While speech-language pathologists perform the objective dysphagia evaluation (VFSS or FEES), other members of the health-care team, specifically registered nurses, licensed independent practitioners, and physicians, are responsible for identifying patients who are at-risk for potential aspiration. From both identification and staffing perspectives, it is beneficial to have an effective screening mechanism in place to determine those patients that need objective testing and those that do not. In support of the use of swallowing screening from an institutional perspective, it has been shown that hospitals with a formal screening protocol in place have both a higher adherence rate to swallowing screening (78% as compared to 57% at sites with no formal screen) and decreased pneumonia rates (2.4% vs 5.4% at sites with no formal screen) (Hinchey et al., 2005).

The justification for both use and need of a screening tool has become more widely recognized and accepted over the past decade. While there is no published practice standard for screening for aspiration risk, the literature supports the use of the 3-ounce water swallow challenge protocol (Suiter & Leder, 2008).

The 3-ounce water swallow challenge requires the patient to demonstrate uninterrupted drinking of 3-ounces of water without overt signs of aspiration. Criteria for passing is drinking 3-ounces of water completely and without stopping and without overt signs or symptoms of aspiration. Criteria for failure include inability to drink the entire amount, stopping and starting, or coughing or choking during or immediately after completion. The 3-ounce water swallow challenge protocol not only has one of the highest sensitivities reported (97%), it is also validated against two criterion standards (FEES) (Suiter & Leder, 2008).

and VFSS (Suiter et al., In Press), and incorporates a large and heterogeneous population sample (n=3,000) that far exceeds the breadth of subjects used by other studies (Suiter & Leder, 2008). While there are other screening tools with reasonable sensitivities and adequate methodology, when consideration is given to the strength of the research design, the time needed to administer, and ease of administration, this swallow screen is superior to other screens presented in the literature.

Additionally, there is strong and novel evidence that if a patient passes the 3-ounce screen, the clinician can confidently recommend an oral diet without need for further objective testing (Suiter & Leder, 2008; Leder et al., 2012). These data support a significant clinical change in the way clinicians have utilized screening tools in the past in that rather than using the screen solely to determine the need for additional testing, the screen can be utilized clinically to make diet recommendations. Taken collectively, the body of literature published about the use of the 3-ounce screen, the clinical considerations that must be taken into account, and the ability to recommend oral alimentation across heterogeneous diagnoses and at critical levels of patient care, demonstrate a programmatic line of research that is unmatched by other screening tools (Leder et al., 2012).

Statement of the Problem

There is a foundation of published literature that supports the involvement of health care professionals other than speech-language pathologists in the swallow screening process. Unfortunately, much of this literature is based on poorly supported screens, non-validated test stimuli, inadequate statistics, and questionable methodology. There are numerous clinical advantages to enlisting the assistance of a diverse range of health care practitioners. In most institutions, speech-language pathology is a consult service and clinicians do not have independent direct access to each and every patient who could potentially benefit from a swallow screen. The usual procedure is to rely on some form of screening process to determine the need for a referral. Unfortunately, these screening processes vary widely, are not standardized even within the same institution, most are non-evidenced based, and, therefore, unsubstantiated by the literature (McCullough et al., 1999). Registered nurses are a logical first choice to provide assistance in implementing the screening process to determine aspiration risk as they are the professionals that have the most direct, frequent, and continuous contact with patients.

There have been no data published, to date, that combine these two critical aspects of screening for aspiration risk, that is, nursing administration and interpretation of the 3-ounce water swallow screen. Research in this area could potentially alter the way screening for aspiration risk is conducted, how recommendations for appropriate oral diets are made, and, perhaps, move the field toward a much needed best practice model for swallow screening.

Purpose of the Study

This study has two aims: 1. to investigate the ability of registered nurses to administer and interpret the 3-ounce water screen; and 2. to describe the necessary training required for nurses to administer this screening tool. Specifically, the goal of the study is to answer the research question: Can registered nurses who have completed a web-based training module administer and interpret the 3-ounce water swallow screen reliably when compared to speech-language pathologists? A favorable outcome would permit this research to support both a successful training algorithm for nurses regarding instructions on how to administer and interpret the 3-ounce challenge that can be used by nursing administration for widespread implementation as well as the nurse's ability to utilize this screen in clinical practice.

CHAPTER II

Review of the Literature

Purpose of Screening

Dysphagia is a symptom of varied diagnoses. However, due to the nature of the disorder, much of the literature on dysphagia screening is focused upon stroke. The prevalence of dysphagia in stroke patients ranges from 30% to 67% (Edmiaston, et al., 2009). The reason for such variance in reported prevalence is multifactorial. Reported estimates vary based on the operational definition of dysphagia being utilized, the method of swallowing assessment, variability in the time post-onset that the evaluation was performed, as well as the type of stroke and number of subjects investigated (Mann et al., 2000).

Dysphagia can ultimately result in aspiration which has been reported to occur in 20 to 25% of stroke patients (Edmiaston et al., 2009). Daniels and colleagues (1997), however, reported estimates of 30% to 70% of stroke patients aspirating when VFSS determined aspiration status. There is a strong correlation between aspiration and the development of aspiration pneumonia (Daniels et al., 1997; Edmiaston et al., 2009). It has been shown that pneumonia is 7.5 times more likely to develop in stroke patients who aspirate than in those who do not (Holas et al., 1994). Hinchey and colleagues (2005) reported that of the deaths that occur after a stroke, approximately 35% are caused by pneumonia. As is demonstrated by these statistics, the occurrence of dysphagia is a serious medical issue that requires a thoughtful evaluative process. Despite variability in percentages of patients impacted, it is evident that the sequelae associated with dysphagia have a significant impact on both the health of the patient and the cost associated

with their care. Complications associated with dysphagia and potential aspiration pneumonia can increase length of stay, delay rehabilitation efforts, increase risk of mortality, and therefore, should be avoided whenever possible (Smithard et al., 1998).

From an institutional perspective, it has been shown that hospitals with formal swallowing screening protocols in place have a significantly decreased risk of pneumonia and higher compliance rate for conducting the swallowing screen (Hinchey et al., 2005). Hinchey et al. (2005) conducted a study that included 15 acute care institutions and investigated adherence rates for screening, the type of screen used, and the rate of development of in-hospital pneumonia in the stroke population. While investigators did not control for type of screen used, they found that of those sites with some type of formal screen in place, adherence rates for the screening process itself were 78% as compared to 57% at sites with no formal screen. The pneumonia rate at sites with a formal screen was 2.4% vs 5.4% at sites with no formal screen. This study supports the use of a formal screening protocol to minimize the risk of developing nosocomial acquired pneumonia.

Terminology of Swallow Screening

Further complicating review of this body of literature are the semantics and definitions of terminology surrounding dysphagia evaluation and screening. In an effort to decrease confusion and facilitate comparison of studies and procedures across the literature, operational definitions for common dysphagia terminology will be defined here.

A swallowing screen is defined by the American Speech Language and Hearing Association (ASHA) as "...a pass/fail procedure to identify individuals who require a comprehensive assessment of swallowing function or a referral for other professional and/or medical services" (ASHA, 2004, p. 3-10). Logemann et al. (1999) agrees in her report that a screening procedure is used to identify the presence or absence of a symptom, in this case, aspiration. If the procedure is being used to identify the abnormal anatomy and physiology causing the swallowing problem, this is considered a diagnostic tool.

Suiter and Leder (2008) further defined guidelines for screening in clinical practice. They recognize the two aforementioned criteria for a screening test: to determine the likelihood of aspiration and to determine the need for further evaluation. They proceeded to add a third clinical criterion: that the screening tool identify when it is safe to recommend oral intake.

For the purposes of this review, the term screening will be used to describe a tool which is relatively quick to administer, that seeks to determine presence or absence of a risk factor or symptom i.e. aspiration risk, has a goal of passing or failing, and is able to make recommendations for the need for further objective testing or recommendations for an oral diet as appropriate.

The term swallowing evaluation will be reserved for procedures which define the anatomy of the swallow and can identify characteristics such as bolus flow and swallow physiology. Based on this definition, only objective measures i.e. VFSS or FEES will be considered true diagnostic evaluations.

The more difficult distinctions that need to be made are surrounding those of the bedside or clinical examination. Also called the clinical bedside evaluation, the definition for this procedure is seemingly even more variable than for that of screening. Many clinicians use this terminology to describe what is perceived to be a more thorough evaluation of swallowing without using an objective measure. Table 1 shows those studies that use the clinical bedside terminology indicated by CBE (clinical bedside evaluation). The criteria for CBE vs screening is not well defined in the literature and it seems that it is left to the discretion of the author how they choose to classify their assessment of choice. In general, it seems that the CBE more consistently has additional assessment measures. For example, most CBEs reviewed had a history component, an oral mechanism component, and oral feeding trials. However, this was not true of all CBEs that were reported on and there were additional CBEs that did not contain all of these components. Furthermore, there were screening tools that contained these measures. If one removed the CBE label it might be difficult to determine how the authors classified their swallowing tool simply by looking at assessment measures alone.

In order to clarify any semantic confusion, for the purposes of this critical review, the clinical bedside evaluation will be considered a screen. By the definition of evaluation above, a

CBE cannot be classified as a true evaluation as it cannot identify anatomy, physiology, or bolus flow characteristics. A CBE can only determine the presence or absence of aspiration risk.

Furthermore, dysphagia will be a term reserved for a swallowing disorder that is defined by abnormal anatomy, physiology, and bolus flow characteristics. The term swallowing will be utilized to identify more general criteria such as aspiration risk. While the use of this terminology may vary due to the way the authors have chosen to define statistical measures, the reader should be mindful of these definitions when interpreting the discussion portions of this review.

It should also be noted here that the unit of measurement for liquid volume varies by author. Both milliliters (ml) and cubic centimeters (cc) are used in this literature. One milliliter equals one cubic centimeter and thus the units can be used interchangeably.

Evidence Based Screening

There is significant variability in the way screenings are conducted. Current literature does not provide a consensus on methodology, stimuli used, criteria for passing/failure, or variables that should be included as part of a swallowing screening tool. McCullough and colleagues (1999) conducted an investigation of clinician preferences and practices for both clinical swallowing examinations and VFSS in the adult neurogenic population. Their findings demonstrated the variability in clinical methods that are used to assess swallowing. While they found that clinicians typically utilize those methods that “they believe they should use”, they identified a lack of consensus regarding the perceived importance and use of oral mechanism examination and trial swallows as part of their evaluation. The study also found that clinicians are relying on the oral mechanism examination to assess for dysphagia risk despite the fact that no evidence has been found to support its use in detecting dysphagia.

This same study (McCullough et al., 1999) elegantly reported all preferences and practices for conducting a clinical swallowing examination by breaking down the bedside assessment into four categories of history, oral motor, voice, and trial swallows. Each category was assigned evaluation measures i.e., history was made up of 14 measures, oral motor 15, voice 7, and trial swallows 21. For example, oral motor measures included tongue strength, gag, and voluntary cough while trial swallows included 3-ounce swallow, puree, ice chips, swallow delay, and laryngeal elevation among others. Alarming, only 56% of those methods utilized for assessment had support from the literature. While helpful in identifying a call for research, the

results of this study elucidated the fact that current practice in swallow screening is not particularly well supported by current evidence in the literature.

McCullough et al (2000) conducted a subsequent study that investigated the reliability of clinicians conducting clinical swallowing examinations. Many studies determine reliability of clinical examinations as compared to objective assessment measures. This study, however, posed a novel and valid research question. No prior studies considered reliability for measures of the clinical examination itself. Without this information, it is difficult to draw conclusions about whether or not a clinical examination is useful for detecting aspiration. Overall, McCullough and colleagues (2000) found that clinicians demonstrated sufficient inter- and intra-judge reliability on less than 50% of those measures typically used on a clinical evaluation. Measures were again categorized as they were in their prior study (McCullough et al., 1999). In the oral motor category the most unreliable measure was gag reflex and the most reliable were oral mechanism and speech intelligibility ratings. In the voice category, the most unreliable measure was sustained /a/ and most reliable was voice rating using a speech task. Trial swallows were found to be largely unreliable. Only four out of 26 measures were rated with sufficiently high inter and intra-judge reliability. The only factors involving swallow trials that were found to have sufficient reliability were duration for thin liquid, estimate of oral stasis, 3-ounce water swallow test, and overall dysphagia rating. Results of this study are important when conducting a critical literature review as it calls into question the reliability and validity of the many subjective variables and tasks that are frequently utilized as part of many published swallow screens.

In a time when clinicians are called to a heightened level of accountability through evidence based medicine, a choice for best practice should be made only after review of the literature. Careful consideration of these two studies (McCullough et al., 1999; McCullough et al., 2000) highlight the lack of evidence and reliability in much of the current screening practice.

Table 1 is a summary of reviewed literature in swallowing screening. This table can also be used to support the need for evidence based dysphagia screening as it demonstrates significant variability in both the stimuli used as well as in the sensitivity and specificity of screens reviewed. This is not to imply that there is only one effective screening tool. However, variability in methodology, sample size, population, and swallowing stimuli makes analysis of the current evidence challenging to interpret without a more detailed consideration of methodology and statistical viability of each study.

Table 2 is a report of studies which identified sensitivity and specificity for each variable administered on a clinical examination/screening. Only a small subset of the reviewed studies conducted this statistical analysis. Results for use of these individual variables in general are poor and not well supported by current evidence.

Definition of an Effective Screen

As defined from a methodological perspective, a successful screen should be simple to administer, cross-disciplinary, cost effective, acceptable to patients, and able to identify the target in question by yielding a positive finding when aspiration risk is present and a negative finding when dysphagia is absent (Cochrane & Holland, 1971). An effective screen must be defined by more than methodology – it must also be defined statistically. There are a number of standard statistical measures that are prevalent in the swallowing screening literature.

Appendix A shows a 2x2 contingency table that is frequently used to conceptualize the statistical measures involved in determining the effectiveness of a screening tool. In each case, the conditions under which the statistics are to be interpreted must be defined. True and false positives are the values from which all other statistical calculations are made. Suiter et al (2009) reports results on the 3-ounce water swallow challenge based on results of FEES. This example can be used to conceptualize the definitions of true and false positive and negatives. A true positive resulted if a patient failed the 3-ounce challenge and demonstrated aspiration on FEES. A true negative resulted if a patient passed the 3-ounce challenge and did not aspirate on FEES. A false positive resulted if a patient failed the 3-ounce challenge but did not aspirate on FEES. A false negative resulted if a patient passed the 3-ounce challenge despite aspirating on FEES (Suiter et al., 2009). These true and false positives and negatives are used to calculate statistical measures such as sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) that provide information about the validity of a given test.

Sensitivity of a particular clinical sign (wet voice after swallow for example) for detecting a sign on objective measure (aspiration on VFS) is defined as the proportion of patients who have the sign (aspiration) who also have the clinical sign (wet voice). Sensitivity measures a test's ability to identify an individual with the disease as positive. Tests that are found to be highly sensitive indicate that there are few false negative results, and thus fewer cases of disease are missed (Rosenbek et al., 2004).

Specificity of a particular clinical sign (wet voice after swallow) is the proportion of patients who do not have the sign on objective measure (aspiration on VFS) who also do not have the clinical sign (wet voice). Specificity measures a test's ability to identify an individual without the disease as negative. Tests that are found to be highly specific indicate fewer false positive results, and are able to help 'rule in' the diagnosis. (Rosenbek et al., 2004).

Positive and negative predictive values are also utilized as measures to determine the value of a test. Positive predictive value (PPV) is defined as the proportion of the patients who are positive for a clinical sign that who also have a particular disease and negative predictive value (NPV) is the proportion of patients who are negative for a clinical sign who do not have a particular disease (Go, 1998). Reporting these parameters allows statistical interpretation to shift from reporting the percentage of people who aspirate and have the clinical sign to how well a clinical sign predicts aspiration. (Rosenbek, 2004)

Likelihood ratios (LR) allow interpretation of how 'likely' something is to occur. They express the likelihood related to a particular condition. For example, in a given context, a LR of 1.4 means that the likelihood of a cough after swallow is 1.4 times greater for someone who

aspirates than someone who does not. A negative likelihood ratio can also be expressed, providing information about a condition being less likely based on a clinical sign (Rosenbek, 2004).

While ideally each screening tool should have equally high sensitivity and specificity, this is not often the case. In the swallowing screening literature, the higher the sensitivity, the lower the specificity. That is, the screening tests that are successful at identifying those patients with a particular disorder often over-identify patients who do not have the disorder (Logemann, 1999). While favorable from an identification perspective, the clinical implication of a low specificity is over-referral for objective testing, withholding of oral feeds and medications, and unnecessary use of naso-gastric feeding tubes (Leder et al., 2002). While these potential drawbacks must be considered when determining criteria for a statistically sound screening tool, Leder et al. (2000) accurately report that each clinician must determine their own ethically driven parameters for accuracy. While there is no consensus for a defined target parameter for sensitivity, Leder et al. 2002 report that an effective screening tool should have a sensitivity of 95% or greater.

Critical Caveat: Silent Aspiration

Without silent aspiration, current practice in swallowing assessment would be significantly altered. The role of objective assessment would likely be different. The discussion surrounding screening versus evaluation would likely be unnecessary. However, silent aspiration remains of critical importance in the determination of what evaluative process should be considered in the assessment of dysphagia. Clinicians that are called upon to make an ethical determination about acceptable risk when evaluating swallowing must not only understand the validity of the tests they are using but also possess the knowledge of the prevalence and characteristics of silent aspiration.

Silent aspiration occurs when material passes below the level of the true vocal folds without overt signs e.g., coughing or choking that can be detected solely by observation of the patient without visualization of the aspiration event. By this definition, silent aspiration would be undetected on all screening tools or clinical evaluations of swallowing. Leder et al. (2002) attributed the unreliability of clinical examinations to the high incidence of silent aspiration i.e., 28%-52% in patients referred for swallow testing. In theory, then, a clinician would have to consider that up to 50% of patients could be potentially missed via silent aspiration by any given screening tool.

In 2011, Leder, et al. published a study that contributed novel evidence that silent aspiration is volume dependent, thus expanding the utility of the 3-ounce water swallow challenge protocol. In this study (n=4,102), Leder and colleagues used the 3-ounce water swallow challenge in conjunction with FEES to demonstrate that silent aspiration is volume

dependent. That is, patients who were shown via FEES to be silent aspirators of smaller volumes (up to 5ccs) demonstrated overt signs of aspiration when they subsequently drank larger volumes (3 ounces or 90ccs). This finding helped to elucidate why particular swallowing screens had higher sensitivities and lent support for screening tools which include PO trials with larger volumes of liquid intake.

Swallowing Screen Stimuli

As can be seen in Table 1, screening modality varies significantly across the swallow screening continuum. Screening tests consist of both the assessment of items that are associated with swallowing i.e., water swallow test, and those that do not require a swallowing task, i.e., a questionnaire, facial asymmetry, lingual deviation, or dysarthria.

Individual Efforts Associated With Swallowing

Most, but not all, swallowing screening tools utilize some form of a swallowing trial. There is, however, high variability in what per oral feeding (PO) trials are chosen for administration. Swallowing trials may take the form of a water swallow test (WST) or a bolus swallow test (BST). Water swallow tests are defined by swallow trials that involve water only and are generally identified using this nomenclature by the authors. Bolus swallow tests (BST) are those that involve administration of multiple and alternative PO consistencies. For the purposes of this review, the above definition of BST will be used for descriptive purposes regardless of whether or not authors designate the screen as such. Additionally, there are several tests that use specialized food types with reported justification for enhancing the swallowing response, e.g., sour milk (Westergren et al., 1999) and lemon ice (Weinhardt et al., 2008). In some screens water swallow trials are also embedded within the BSTs, however these are not considered WSTs proper. Some screens consist only of a water swallow test and some of bolus swallow test. Others use novel combinations thereof. Table 1 demonstrates this variability across swallow screens as well as variability of volume that is administered during trials.

Volumes administered for the water swallow test range from 3 milliliters (mls) to 100 mls with varying increments of 3mls, 5mls, 30mls, 60mls, 90mls, and 100mls. Bolus swallow tests vary by both consistency and volume. Consistencies offered include thin liquids, thickened liquids, puree, semi-solids, and solid consistencies. Volumes of consistencies are also variable. Quantities range from 1ml to 60mls, with reported increments of 1ml, 0.3teaspoons (tsp), 0.5tsp, 3mls, 5mls, 10mls, 20mls, 50mls, and 60mls.

Individual Efforts Not Associated With Swallowing

Swallowing screening tools often, though not always, include other measures that are purported to predict aspiration risk. Screens can include questionnaires, a medical history, subjective variables, oral mechanism evaluation, and other measures such as pulse oximetry and scores from scales such as the National Institutes of Health Stroke Severity Scale (NIHSS). Oral mechanism examination can include assessment of labial and lingual strength and range of motion, dentition status, and palatal range of motion. The category of subjective variables is arguably the most varied including measures such as gag reflex, voluntary cough, dysphonia, dysarthria, secretion management, orientation, wet vocal quality, and command following.

The matrix of these efforts is highly variable across screening tools. Logemann et al. (1999) used a total of 28 variables on their screening tool, 21 of which fall under the category of efforts not associated with swallowing, with the majority of those being subjective. In contrast, a screening procedure by Hammond et al. was published in 2009 which utilized sophisticated measurements of the voluntary cough alone to predict aspiration risk. Though these examples are outliers in terms of numbers of variables included, they demonstrate not only the wide range but the lack of standardization of the screening tools available to clinicians.

Water Swallow Tests

DePippo et al. (1992) were the first to report data on the 3-ounce water swallow test. The study investigated 44 stroke rehabilitation patients and used VFSS as the objective measure by which to validate the 3-ounce water swallow test. This screening test required the patient to demonstrate uninterrupted drinking of 3 ounces of water without overt signs of aspiration. Failure criteria were inability to drink continuously, cough during or up to 1 minute after completion of test, or a wet/hoarse voice quality post-swallow. The authors reported the relationship between cough or wet/hoarse voice on the screen and aspiration on VFSS to have a sensitivity of 76% and specificity of 59%. Sensitivity and specificity of these same signs on the 3-ounce screen were revised to 94% and 26% respectively when considering the relationship between these signs and aspiration of greater than 10% of the bolus on VFSS. They went on to report that the relationship between these overt signs on the 3-ounce swallow and aspiration of thickened liquids or solids on VFSS was found to have a sensitivity of 94% and a specificity of 30%.

This study is methodologically sound in that it uses an objective measure to determine aspiration status. Clinical conclusions gleaned from statistical data are reasonable in terms of reporting accurately what information the screen is able to provide. Drawbacks of the study include a small *n* of only 44 patients. The authors did not include a power analysis to determine if this sample size was adequate to avoid a Type II error. Additionally, methodology indicates that reliability was addressed in the study, but the authors failed to report any reliability data to substantiate their methodology. Methodology does not indicate that the study was blinded which

further calls into question the reliability of the results presented. It is also questionable how the authors determined the percentage of material aspirated. They only reported that examiners ‘estimated’ the percentage. Without the ability to quantify the amount aspirated, it is challenging to justify clinical use with this reported statistic.

While the statistical analysis of this study was interesting from a research perspective, one must call into question the clinical utility of the second two conclusions i.e., results that indicated presence of greater than 10% of the bolus and aspiration of thickened liquids. Once completed, it is difficult to ascertain what these hypotheses add to the clinical information that the clinician is receiving. If the patient coughs on the 3-ounce screen, the clinician is most interested in overall aspiration risk. The other measures are clinically irrelevant and redundant. To report higher sensitivities under these conditions gives the non-discerning clinician the false impression that the sensitivity is higher than it truly is from a clinical utility perspective. This leaves the clinician with one clinically relevant conclusion – that with a sensitivity of 76%, the 3-ounce swallow test will potentially miss 24% of patients who aspirate.

In 1994 DePippo et al. again investigated the 3-ounce water swallow test in an alternative context. Named the Burke Dysphagia Screening Test (BDST), it consisted of a screening questionnaire in conjunction with the previously reported 3-ounce water swallow test. The questionnaire screened for presence of bilateral stroke, brainstem stroke, history of pneumonia during acute phase of hospitalization, persistently decreased oral intake (less than half of meals consumed), prolonged time to eat, non-oral feeding in place. If the patient screened positive for any one of these factors or failed the 3-ounce screen, they failed the BDST. If they screened

negative for all of these factors and passed the 3-ounce screen, they passed with BDST and were ordered a regular diet. Subjects in this study were also stroke rehabilitation patients (n=139). Unlike the previous publication, DePippo did not report on risk of aspiration or dysphagia. The results of the BDST were instead compared with the development of pneumonia, recurrent upper airway obstruction, and death.

Statistical analysis demonstrated that the relative risk of developing any of these above complications was 7.65 times greater for those failing versus passing the BDST and that the relative risk for these complications was 2.28 times greater for those failing the 3-ounce alone.

With the exception of lack of reliability data, this study appears to be grossly methodologically sound. Criteria for classification of pneumonia were quite rigorous. Upper airway obstruction was recorded only if recurrent i.e., if patient required the Heimlich maneuver more than once during their stay. It should be noted that complications were recorded at time of discharge and so covered only that period that the patient was on the rehabilitation unit. The length of stay was not reported.

Despite these minor methodological issues, this study provides interesting information about a patient's risk of developing pneumonia, upper airway obstruction, or dying during their rehabilitation stay. However, it fails to give the clinician any practical information about aspiration risk at the time of the screen. Certainly, when combined with DePippo's earlier study (1992), a less than optimal sensitivity of 76% can be considered when utilizing the 3-ounce portion of this test, but the BDST proper offers no novel information about aspiration risk per se. Additionally, it is difficult to determine what relative risk really means in this study. Patients

that fail the BDST went on for further swallowing assessment and modification of oral intake. It is unclear what subset of the population that failed developed such complications. As patients were recommended an oral diet based on the results of the study, it might have been clinically beneficial to provide data on outcomes specifically related to their oral intake. The fact that patients have an increased risk of developing these complications if they fail this screen is more clinically meaningful if the reader knows the PO status of those patients who failed.

In 2008, building on DePippo's prior research, Suiter and Leder published data on utilization of the 3-ounce water swallow test in 3,000 hospitalized patients. Given the limitations of sample size in previous studies by DePippo, the goal of this study was to determine if the 3-ounce water swallow test was able to be generalized to a larger and more heterogeneous patient population. Additionally they sought to determine if PO recommendations could confidently be made based on the result of the screen – a critical component of the literature that was also not addressed in prior studies.

Suiter et al. (2008) used FEES as the objective measure in this study. Each of the 3,000 participants received both a FEES and a 3-ounce water swallow test. The tests were performed immediately one after the other. Subjects were acutely hospitalized patients in 14 different diagnostic categories. Criteria for passing and failure were the same as in earlier studies, that is, the patient had to drink 3-ounces of water uninterrupted and without overt signs or symptoms of aspiration. Criteria for failure of the test included inability to drink the entire amount, stopping and starting or coughing or choking during or immediately after completion. Findings were favorable and in support of generalization of the use of the 3-ounce water swallow screen in a

heterogeneous population with a sensitivity for predicting aspiration of 96.5% and specificity of 48.7%. This indicates that the 3-ounce water swallow test is an excellent predictor of aspiration in a heterogeneous population, potentially missing only 3.5% of patients. As discussed by Suiter and Leder, the 3-ounce water swallow test does not meet the strict criteria of an efficient screening tool as it does not have both high sensitivity and high specificity. The result of low specificity indicates that patients will be over-referred for objective testing. That is, it unnecessarily restricts liquid intake for nearly half of the people tested. Furthermore, approximately 70% of the individuals who failed the 3-ounce went on to tolerate some form of oral intake successfully based on FEES results.

Methodology of this study was sound. The authors not only reported results for the test as a whole, but also reported statistics on each of the 14 diagnostic categories. Given the unprecedentedly large sample size, each diagnostic category had meaningful statistical measures to guide clinicians on use of this screening tool in specific populations. Range for sensitivity was 92.7% in right stroke to 100% in head and neck surgery, neurosurgery, brainstem stroke, Parkinson's disease, and dementia. Specificities were relatively low across categories (range of 25.4% to 67.3%). Positive and negative predictive values and likelihood ratios were also reported for each category.

The second question that Suiter and Leder (2008) answered is of critical importance in clinical practice. What clinical information does one receive from the passing of a 3-ounce water swallow challenge? Based on the results of this study, clinicians can confidently recommend an oral diet based on the 3-ounce water swallow test alone without the need for additional objective

assessment. Statistical support indicates that sensitivity of the 3-ounce test for identifying individuals deemed safe for oral intake based on FEES is 96.4% and specificity is 46.4%. This study was the first published to provide information about diet recommendations that can be made based on the results of the 3-ounce water swallow test. Importantly, this not only allows patients to resume earlier PO intake but also eliminates the need to use thickened liquids and reduces the potential risk associated with tube feeding while awaiting objective assessment.

While this study undoubtedly made significant contributions to the literature on swallowing screening, it is lacking adequate reliability data. Intra-rater reliability was addressed via a previous study, however, no inter-rater reliability data specific to this study was provided. Although lack of blinding is an issue, it was mitigated by additional post-study blinded FEES testing resulting in 100% identification for tracheal aspiration (Leder, Suiter, Murray & Rademaker, 2013; Leder, Judson, Sliwinski, & Madson, 2013). Also, a recent prospective double-blinded videofluoroscopic study confirmed the clinical usefulness and validity of the 3-ounce challenge for determining aspiration risk (Suiter, Sloggy, & Leder, In Press).

Additionally, while a diet was recommended based on results, decision for PO status was only described in terms of dentition status i.e. patient was to receive a regular or soft diet if dentate and a puree diet if edentulous. The authors do discuss the importance of clinical judgment in that 98.5% of patients who passed were recommended for oral diet, and the remaining 1.5% were not due to patient specific factors that deemed those patients unsafe. While this discussion was insightful from a clinical standpoint, more direction and statistical evidence for consideration of these factors would be helpful for the practicing clinician who is

attempting to make diet recommendations. Additionally, no longitudinal data was reported to determine if patients tolerated the recommended diet.

In 2009 Leder et al. began to address some of these issues by publishing a study that supported the use of a brief cognitive screen in conjunction with assessment of swallowing. This study highlights the importance of considering additional factors when assessing for aspiration risk. It was reported that the odds of liquid aspiration were 31% greater for patients not oriented to person, place, and time. They also found that for patients who were unable to follow 1 step commands, the odds of liquid aspiration, puree aspiration, and being deemed nil per os were 57%, 48%, and 69%, greater, respectfully, than for patients who were able to do so.

Following this study, Suiter, Leder, and others went on to expand the utility of their original study using the 3-ounce water swallow challenge (2008). In 2011 and again in 2012, Leder and colleagues reported more substantial evidence about diet recommendations based on the 3-ounce water swallow challenge. These studies, Leder et al. 2011 (with n=75 stroke patients) and Leder et al. 2011 (n=493 intensive care and step-down unit patients) investigated patient's ability to tolerate a recommended oral diet 24 hours following passing the 3-ounce water swallow screen. Results indicated that diet recommendations were followed and that patients were tolerating recommended diets without overt signs or symptoms of aspiration with 100% success after 24 hours. A comparison study similarly supported the use of the 3-ounce swallow screen for making diet recommendations in 1,000 general hospitalized patients (Leder et al., 2012), thereby expanding the use of this swallow screen protocol to virtually all hospitalized patients. An additional study, conducted by Suiter et al. in 2009, indicated that the 3-ounce

water swallow test can be successfully utilized in the pediatric population with a sensitivity and specificity of predicting aspiration during FEES of 100% and 51.2% respectfully. It was also shown that this screen can be utilized to make diet recommendations demonstrating a sensitivity of 100% and specificity of 44% for predicting oral intake.

The combined results of this series of studies on the 3-ounce water swallow challenge offer compelling evidence in support of widespread use of this water swallow screening tool for detecting aspiration risk in the vast majority of hospitalized patients.

Daniels et al. (1997) utilized another form of water swallow test combined with assessment of subjective clinical features to determine if the screening tool could distinguish patients with mild dysphagia or normal swallowing from those with moderate to severe dysphagia. The authors defined moderate dysphagia as two or fewer instances of aspiration on a single consistency and severe dysphagia as aspiration of more than one consistency. The screening tool consisted of a water swallow test using 5ml, 10ml, and 20 ml volumes. The clinical features that were coded for statistical purposes included dysphonia, dysarthria, abnormal volitional cough, abnormal gag, cough following swallow, and voice change following swallow. Results of the screen were compared with VFS. Subjects included 59 acute stroke patients.

Results of Daniels' study indicated that the presence of two out of six of these clinical features is able to distinguish dysphagia severity with a sensitivity of 92.3% and a specificity of 66.7%. To clarify, the presence of two out of these six features accurately predicted with 92.3% certainty which patients had moderate to severe dysphagia. It is of interest to note that taken

individually, the sensitivities and specificities of each of these clinical features were poor (sensitivity range of 30% to 76% and specificity range of 60% to 87%). See Table 3 for details.

Relative strengths of this screening tool included use of an objective assessment (VFSS) to corroborate severity of dysphagia, adequate statistical measurements, and report of both inter- and intra-rater reliability. Drawbacks include a lack of power analysis to determine if sample size was adequate to avoid a Type II error (n=59 patients with 6 variables). Additionally, the average time between swallow screen and formal swallow evaluation was 48 hours. This is too long a period in an acute stroke patient given this population's susceptibility to rapid change in medical and neurological status, thus potentially confounding results of the swallowing screen or evaluation. While not a criticism of the study itself, clinicians should be cautioned that a sensitivity of 92% means that almost 10% of the time this screening tool is not going to be able to identify someone with a moderate to severe dysphagia.

A subsequent study by Daniels et al. in 2000 aimed to evaluate outcomes in acute stroke patients using this same screening tool, that is six clinical features paired with 5ml, 10ml, and 20ml water swallow trials. The sample consisted of 56 stroke patients. Methodology was similar to the previous study in that these results were compared with VFS. An addition to this study was the consideration of outcome measures, namely dietary status at discharge and development of aspiration pneumonia. Chi square analysis revealed that the presence of two or more clinical predictors significantly distinguished patients with normal swallow or mild dysphagia from those with moderate to severe dysphagia. Results of outcome measures were

that no patients developed complications or pneumonia during their hospital stay and that 52/54 patients had returned to a regular diet at the time of discharge.

The results of this study support Daniels' prior work (1997). Of concern again in methodology is the lack of power analysis to determine adequate sample size, lack of reliability data, as well as timing of screen vs. VFSS. In this study the criteria was that the VFSS be completed within five days of the screen, a latency even greater than in the previous study which is unacceptable in the acute stroke population due to their rapidly changing neurologic status.

Leder and Espinosa (2002) published a study considering the "two out of six" clinical variables described by Daniels (1997). In this study they compared the clinical examination which consisted of the same clinical factors: dysphonia, dysarthria, abnormal gag, abnormal volitional cough, cough following swallow, and voice change after swallow with FEES. The PO trials consisted of single sips of water boluses via a straw. A rating of no aspiration risk was made if zero or one of the clinical identifiers were present and a rating of aspiration risk was given if two or more of the clinical identifiers were present. Results indicated that the water swallow screen had a sensitivity of 86% and a specificity of 30%.

Leder and Espinosa (2002) accurately concluded that this clinical assessment underestimated aspiration risk in patients with aspiration and over-estimated aspiration risk in patients who did not exhibit aspiration. Clinically they cautioned the clinician against using this clinical bedside measure and, as this was prior to the publication of robust data that was presented on the 3-ounce water swallow test, recommended objective testing to determine aspiration status.

This study was methodologically sound, however, the sample size was small and no power analysis was reported to ensure avoidance of a Type II error. Additionally generalizability of the results is unknown as this sample was homogeneous in the diagnosis of stroke patients. Overall, results are sound and authors communicate appropriate clinical cautions about utilization of such a clinical screening tool.

Massey et al. developed the Massey Bedside Swallowing Screen in 2002. This water swallow screen consisted of two trials, first one teaspoon and second 60 mls of water. Screening for aphasia, dysarthria, gag, and voluntary cough, as well as an oral mechanical examination were also conducted. The sample consisted of 25 stroke patients. Chi square analysis was reported and it was found that this screening tool demonstrated the statistically significant ability to differentiate between dysphagia present and dysphagia not present.

Massey and colleagues demonstrated adequate reliability data for this study. They reported that VFSS was used to confirm presence or absence of aspiration, however, this was not explicitly stated in the methodology of the study. The authors again do not provide a power analysis to support the use of an adequate sample size given the number of variables used. However, given the given the small n , the authors recognized that a larger sample size would be desired. The authors also reported specificity and sensitivity values of 100% in the abstract and summary, but this information was not found explicitly in the results section in the body of the text. Therefore, it is impossible to interpret the accuracy of this statistic.

It should be noted that while Massey et al. (2002) included a questionnaire for subjective variables in their screen, the presence of such variables (aphasia, dysarthria, facial asymmetry,

etc.) appears to be utilized as only a screening tool to trigger the need for a speech consult to be placed and had no impact on the swallowing screen proper. The authors took care to provide detail about the content validity of the screen. Much of the justification for this validity is irrelevant, however, as the majority of the content of the screen is not being utilized to determine swallowing success. Essentially this screen is a WST with a single teaspoon and 60ml trial.

In 2004 Wu et al. reported findings on a 100 ml water swallow test. The screen consists of the WST only and was conducted on 59 stroke patients. Presence or absence of aspiration was confirmed via VFS. Wu chose to examine the variables of swallowing speed and choking/dysphonia following the water swallow trial. It was demonstrated that the sensitivity of swallowing speed in detecting dysphagia was 85.5% and specificity was 50% while the sensitivity of choking or dysphonia post-swallow for detecting aspiration or penetration was 47.8% and the specificity was 91.7%.

While methodology is generally sound, the utility of the reported results are questionable. This study defined abnormal swallowing speed as below 10 ml/second. Measurements are reportedly more complicated when the entire 100ml is not consumed and are irrelevant if the patient coughs prior to drinking 10mls. The remaining statistics were not supportive of use of the overt signs of choking or dysphonia post-swallow to determine aspiration /penetration status. A sensitivity of 47.8% is inadequate to justify use of this screening tool based on this data alone. The authors' claim that "choking in the 100ml WST can be used as an alternative approach to follow up on aspiration status" (Wu et al, 2004) is erroneous. The purpose of a screen does not change based on whether it is an initial screen or follow up. Potentially missing 52% of patients

at-risk for aspiration is unacceptable. The reason for such poor sensitivity cannot be confirmed, however, the small sample size and use of both penetration and aspiration as outcomes should be considered when comparing statistics in similar studies. In addition, no screening test can determine the bolus flow characteristics of laryngeal penetration.

McCullough and colleagues (2005) investigated use of a screen which consisted of four sections: history, oral mechanical examination, voice/speech praxis, and trial swallows with 5ml, 10ml, and 3-ounces. The screen was conducted on 165 stroke patients. The authors reported adequate inter- and intra-rater reliability. Statistical reporting was involved. Specificities and sensitivities for each measure were reported individually. When measures were considered individually, sensitivities were quite poor and not clinically useful. This data can be found on Table 2. The authors reported the remaining statistics by categories of the screen. The two history measures most useful in detecting aspiration were pneumonia and poor oral hygiene with sensitivities of 9% and 14% and specificities of 98% and 97%, respectively. The low sensitivities and high specificities can be interpreted clinically to mean that the presence of both of these factors are much more meaningful than their absence. For the category of voice, the two best measures for detecting aspiration were found to be breathy voice and wet/gurgly voice (specificity of 98% and 96% and sensitivity of 16% and 22%, respectively). Regression analysis demonstrated that the best measures in the model for detecting aspiration overall were failure of the 3-ounce water swallow test, unilateral jaw weakness, and dysphonia. The 3-ounce water swallow test had a likelihood ratio of 9.5.

This study is methodologically sound and statistically well organized. An additional strength of this study is the fact when VFSS was used as an objective measure, the 3-ounce swallow was used as point of comparison. This methodology is not frequently seen as investigators tend to conduct a standard VFSS with smaller quantities or ill-defined larger quantities of liquid swallowing for testing. Such a rigorous assessment on VFSS is more consistent with the screening tool that is being utilized. While much interesting and important screening information can be gained from this study, a disadvantage from a clinical utility perspective is that the identified best predicting measures were not investigated together as a novel screening tool. This may be of interest for future research.

In 2009, Turner-Lawrence et al. published a feasibility study investigating the sensitivity of a water swallow screening tool to be used by emergency room physicians that consisted of a questionnaire and a 10 ml water swallow test. The questionnaire included inquiry about swallowing complaints, voice changes, facial droop, and aphasia. A positive screen for one of these factors resulted in the failure of the screen. If passed, a cup containing 10mls of water was offered to the patient. Criteria for failure of the water swallow portion of the screen were coughing or choking during the test or voice change or a decrease in pulse oximetry greater than 2% after the test. Authors indicated that this screening tool offered promising preliminary results for physician swallow screening in the emergency department. Reportedly sensitivity of this tool was 96% and specificity was 56%. Results should be interpreted with extreme caution. The authors recognized that there were significant limitations associated with this study.

As the authors recognized, the sample size (n=84) was small. More concerning, however, was the volume of water tested. Ten milliliters is an insufficient quantity of water to assess for aspiration risk when considering previously reported results that silent aspiration is volume dependent (Leder et al., 2011). Equally as problematic was the fact that no objective measure was reported to have been used as criterion standard for comparison with this screening tool. Investigators reported that the results of the physician screen were compared with those of the standardized dysphagia assessment performed by a speech-language pathologist, but they do not explicitly state what this assessment entailed. Without this information, it is impossible to interpret the results of the study with any degree of certainty. Screening questions chosen are, in general, subjective in nature with limited evidence in the literature to support use of such factors in assessing for aspiration risk. The use of oxygen saturation as an indicator for aspiration is not well supported in the literature. Leder (2000) reported that there were no significant differences in SpO₂ based on aspiration status during FEES and Ramsey (2006) found a poor association between oxygen desaturation and aspiration on VFSS. While this feasibility study does appropriately support the use of a screening tool by other medical professionals, i.e., physicians, the limitations of this study make it difficult to put this screening tool into clinical practice without further investigation.

The Toronto Bedside Swallowing Screening Test (TOR-BSST) is a swallowing screen developed by Martino et al. in 2009. This screen was investigated using 311 stroke patients and consists of five items: voice before water swallow, lingual movement, pharyngeal sensation, water swallow, and voice after swallow. The results of this tool were compared with VFS as an objective measure. Martino et al report 'excellent' validity with sensitivity at 91.3 %.

Strengths of this study included reliability data provided and defined parameters for inclusion as well as the use of rating scales such as the NIHSS to define the patient population. The investigators made efforts to formally educate screeners in a four hour session to ensure they were adequately trained to administer the screen. The authors also provided evidence from the literature to support chosen test items. Also favorable is the authors' support of nurses performing the screen.

This assessment is frequently cited in the literature, however, methodology is lacking in that the procedure for performing the assessment is not explicitly stated in the methods. The authors discuss the development of the test using the Kidd water swallow test (50ml water swallow test) and the four other items selected for the test, however they fail to define the actual procedure or criteria for pass/fail. The lack of this information leaves a clinician who is interested in utilizing this tool less than confident about the procedure that would need to be put in place to ensure the TOR-BST was being carried out properly. A web search revealed that a clinician must pay \$400, attend a training course, and be certified to administer the TOR-BSST.

One important criticism of this study is the fact that only 20% of the patients received VFSS for comparison to the TOR-BSST. The authors explained that limits on unnecessary irradiation exposure limited ability to evaluate all subjects. While power analysis was reported at 80%, it seems that calculations of sensitivity would have been more accurately reported with a higher percentage of objective measurements conducted.

Edmiaston and colleagues in 2010 developed the Acute Stroke Dysphagia Screen (ASDS). The ASDS is comprised of a nurse administered 3-ounce water swallow test with four

screening questions regarding facial, lingual, palatal asymmetry and weakness and a Glasgow Coma Scale (GCS) rating less than 13 as a cutoff. The subject sample consisted of 300 acute stroke patients. The study found that sensitivity and specificity for aspiration were 95% and 68% respectively and the sensitivity and specificity for dysphagia were 91% and 74% respectively.

Again, this study supports the use of nurse screeners. The nurses were provided training on the screening tool to ensure proper administration. Adequate reliability data was provided. The authors drew reasonable conclusions about the data presented and make realistic recommendations about how the screen should be used. The chief criticism of this study is that there was no objective criterion standard used.

The Mann Assessment of Swallowing Ability (MASA) (Mann, 2002), a clinical/bedside swallowing evaluation was used as the comparison measure. While the MASA has been validated against VFSS, the authors should not assume that it is an acceptable replacement for an objective evaluative measure. Additionally, mean times for administration of the assessment were 8 hours for the ASDS vs 32 hours for the MASA. In acute stroke patients this latency is problematic especially when considering the assessment performed by the speech-language pathologist (MASA) requires no special equipment or scheduling other than the availability of the patient participant. The lack of use objective criterion standard should caution clinicians when interpreting the seemingly favorable results of this assessment.

In 2013, Edmiaston and colleagues presented a similar screen called the Barnes-Jewish Hospital Stroke Dysphagia Screen (BJH-SDS). An improvement over the first publication, VFSS was used as criterion standard. The BJH-SDS also screened for GCS less than 13, facial,

lingual, and palatal asymmetry. If the screening questionnaire was passed for these parameters, a 3-ounce water screen was administered. Any positive screening item resulted in a fail and no 3-ounce swallow screen, ultimately leading to a high fail rate. While a sensitivity of 94% and specificity of 66% were reported for detecting dysphagia, it is unclear how authors came to this conclusion based on the data presented. As would be expected with such strict use of non-validated screening items, many of participants failed the screen based on the questionnaire alone. While improved use of a criterion standard was noted in this study, the authors fail to provide enough information about how statistics were calculated given such a high fail rate.

Patterson et al. (2010) investigated the validity of a timed 100ml water swallow challenge as compared with FEES. They utilized this assessment in a subject sample of 173 head and neck cancer patients who were referred for irradiation or chemoradiotherapy. Data were gathered at pre-treatment and 3, 6, and 12 months post-treatment. Measurements used for criteria for performance on the swallow test were swallow volume (ml/number of swallows) and swallow capacity (ml/time taken to swallow). Sensitivity and specificity were determined for each time point and overall sensitivity and specificity were found to be >67% and >46%, respectfully.

Sufficient data on inter- and intra-rater reliability were provided. While this study does use an objective assessment for criterion standard, the authors erroneously report that the 100 ml swallowing test is a good tool for identifying patients with aspiration. This is simply not so. Potentially missing over 30% of patients does not meet the criteria for the definition of a good screening tool (Leder & Espinosa, 2002). Additionally, sensitivities and specificities varied widely across the time interval studied making results even more difficult to interpret for

meaningful use. Results of this timed 100ml test should not be compared with those of the 3-ounce water swallow test, which uses a similar volume (90ccs), as the criteria for pass/ fail are dissimilar.

Steele and colleagues (2011) published a study with the intent of comparing a swallow screening protocol with simultaneous VFSS. The swallow protocol included screening for tongue lateralization, voluntary cough, 3-5cc sips with phonation, cup drinking in 3-4 sequential sips (volume unspecified), and phonation. Findings included sensitivity and specificity as evaluated by speech-language pathologists to be 64% and 16%, respectively, for aspiration risk and 71% and 13%, respectively, for dysphagia.

Despite poor sensitivity for detecting aspiration or dysphagia, this study had a number of strengths. First, the authors chose to use same time / same swallows. This is ideal as it eliminates the question of time-lapse or variability of swallow when comparing two assessment tools using the same patient. The study provided strong data on reliability and provided data on both results of speech-language pathologists as well as nursing raters. Ultimately the authors drew accurate conclusions about their data stating that "...swallow screening decisions based on a series of 3-4 thin liquid swallows do not have good clinical utility for detecting penetration-aspiration or dysphagia." (Steele, 2011).

This study used a relatively small *n* of 40 when taking into consideration the diagnostic heterogeneity of the sample. While using simultaneous swallows is advantageous comparison purposes, the authors acknowledge that using this methodology does limit the objective measure in that the VFSS becomes an objective assessment of a screening tool, not the actual criterion

standard itself as it is traditionally conducted. Overall, conclusions about this assessment tool render it inadequate for clinical use.

The Swallow Provocation Test (SPT) was investigated by Teramoto and colleagues in 2000. The SPT involves injecting 0.4ml and 2ml of water into the pharynx via a nasal catheter while patient is supine. Latency of swallow onset was recorded. Both the SPT and a water swallow test (10ml and 30ml volumes within 10 seconds) were conducted. Criterion for passing the water swallow was drinking without interruption, no cough, and no voice change. The statistical analysis was broken down by step 1 and 2 of each test, based on volume administered. Results were reported as sensitivity and specificity of SPT and WST for the detection of aspiration pneumonia. For 0.4ml SPT sensitivity was 100% and specificity was 83.8% for detecting aspiration pneumonia. For 2ml SPT results were sensitivity of 76.4% and specificity of 100%. For 10 ml and 30 ml water swallow tests the sensitivities were 71.4% and 72%, respectively and the specificities were 70.8% and 70.3 %, respectively.

There are multiple fatal flaws in methodology and statistical analysis in this study. The authors conducted their statistical assessment based on the screening tool's ability to detect aspiration pneumonia, however, the authors define the diagnosis of aspiration pneumonia using the following criteria: an episode of aspiration, a chest x-ray showing an inflammatory response, and a chest x-ray showing an increased white blood cell count. A single episode of aspiration is not a criterion for a clinical diagnosis of aspiration pneumonia. Additionally, a chest x-ray is not capable of demonstrating white blood cell counts. Therefore, two of the three criteria are unusable. Furthermore, the authors then report that SPT is more useful than WST for

differentiating patients that are pre-disposed to aspiration. This clinical leap cannot be made without further substantiation of their findings. Aspiration pneumonia, however ill defined, is not the same clinical indicator as aspiration risk.

Additionally, the SPT uses laryngeal movement as a measure of swallow success. Laryngeal movement does not always equate to a swallow and without additional criteria, it also does not equate to a successful swallow. It is unclear what the clinical correlate of the SPT is. The authors did not provide evidence nor explain how injecting less than 0.5mls of liquid is supposed to be indicative of a clinical swallow. Given such small volumes and unconventional method of administration, patients with dysphagia enrolled in this study would have been at high-risk for silent aspiration. These irrevocable flaws overshadow other considerations such as no objective measure and small n (26 patients with aspiration pneumonia and 26 age matched controls) that ultimately make the results of this study clinically ineffectual.

Yeh et al. (2011) published a study investigating the question of whether or not dysphagia screening decreases pneumonia in the stroke ICU using the 3-step swallowing screen (3-SSS). The first step essentially defines the exclusionary criteria of decreased consciousness, prior dysphagia, need for tube feeds, intubation, poor oxygen saturation, or frequent choking on saliva. If the patient fails the first step the patient is made nil per os, placed on tube feeds, and re-evaluated in 7 days. The second step is 3mls of water over 3 trials. The third step is 100mls of water over 2 trials. The criteria for failure of the water swallow portion of the test include choking, wet voice, or slow swallowing. If the patient fails the water trial portion, the speech-language pathologist is consulted for a formal evaluation. Logistic regression demonstrated that

higher NIHSS scores, older age, and ever having a naso-gastric tube were independent risk factors for stroke associated pneumonia. Dysphagia screening was found to have a borderline effect of reduced pneumonia occurrence in stroke patients with an odds ratio of 0.42.

Clinically, this protocol is lacking in terms of patient management. If the patient fails one of the screening criteria they are made nil per os and the screen is repeated 7 days later. This is an unnecessary latency. This means that a patient admitted with altered mental status or prior history of dysphagia has an automatic seven day delay in potential resumption of oral intake. This is unacceptable clinical management and fiscally irresponsible as it increases length of stay and use of enteral tube feeding for nutrition.

Statistical results of this study should be interpreted with caution. There is no objective criterion standard to which the 3-SSS is compared to determine if aspiration is actually occurring. Subjects were divided into two groups so as to examine outcomes prior to the 3-SSS implementation and after it was put into place. While the authors provided solid criteria by which to define nosocomial pneumonia and provide data and pneumonia rates, they did not clarify whether attempts were made to determine the type or etiology of pneumonia. Not all hospital acquired pneumonia is aspiration related and this differential diagnosis is a critical distinction that is needed for meaningful interpretation of the results of this study. Without this distinction it is impossible to determine if the screening protocol did not impact pneumonia rates because it was ineffective or if it was because the patients did not have an aspiration pneumonia to begin with. These fatal flaws make meaningful interpretation of this data impossible.

Gottlieb and colleagues (1996) published a purported validation of the 50 ml water swallow test in 180 stroke patients. They conducted an oral mechanism evaluation, observed the rate and appearance of the pharyngeal swallow reflex, auscultation to voice, palatal movement, as well as a 50ml swallowing test. The authors reported that the presence of pneumonia in dysphagic patients was 2.5 times more likely than in non-dysphagic patients and concluded that use of this assessment tool allowed clinicians to reduce the frequency of pneumonia after stroke.

Unfortunately there are several methodological flaws that make results of this study difficult to interpret. Initially, the authors claimed that using their water protocol procedure was akin to the 90ml water swallow by DePippo (1992). Not only are the volumes of water used in these two tests different, but the methods are also incongruous. DePippo clearly stated that one criterion for passing the 3-ounce water swallow challenge was continuous drinking of the entire volume. The authors of the current study report that “in order to drink 50ml the patient would usually need to sip four or five times” (Gottlieb, p.530). This is not the same assessment and the authors erroneously utilize statistics from DePippo’s study to substantiate the use of this tool.

Criteria for failure of this assessment is a cough during swallow. Patients with a positive result were treated with a plan for feeding with aspiration precautions, alteration of diet, and swallow therapy. The development or resolution of pneumonia was monitored and the authors drew conclusions about a relationship between pneumonia and dysphagia. No objective measure was used to determine the absence or presence of dysphagia. Additionally, the authors admitted that it was impossible to relate all pneumonia cases to aspiration since half of those pneumonia cases occurred in patients without a swallowing impairment. This fact alone negates the results

of this study which draws conclusions based on pneumonia rates. While a clinically interesting question was posed, flawed methodology renders the results of this study meaningless.

Lim et al. (2001) reports similarly faulty data investigating a 50 ml water swallow test in acute stroke patients. Methodology involves a water swallow test completed in 10 ml aliquots, i.e., not requiring continuous drinking and in conjunction with pulse oximetry. The results of this screen were compared with FEES. Authors reported a sensitivity of 76.9% and specificity of 83.3% for the pulse oximetry, 84.6% and 75%, respectively, for the '50ml' water swallow challenge, and a sensitivity of 100% with a combined test of the pulse oximetry and the 50 ml water swallow test.

Although the authors reported these statistics, they did not provide any statistical justification for how combining two poorly performing screens can yield an effective screen with 100% sensitivity. Without a plausible explanation, statistical analysis is in question. Methodology is also flawed in this study. As noted above, sequential drinking was not required of the 50 ml water swallow. Additionally, methodology dictated that pulse oximetry measurements be taken ten minutes following the 50 ml water swallow test and using 10mls over three trials. FEES assessment was completed at a separate interval within 24-48 hours. Simultaneous assessment of these tests would have provided a more meaningful result had the statistical analysis been sound. These drawbacks make it impossible to draw clinically useful conclusions about this screening tool.

Bolus Swallow Tests

The remaining swallow screens involve swallowing trials with some combination of boluses of water, thickened liquids, puree, and/or solid consistencies. Many of these screens are considered clinical/bedside swallow evaluations, that is, they may contain a history component, an oral mechanism examination, subjective variables, and bolus swallow trials. Despite their seemingly more involved administration, the reader should be mindful that these are still considered screens as they are not able to define the anatomy and physiology of the dysphagia that is present or bolus flow characteristics. It would also be prudent to recall the administrative and methodological simplicity of the water swallow tests when considering the remainder of this literature review. For example, Logemann (1999) presents a 28-item screen and Splaingard (1988) presents a tool that takes 35-40 minutes to complete. While not all clinical screening tools stated the specifics of the administrative burden on the clinician, one should carefully consider what is to be gained from simply conducting a longer screen with more components prior to making a choice about which screen is most effective for their patient population.

One of the most frequently cited studies on the clinical bedside swallow is that of the work of Splaingard et al. (1988). This study investigated the ability of the bedside swallow assessment to predict aspiration on VFSS in 107 rehabilitation patients. VFSS was conducted within 72 hours of the bedside assessment. Bedside assessment in this study involved a case history, oral mechanical examination, voluntary cough reflex, vocal quality, respiratory quality, and gag reflex. PO trials included liquids (juice, nectar, frosty), pureed foods (applesauce, pudding), ground meat, and solids (fruit, meat, bread crackers). Patients were evaluated on vocal

quality, throat clearing, cough, signs of respiratory distress, swallow reflex, bolus control, and oral transit of bolus. Clinicians were asked to make a determination about pocketing or residue in the valleculae or pyriform sinuses and rated the subjects on a five point dysphagia severity scale that consisted of the following ratings:

1. Within normal limits: no detected swallowing abnormalities
2. Mild: slight delay in swallowing reflex but good apparent oral and pharyngeal peristalsis
3. Moderate: delayed swallowing reflex, suspected pooling in valleculae or pyriform sinuses before the swallow, heavy coating on pharynx post-swallow
4. Severe- trace aspiration with delayed swallowing reflex, pooling in valleculae or pyriforms
5. Profound: >10% aspiration, absent swallowing reflex, pooling with poor pharyngeal peristalsis

Results of this study indicated that the bedside evaluation identified only 42% of those patients that aspirated on VFSS. PPV was found to be 75% and NPV was found to be 70%. Clinically, this translates to 75% of patients who fail the bedside swallow will aspirate on VFSS and that 70% of patients who pass the bedside evaluation will not aspirate on VFSS.

Splaingard et al. (1988) accurately concluded that the bedside evaluation underestimates the frequency of aspiration in the neurological population. While the methodology was ultimately sound in terms of using an objective measure to determine aspiration status and in the

use of a robust clinical bedside examination, the authors demonstrated flawed reasoning in the assessment and scoring of their clinical examination. Clinicians were required to determine bolus flow characteristics, such as pooling in the valleculae and pyriform sinuses, which is impossible without an objective measure. Additionally, they require the use of a rating scale that contains similar descriptive criteria such as trace aspiration, greater than 10% aspiration, and heavy coating in pharynx post-swallow, that in no way can be determined via clinical bedside assessment. Thankfully, the signs that clinicians were using to rate aspiration risk were generally overt signs such as coughing and voice change that would have been unlikely to change the outcome of the study had the logic behind the ratings been sound. Despite this flaw in logic, methodology was sound enough to draw appropriate conclusions about this bedside assessment. Based on these results, it would appear that there would be little clinical utility for the implementation of this type of assessment.

In 1999, Logemann et al. published findings on a 28-item screening tool. This screening procedure was compared with VFS and contained screening items over five different categories: medical history, behavioral, gross motor, oral mechanism examination, and swallow trials (1ml of thin liquid, puree, and ¼ of a cookie). Ratings for PO trials were safe and unsafe. Summary variables were also used for statistical analysis. Results indicated that the best single predictor of the presence of aspiration was a throat clear or cough during swallowing trials with a sensitivity of 78% and a specificity of 58%. Logemann also provided statistics on oral and pharyngeal stage of the swallow as well as sensitivities and specificities for the majority of variables used. These can be found in Table 2. Additionally, statistics were provided for summary variables

i.e., unsafe on more than 8 of the 28 observations with a sensitivity of 62% and specificity of 65%.

Logemann thoughtfully and accurately discusses the dichotomy between a dysphagia evaluation and swallowing screen. The screen presented is thorough and considers all phases of the swallow, however, there is not a significant amount of compelling evidence for use of this screening tool. A sensitivity of 78% and a specificity of 58% leaves much to be desired in terms of confidently identifying those patients at risk for aspiration. Statistical measures are involved but too complicated to be used meaningfully. For example: logistic regression identified unsafe ranking on at least 2 of the 3 following variables: unsafe on 8 of the 28 observations, observation of swallow delay, and facial weakness. This resulted in a sensitivity of 71% and specificity of 73% in correctly classifying 72% of patients as either having or not having a pharyngeal delay. Even if the statistics were favorable, a practitioner in clinical practice would not have the time to unravel this statistical milieu to abstract clinically meaningful information from these results. While well thought out and thorough in methodology and statistical analysis, use of this screening procedure does not appear to be clinically favorable given other options in the literature.

Following her earlier investigation of the water swallow test, Daniels et al. (1998) published a study which considered utilization of a bolus swallow test for screening in acute stroke patients (n=55). This screen consisted of an oral mechanism examination and a clinical examination which involved PO trials with liquid (5ml, 10ml, 20ml), semisolid, and solid boluses. Swallow was assessed for oral transition, oral retention, initiation of laryngeal

excursion, voice quality after swallow, and spontaneous cough. Criteria for failure was cough or voice change after swallow. The clinical bedside examination was compared with VFSS as the objective criterion standard. Results indicate that sensitivity and specificity for determining aspiration was 69.6% and 84.4%, respectively. Logistic regression revealed that the combination of abnormal volitional cough and cough with swallow predicted aspiration with 78% accuracy. In keeping with the reporting of her prior study, Daniels reports that 90% of patients with aspiration presented with two or more clinical features, suggestive that two or more clinical features may be predictive of aspiration status.

Methodology of this study was sound, however it is unclear what information was gleaned from the more involved bedside evaluation given that the criteria for failure was cough and voice change post-swallow and does not appear to take performance on other measurements into account when administering the screen in its entirety. The authors did report adequate reliability data and they conducted a 3 month follow-up to determine incidence of aspiration pneumonia, a step that many studies fail to complete. Statistical analysis was detailed, but conclusions drawn from them are problematic. The authors reported that ability to predict aspiration was significantly improved by the presence of both abnormal volitional cough and cough after swallow. The 78% accuracy reported may be statistically significant, however, it is not particularly clinically helpful as one would be potentially unable to predict aspiration in 22% of cases. This is not acceptable from a clinical perspective. Additionally, the authors report but fail to comment on the sensitivity and specificity of this assessment tool in their conclusions.

These statistics demonstrated poor predictive value and should not be omitted from discussion or conclusions about this screening tool. Overall, other screening tools demonstrate better predictive value with much less detailed assessment tools.

Mann et al. (2000) presented a variation on a water swallow test which included thickened liquids. Components of this clinical screen include history, oral mechanism examination, trial swallows with thin liquids (5ml, 20ml), and thickened liquids, as needed. The results of this assessment were compared with VFSS in 128 acute stroke patients. Inter-rater reliability is provided for both the clinical screen and the VFSS. Results indicated that the sensitivity and specificity for identifying any swallow dysfunction or abnormality were 73% and 89% respectively. For the detection of aspiration alone, the sensitivity was 93% and the specificity was 63%.

Mann et al. concluded that the clinical bedside evaluation underestimates the frequency of swallowing abnormalities and overestimates the frequency of aspiration as compared with VFSS. Criticisms of the study include the timing of the swallowing screen and VFSS as well as the small volume of water used. The bedside assessment was conducted within 3 days of symptom onset and the VFSS was conducted within 10 days. This is too long a time period between testing in the acute neurological population as neurological status often changes rapidly in the acute care setting. As has been stated previously, silent aspiration is volume dependent (Leder et al., 2011), leaving open the question of whether the results of this assessment may have been different if larger volumes of water were used.

In 2001 McCullough et al. investigated a clinical bedside examination that consisted of a history, oral mechanism evaluation, voice, and trial swallows with thin, thick, puree, and solid consistencies in 5 cc boluses. Investigators examined delayed swallow, swallow duration, laryngeal elevation, and swallows per bolus in each of the 60 acute stroke patients and compared results to VFS. Inter- and intra-judge reliability were presented. Sensitivities and specificities were presented for each individual variable utilized in the bedside swallow assessment. These results can be found in Table 2.

When assessing data for clinical utility, McCullough and colleagues (2001) identified an arbitrary criterion for minimal acceptability values for sensitivity and specificity of 60%. Results indicated that no history items, oral motor tasks, or voice signs met even this minimal criterion. There were two signs on trial swallows that met criteria for sensitivity and specificity: spontaneous cough with swallow (sensitivity of 68% and specificity of 81%) and overall estimate of presence or absence of aspiration (sensitivity of 77% and specificity of 63%).

McCullough et al., drew reasonable conclusions about lack of clinical utility of the screen based on their data. It is again unfortunate that larger volumes of water were not trialed. Five milliliters is not a sufficient measure of swallow capability and again with this evaluation one is left wondering if the outcome of the study would have been different with the utilization of a larger volume of liquid.

Rosenbeck and colleagues (2004) presented data (n=60) on clinical evaluation consisting of the same evaluative measures as that of McCullough (2001). The only difference in methodology appears to be the addition of a 3-ounce water swallow that is reported in a table

within the text but is not discussed in the procedures. Rosenbeck also reports specificities and sensitivities of the individual evaluation measures. These can be found on Table 2. Overall rating of dysphagia based on trial swallows was found to have a sensitivity of 91% and a specificity of 47%. Additional results were calculated by combining signs and reported as likelihood ratios. There were four historical signs, i.e., incidence of pneumonia, nutrition status, presence of a feeding tube, need for suction, three oral motor signs, two speech signs, and two trial swallow signs. Examples of presentation of results were: if two out of four history signs are present the person is 12 times more likely to aspirate than someone without two out of four history signs or if dysarthria or dysphonia are present, the person is 1.3 times more likely to aspirate.

Methodology of this study is adequate, although clarification about the inclusion of the 3-ounce water swallow would have been advantageous in the interpretation of results. The calculation of likelihood ratios is interesting and gives the clinician a better sense of what result to expect, however, these statistics are limited in their practical clinical application.

Clinicians who use this tool should be aware that there is the potential to miss almost 10% of their patients based on sensitivity data, but it is unclear how one would clinically apply the information gleaned from the likelihood ratios that a patient is exhibiting for a particular grouping of signs.

Another frequently cited study is that of the Gugging Swallowing Screen (GUSS) as investigated by Trapl et al. (2007). In this study 50 acute stroke patients were assessed using a two part screen: saliva swallow, food bolus swallows of pudding, solids, and thin liquids

swallows (3ml, 5ml, 10ml, 20ml, and 50ml). Subjects were rated on cough, saliva management, drooling, voice changes, and laryngeal elevation. Results were compared with FEES. Results reported by the authors support the use of this study for use in identifying dysphagia and aspiration risk.

The use of an objective measure for validity, aspects of methodology, i.e., time between tests was less than 2 hours, and inter-rater reliability data were all relative strengths of this investigation. The use of nursing screeners is also a strength although training of nurses who conducted the screen was not discussed. The sample size was small prior to the questionable decision to further divide subject groups for external validation by the nurses, making it more difficult to draw confident conclusions about the data presented. Another aspect of methodology that is called into question was the ratings for the PO trials. The criteria for rating each consistency is not consistently clear. For those criteria provided, the measures should be called into question. For example, the semi-solid swallowing trial is supposed to be terminated if 1 of the 4 aspiration signs is present (deglutition, cough, drooling, and voice changes). Deglutition should not be an exclusionary criterion. Given these concerns, the reported sensitivity of 100% and specificity of 50% should be interpreted with caution.

In 2011 Schultheiss et al. investigated 62 subjects using the saliva swallow test (SST), a WST (5ml, 10ml, and 20ml) and a bolus swallow test (1/3 tsp., 1/2 tsp., and 1tsp of jelly) to assess presence of aspiration as compared to FEES. Results indicated the sensitivity of the water swallow test was 70.7% and the specificity was 82.5%. The water swallow test combined with

the bolus swallow test demonstrated a sensitivity of 76.2%. A combination of the bolus swallow test and the saliva swallow test yielded a sensitivity of 89.6% and a specificity of 72.7%.

None of the three subtests demonstrated acceptable statistical measures for independent use. The authors went to great lengths to combine tests in an effort to report improved statistical measures seemingly in an effort to demonstrate the effectiveness of this screening tool.

Additionally, when the results of the clinical swallow were compared with those of FEES, there were statistically significant differences for the WST and the BST indicating that neither test was accurate when compared with the objective criterion standard. Therefore, the authors' statement that the combination of the BST and the SST resulted in a sensitive clinical instrument for detecting aspiration should be called into question.

The Brief Bedside Dysphagia Screening Test was developed by Mandysova et al. and published in 2011. This screening tool was developed for use by nurses and includes 32 measures from portions of the Massey Bedside Swallowing Screen and the Gugging Swallowing Screen. Results of the screen (n=108) were compared to FEES. Findings indicate an overall sensitivity of 87% and specificity of 30.4% in the ability of the screen to detect aspiration risk. Sensitivity for the neurological population was found to be 95% and for the ear nose and throat population to be 60%. The authors appropriately recognized that additional investigation was needed to generalize these results to other populations given the large range in sensitivity results.

Although this study advocates for nursing screeners, results do not support widespread use of this screening tool across the general hospital population. Of concern is that some test items have no reported correlation with swallowing, i.e., bleeding and shoulder shrug and the

fact that test items were excluded if the patient was unable to follow directions resulting in data missing from the analysis. Overall, while basically methodologically sound, there are other screening tools with more compelling data for use in the general hospital population.

Other Swallowing Screens

Hammond et al. (2009) used a single clinical measure of voluntary cough to determine if quantification of this measure could be used as an indicator for aspiration risk. Investigators used measures associated with voluntary cough, such as inspiratory peak flow, inspiration volume, duration of glottis closure, expulsive peak flow and cough volume acceleration and compared results to FEES or VFSS. Results demonstrated that peak flow of the inspiratory phase, sound pressure level, peak flow of the expulsive phase, expulsive phase rise time, and cough volume acceleration were significantly impaired in severe aspirators as compared with non-aspirators. While findings are interesting, the use of a non-swallowing variable, the need for specialized equipment and complex analysis calls into question the clinical utility of this screening tool as a practical option for identifying those patients at risk for aspiration.

Ramsey et al. (2006) also used a single measure to detect aspiration. This study of 189 stroke patients investigated the ability of pulse oximetry to determine the presence of aspiration risk. The authors conducted what was described as a modified Bedside Swallow Assessment (mBSA). During VFSS, 5mls and 75mls of contrast were administered and aspiration status determined. Pulse oximetry was measured during the assessment. Findings indicated that there was a poor association between oxygen desaturation and aspiration status. It was demonstrated that using VFSS as the criterion standard, sensitivity and specificity for detecting aspiration were 47% and 72% respectively for mBSA, 33% and 62% for desaturation greater than 2%, and 13% and 95% for desaturation greater than 5%. The use of objective measures, simultaneous bedside swallow during VFSS and pulse oximetry, and solid statistical analysis allow clinicians to

interpret these poor results with confidence and conclude that pulse oximetry is not a good indicator of aspiration risk.

The remaining studies to be reviewed demonstrate either lack of adequate statistics, significantly flawed or poorly defined methodology, or report mis-interpreted or uninterpretable results.

In an early study (1983), Linden et al. reported results based on the use of a clinical swallowing screen versus a VFS in predicting laryngeal penetration. Reported statistics were ratios only perhaps secondary to the small sample size (n=15). Linden et al. found that 11/15 patients demonstrated laryngeal penetration during swallowing. They found a high incidence of impaired pharyngeal gag and wet vocal quality in this group. Cough was found to be an unreliable indicator of laryngeal penetration. These results were interesting but not clinically significant nor generalizable due to inadequate statistics.

A study by Clave et al (2008) yielded a similar conclusion to that of Ramsey (2006) despite a larger sample size (n=85). Clave and colleagues looked at the Volume-Viscosity Test (V-VST) and VFSS. The V-VST used liquid, nectar, and honey consistencies in different aliquots (5ml, 10ml, 20ml) to assess swallowing. The clinical measures identified were cough, change in vocal quality, and pulse oximetry. Results indicated that the sensitivity for detecting impaired swallow was 88% and the specificity was 64%. For the detection of aspiration, sensitivity was 100% and specificity was 28.8%. Conflicting statements were reported regarding the pulse oximetry findings and make interpretation impossible. Specifically, the authors first reported that there was no difference in basal oxygen saturation among patients with

and without impaired swallow on VFSS. However, they proceed to incorrectly state that pulse oximetry was a viable measure for determining aspiration risk. As pulse oximetry was one of only three measures used to determine swallowing success, questions surrounding contradictory results leave one unable to interpret the remaining statistics.

Nishiwaki et al. (2005) used a swallowing screen consisting of a saliva swallow test, a 30-ml water swallow test that was modified secondary to concern for aspiration with the previously published 90ml water test, and various oral mechanism factors. The results of this evaluation were compared with VFS and indicated that the only variable significantly associated with aspiration on VFSS was cough/voice change in WST with a sensitivity of 72% and a specificity of 67%. The authors erroneously concluded that this swallow screen should be used to detect aspiration. Methodology was flawed in that VFSS was performed much later, i.e., within 7 days of the bedside evaluation. This latency is unacceptable in acute stroke patients. Given concerns about methodology, erroneous conclusions, and the lack of significance found with the remaining measures of the screening tool, the use of this tool in clinical practice would not be recommended.

Smith et al. (2000) reported results of a bedside swallowing evaluation combined with oxygen desaturation in 53 stroke patients. Results were compared with VFSS. Findings indicated that the bedside assessment and changes in oxygen desaturation had a 73% sensitivity and 76% specificity for predicting aspiration risk. Oxygen desaturation of greater than 2% alone had a sensitivity of 87% and a specificity of 39%. The flaw in this study was extremely poorly

defined methodology for the bedside assessment. Without this information it is impossible for the clinician to put this research into clinical practice.

Weinhardt et al. (2008) proposed to validate a screening tool using both nurses and speech-language pathologists that involved inclusionary criteria of a minimum NIHSS score and a bolus swallow test. The bolus swallow test consisted of 1 teaspoon of lemon ice, 1 teaspoon of applesauce, and 1 teaspoon of water. Criteria for passing were no cough or change in vocal quality with PO trials. The fatal flaw with this study was that there was no objective measure of swallowing used to detect the presence or absence of aspiration. Without objective corroboration, validity for this tool cannot be determined. The only useful information remaining was reliability data, i.e. 94% agreement between nurses and speech-language pathologists. While efforts to validate a screen using nurses should be encouraged, the current methodology and the rationale are inadequate to answer the proposed research question.

The authors provide erroneous rationales for their chosen methodology. The authors stated, "Because even a normal swallower aspirates liquids more frequently, water is not the first test item presented to the patient." (Weinhardt et al., 2008 p. 249). They reported that in order to decrease the risk of aspiration, they use lemon ice as a stimulus, claiming that, "...even when melted completely, lemon ice remains thicker than water thus reducing aspiration potential" (Weinhardt et al., 2008 p. 249). Neither of these statements are based on objective viscosity measures and lend no support to their choice in methodology which remains limited in both number and volume of boluses delivered. Flawed methodology combined with the lack of objective measure by which validity must be determined make the results of this study unusable.

Bravada and colleagues (2009) reported on the use of a nursing screen that involved administering a questionnaire to 101 stroke patients which is then compared to NIHSS severity. The questionnaire included information about orientation, command following, facial weakness, saliva management, dysphonia, weak cough, dysarthria and voice change after swallow. If any one of these measures was found to be present it was a criterion for failure and rendered a recommendation of nil per os. The nurses did not administer oral trials in this screening tool. Results indicated that the screening tool had a 29% sensitivity and 84% specificity for detecting dysphagia. The NIHSS, when compared to a speech-language pathology consultation, had a sensitivity of 79% and a specificity of 68%. It was concluded that the NIHSS had better test characteristics than their screening tool in predicting dysphagia.

There are several fatal flaws in this study. The first is the lack of use of an objective swallowing measure. Additionally, nurses do not administer oral trials yet one of the screening items is voice change after swallow. It seems that asking the patient this question would be a potentially inefficient measure of this indicator. The use of the NIHSS was not well specified other than to say patients were classified as low and high. Furthermore, speech-language pathology consultation was ill defined, making it impossible to know what measures were being compared with the NIHSS. Also, not all assessments were completed in the same time frame. Although this was mentioned in the discussion, it was not discussed in the methodology and therefore one cannot determine the impact, if any, on outcome. The lack of sound methodology makes use of this assessment tool clinically ill-advised.

Westergren et al. (1999) reported use of a nursing assessment of dysphagia in 160 stroke patients. This assessment considered either observed or reported swallowing deficits, trial swallows with 30mls of water and sour milk. Criteria for failure of the screen were any one risk question, cough, dribble, or observation that a patient exhibited for an effort to swallow. These results were compared to published rating scales for consciousness and activities of daily living. Authors reported that 77% of patients with dysphagia were identified through this screening method. From this data they concluded that this screening is useful because “most” patients with dysphagia can be identified via this screen alone.

Like many of the nursing screening assessments, this study has poor methodology and lacks statistical analysis. Importantly, no objective measure was used to determine the presence or absence of dysphagia. Reporting that a percentage of patients were identified on the screening tool is not adequate statistical evidence to support use of a screen. Even if statistically sound, 77% is not large enough to make a compelling argument in favor of using this screen. Several methodological issues should also be addressed. Two of the failure criteria are problematic. Rating a person for failure if they “dribble” is unsubstantiated in the literature. The second rating is even more nebulous as it is unclear how one objectively determines that a patient has to “make an effort” to swallow. Less concerning but necessary to point out is the fact that the authors justify the use of sour milk not only because it is thicker and less likely to cause aspiration but also because the sour taste purportedly stimulates oral perception and thereby facilitates the swallow. Even if there were evidence to support this, it is unclear how this translates into functional PO intake. The statistical inadequacies and questionable methodology cause this screen to lack clinical practicality or desirability.

Nursing Administration of Swallow Screening

Swallow screening is a critical component of dysphagia management in the acutely hospitalized patient. The literature provides a plethora of swallow screening options for the speech-language pathologist. The discerning clinician, therefore, must make an informed decision about what screen is desired for use. A successful swallow screen should be simple to administer, cross-disciplinary, cost effective, acceptable to patients, and able to identify the desired target in question (Cochrane & Holland, 1971). The clinician must first consider the quality of the research, ensuring adequate statistical evidence for clinical use, and methodology that is well substantiated by evidence from the literature. Other factors that must be taken into account include ease of administration, time of administration, and necessary equipment to carry out the screen.

Unfortunately, swallow screening practice varies widely and much is unsubstantiated by current research. McCullough et al. (1999) reported that only 56% of methods utilized for assessment had support from the literature. Review of this body of literature yields many studies that contain homogeneous samples, flawed methodology, lengthy and unsubstantiated screening procedures, and erroneous conclusions.

The 3-ounce water swallow screen (Suiter & Leder, 2008) has emerged from the literature as a swallow screening tool which meets the criteria for an effective swallow screen. Specifically, it used a large and heterogeneous patient population sample, was validated against the objective criterion standard of FEES and VFSS, utilized methodologically sound administration techniques and is simple and quick to administer. The 3-ounce water swallow

protocol is supported in an evidence-based manner in this body of literature and should be considered by the clinician seeking an evidence-based practice model of care.

While swallow screening by the speech-language pathologist is accepted practice, the necessity for recognition of swallowing difficulties early in a patient's hospital course supports the need for an evidence-based and registered nurse administered swallow screen (Bours et al., 2008). Unfortunately, the body of literature to date related to nursing administration of swallowing screening lacks sound methodology, rarely demonstrates the use of a criterion standard for comparison with the screen, and has inadequate statistical support to gain wide acceptance.

There is no publication to date that investigates the use of the validated 3-ounce water swallow protocol by registered nurses. In a systematic review, Hines et al. (2011) confirmed that nurses are in an advantageous position to conduct swallow screens. However, due to significant variability in methodology, populations, and interventions of the swallow screens reviewed, no meta-analysis of RN administered swallow screening was possible. The chief criticism of the nurse administered swallow screen literature is the lack of a criterion standard for identification of aspiration risk (i.e. use of VFSS or FEES). Without this critical component, the effectiveness and reliability of a given screen cannot be determined.

In 2008, Weinhardt and colleagues attempted to determine the reliability of nursing's ability to administer a swallow screen as compared to that of the speech language pathologist. While 94% agreement between RNs and SLPs was reported, no criterion standard was used. The screen was significantly lacking from a methodological perspective, consisting of 5 ml boluses of

lemon ice, applesauce, and water and measured success via observation of swallow, cough, and vocal quality. While the reported agreement between RN and SLP was favorable, the lack of a criterion standard and questionable methodology rendered this screen a non-viable option for evidence-based practice.

In another effort to recognize the utility of nurse administered swallow screening, Westergren and colleagues (1999) reported results of an equally questionable screen also with no criterion standard. In this study, sour milk was used as it was claimed that it was thicker than milk and the taste should stimulate oral perception. Criteria for failure of this screen were failure on any risk question or cough or dribble with trial swallows. Unfortunately, with no criterion standard, the results of this study are not able to be interpreted with respect to utility of the screen in determining aspiration risk.

There are several additional studies that lack a criterion standard. Bravata and colleagues (2009) conducted a retrospective study of a nursing administered screening tool versus the National Institute of Health Stroke Scale (NIHSS). In addition to no criterion standard, the screen used in this study was void of per os (PO) trials altogether. Bernard and colleagues (2011) reported on an emergency room nursing screen involving one bite of applesauce, one sip of water, and consecutive sips of water. This study lacked both a criterion standard and data. While advocating for nurse administered screens, the study was purely descriptive and did not contribute to the needed evidence-based practice literature. Perry and colleagues (2001) reported on the use of the Standardized Swallowing Assessment (SSA). While sensitivity and specificity

for this tool were reported (sensitivity of 97% and specificity of 90%), results must be interpreted with extreme caution due to the fact that the criterion standard used for this study was “clinical judgment of dysphagia” as determined via chart review. Daniels et al. (2013) reported on the implementation of a water swallowing screening of stroke patients in the emergency department. Results supported such a screen be used by RNs, that improved screening was noted with implementation of RN training, and that tailored education of RNs improved sustainability of adherence to the screening protocol. Unfortunately, the lack of a criterion standard leaves the clinician without information about the reliability of the screen itself.

Steele and colleagues (2011) reported results of a nursing administered swallow screen that used VFSS as the criterion standard. The study was sound from a methodological and statistical perspective. However the screen, which consisted of both non-swallow items and cup drinking, did not support the use of nursing administered swallow screening. Of note, the sensitivities and specificities of the screen were poor for both SLPs (sensitivity of 71% and specificity of 13%) and RNs (sensitivity of 58% and specificity of 47%) and were not supportive of the use of the screen in general, irrespective of who was administering it.

This body of evidence is lacking studies that use sound methodology, have an objective criterion standard, and demonstrate adequate statistical support for both the presented screen itself and for nursing administration of the screen. Without such support from the literature, moving toward a change in best practice for nurse administered swallow screens will be problematic.

Conclusion

The literature provides a plethora of swallow screening options. The discerning clinician must make an informed decision about what screen is desired for use. Considerations are multifactorial. The clinician must first consider the quality of the research, ensuring adequate statistical evidence for clinical use, and methodology that is well substantiated by evidence from the literature. Other factors that must be taken into account include ease of administration, time of administration, and any necessary equipment to carry out the screen. The screen should provide the clinician with a clear pass/fail choice and explicitly state what action should be taken after completion of the screen.

There is no consensus, to date, on swallowing screening. Although there is not one absolutely correct answer to this question, this critical review has presented one swallow screen that appears to fit most of the desired criteria. The 3-ounce water swallow challenge protocol not only has one of the highest sensitivities reported, it is also validated against the criterion standards of FEES and VFSS and incorporates a large and heterogeneous population sample (n=3,000) that far exceeds the breadth of subjects that other studies have used. While there are other screening tools with reasonable sensitivities and good methodology, when both the time needed to administer and ease of administration are taken into consideration, this swallow screen is an obviously superior choice.

Perhaps most importantly, there is strong evidence that if a patient passes the 3-ounce screen, the clinician can confidently recommend an oral diet without need for further objective testing. This data is in support of a significant clinical change in the way clinicians have utilized

screening tools in the past. When taken collectively, the body of literature that is published about the use of the 3-ounce screen, the clinical considerations that must be taken into account, and the ability to recommend oral alimentation across heterogeneous diagnoses and at critical levels of patient care, demonstrate a programmatic line of research that is unmatched in other screening tools.

CHAPTER III

Methods and Procedures

Participants

This study was approved by the Institutional Review Board, University of Connecticut and the Human Investigation Committee, Yale School of Medicine, New Haven, CT. Two groups of participants were recruited for this study: patient participants (n= 101) and nursing participants (n=52). Inclusion criteria for all registered nurses (RNs) was assignment to care of the patient participant and completion of a web-based training module specific to administration, interpretation, and scoring of the 3-ounce water swallow challenge protocol. In order to reduce selection bias, RN participation was not based on experience or degree status beyond RN. Inclusion criteria for all patient participants was the ability to demonstrate adequate cognitive abilities (Leder et al., 2009) and oral motor functioning (Leder et al., 2012) to participate in the 3-ounce water swallow challenge protocol (Suiter & Leder, 2008).

Patient exclusion criteria included the inability to remain alert for screening, a current order for a modified diet due to a pre-existing dysphagia, presence of a gastrostomy feeding tube, presence of a tracheotomy tube, head-of-bed restriction less than 30 degrees, or nil per os by physician order (Appendix One).

Procedures

Web-Based Training Module

In advance of this investigation, a hospital-wide swallow screening training module was developed, trialed, and approved by the Nursing Educational Leadership Council. All RNs were required to complete a web-based training module incorporated into the hospital's Healthstream

Learning Center. This module included instruction on administration of the 3-ounce water swallow challenge. Specific instructions on how to administer the 3-ounce water swallow protocol was provided (Appendix Two). Pass/Fail criteria as well as information on interpretation of results and scoring were delineated (Appendix Three). RNs were instructed to collaborate with the referring physician, mid-level provider, or speech-language pathologist when determining diet recommendations based on dentition status. In addition to the instructions on how to administer the 3-ounce water swallow protocol, training included a basic anatomic and physiologic overview of swallowing function, criteria to identify patients at-risk for aspiration, and a post-module quiz. Following completion of web-based training, participant RNs were required to attend an in-service conducted by a clinical nurse educator that reviewed the written policy and documentation requirements of the screening protocol. RNs were required to show functional competency via return demonstration of appropriate administration of the 3-ounce water swallow challenge protocol.

3-Ounce Water Swallow Challenge Protocol

Step 1

The 3-ounce water swallow challenge protocol (Suiter & Leder, 2008) was first administered to each patient participant by one of two speech-language pathologists both with > 10 years of experience administering the protocol. The protocol consists of three parts: (1) a brief cognitive screen (Leder et al., 2009), (2) an oral mechanism examination (Leder et al., 2012) and (3) a 3-ounce water swallow challenge (Suiter & Leder, 2008).

A successful 3-ounce water swallow challenge requires drinking 3-ounces of water via cup or straw, uninterrupted, and without overt signs of aspiration i.e., coughing or choking, either during or immediately upon completion. Criteria for failure are inability to drink the entire amount, interrupted drinking, or overt signs of aspiration. Although not part of this study, it should be noted that if passed, an oral diet was recommended based on results of the cognitive screen, oral mechanism examination, and dentition status. If failed, an objective swallow evaluation, either fiberoptic or fluoroscopic, was performed prior to recommending oral alimentation.

Step 2

Within 1 hour of the initial testing completed by the speech-language pathologist, an RN administered the 3-ounce water swallow challenge protocol to the same patient and independently recorded results and diet recommendations. The RN was blinded to both the purpose of the study and to the results of the speech-language pathologist's initial 3-ounce challenge. Out of view but simultaneously with the RN administered water swallow challenge, a speech-language pathologist re-rated both the patient's 3-ounce water swallow challenge for comparison with their initial results and determined the accuracy of the RN administered challenge protocol, interpretation, and diet recommendations.

Data Analysis

Inter and intra-rater agreement

Data was collected and recorded on the data collection sheet in Appendix Four. Results of nursing administered as well as the speech language pathologist administered 3-ounce water swallow challenge protocols were reported. In addition, the performance of the RN's ability to perform the swallow screen was recorded by the SLP as pass/fail, that is, if the nurse performed the screen and interpreted the results correctly.

Intra-rater agreement for the two speech-language pathologists was determined by comparing their initial respective initial 3-ounce challenge protocol results with results of the observed challenge protocol administered by the RNs. Inter-rater agreement between the speech-language pathologists and RNs was determined by comparing results of the observed challenge as administered by the RN and the initial independent ratings of the challenge. Percent agreement between the nurses and speech-language pathologists was reported. Inter-judge and intra-judge reliability was determined using Cohen's kappa. The kappa statistic is considered to be a superior measure of agreement as it is the proportion of agreement between raters after chance agreement has been excluded (Wood, 2007).

Time post web-based training

In order to determine if time post- web-based training impacted nurse's ability to accurately administer and score the 3-ounce water swallow challenge, data on the elapsed time post training was calculated by averaging the amount of time that passed between when the nurse completed the web-based training module and when the study was conducted.

Oral Care

It has been shown that there is a relationship between diligent oral care and reduction in incidence of aspiration pneumonia (Abe et al., 2006). As routine oral care is a part of standard clinical nursing practice, an oral care protocol was not added to the Procedures, however, data were collected on the frequency of oral care as dictated by the guidelines of the medical units on which patient participants were admitted.

CHAPTER IV

Results

Intra- and inter-rater agreements are shown on Table 3. Intra-rater agreement for the two speech-language pathologists was determined by comparing their respective initial 3-ounce challenge protocol results (one speech-language pathologist performed 26 screens, and one performed 75 screens) with results of the observed challenge protocol administered by the RNs. Inter-rater agreement between the two speech-language pathologists and RNs was determined by comparing results of the observed challenge with results of the RN administration and independent ratings of the challenge (n = 101). Both intra- and inter-rater agreements for the two speech-language pathologists were 100%. Inter-rater agreement between the RNs and speech-language pathologists was 98.01%. The discrepancy in percent agreement was attributed to the same RN who incorrectly passed two patients when interrupted drinking occurred, while the speech-language pathologists correctly rated this behavior as a fail.

Table 4 shows nurse pass/fail ratings by participant diagnosis. No diagnostic category was associated with increased difficulty during administration and interpretation of the 3-ounce water swallow screen.

Cohen's kappa was determined to be 1.0 for intra-rater agreement and 0.95 for inter-rater agreement. These values are judged to be excellent agreement. Kappa values range from -1.0 to

+1.0. A kappa of 1.0 indicates perfect agreement while a kappa of -1.0 indicates perfect disagreement. A kappa value of 0 is indicative of random agreement and no relationship between the ratings (Tables 5 A, B and C).

In addition to inter and intra-rater agreements, average elapsed time between RN web-based training and administration of nursing challenge was calculated. It was found that RNs who participated in the challenge were, on average, 26.8 (range 1month -33months) months post web-based training.

It was found that standard nursing care dictated that oral care be performed once per shift. This was consistent across both the intensive care units and on medical floors for patients that are would be candidates for the 3-ounce water swallow challenge, i.e., patients that were not intubated.

CHAPTER V

Discussion

Data presented in this study support nursing's ability to effectively perform the 3-ounce water swallow protocol to detect aspiration risk. These findings are a novel addition to the current literature which is lacking evidence-based support for nursing administration and interpretation of swallowing screening. Given the importance of accurate, reliable, and timely determination of aspiration risk, use of an evidenced-based swallowing screen is of critical importance to determine when to safely initiate oral alimentation and oral medications with the goal of minimizing deleterious sequelae of prandial aspiration.

Identifying which health-care professionals should administer and interpret a swallow screen is of critical importance. Nurses are a logical choice for front-line administration of a validated swallow screening tool (Hines et al., 2001) as they have the most direct, frequent, and continuous contact with patients and many are already performing some type of screen in a non-validated method. Findings of Cohen's kappa of 0.95 support nurses' ability to effectively perform the swallow screen in both neurologically impaired patients as well as the general hospital population. One goal of the 3-ounce water swallow screen is timely initiation of oral intake. In support of this goal, Leder et al. (2012) provided novel evidence that once passed, a diet order can be placed without need for further objective testing. For those patients who pass the screen, nursing administration thus allows for potential near immediate initiation of oral alimentation and medications without delays resulting from the need for an ancillary consult service to conduct the screen.

Support for the use of nursing administration of the 3-ounce water swallow challenge is likely to lead to more widespread use of the screen. It would be prudent to screen at defined points of time in the hospital admission process. For example, in the emergency department, upon arrival in the unit, and at any time a change in medical status is observed. This would allow adverse preventable events, for example aspiration pneumonia, to be potentially avoided due to the ability to monitor aspiration status at any point during the patient's hospitalization.

It is recommended that if a 3-ounce water swallow challenge protocol is failed, the protocol should be repeated within 24 hours because acute care patients often demonstrate rapid improvement in swallowing function (Leder, 1998). If subsequently passed, diet and medication recommendations can then be made confidently without further dysphagia testing. If, however, a second swallow protocol is failed, objective testing, either endoscopic or fluoroscopic, should be performed to determine dysphagia status and make recommendations for safe oral alimentation (Suiter & Leder, 2008).

It should be noted that patients who require a tracheotomy tube for airway maintenance, ongoing mechanical ventilation and pulmonary toilet should not be tested with a 3-ounce water swallow challenge protocol. Silent aspiration occurs more frequently due to laryngeal desensitization from chronic aspiration of secretions (Link et al., 2000, Donzelli et al., 2003). Although a tracheotomy is not causal for aspiration (Leder & Ross, 2000, 2010), if aspiration occurs, it is often silent, that is, without overt signs such as coughing, leading to higher false negative rates using a 3-ounce swallow challenge. Therefore, objective endoscopic or fluoroscopic testing is recommended for this patient population.

Medication management is another important factor when considering the impact the 3-ounce swallow screen has on clinical practice. One goal of implementation of the screen would be to safely administer oral medications in a more timely fashion. If the screen is failed, however, patients should remain nil per os until an objective assessment can be completed. When any swallowing problem is suspected, it is very important to screen for aspiration risk prior to oral ingestion of medications (Cichero et al., 2009). A 'nil-per-os order except medications' in the dysphagic patient is imprudent and potentially dangerous (Leder & Lerner, 2012). Difficulty taking medications is common in patients with dysphagia, and formulation was reported to be an important variable (Kelly et al., 2010). Once the dysphagia is objectively assessed, changes in formulation or compensatory strategies to improve swallowing success can be implemented. Such interventions may include varying pill size, shape and texture; changing to either liquid or chewable formulations; use of viscosities other than water, for example nectar-like, honey-like and pudding consistencies (Kelly et al., 2010).

There is some evidence to suggest that oral care plays a role in preventing aspiration pneumonia (Abe et al., 2006) and as such, data on oral care were collected for the patient units that participated in the study. While there is no published standard dictating parameters for such care, from a nursing practice perspective, once a shift oral care was deemed to be adequate to initiate the 3-ounce water swallow screen without need to add a formal oral care component to the screen itself.

A critical caveat regarding this study involves the training module that was in place prior to the start of the study. A robust training mechanism was a necessary component to the success of the ability of nurses to successfully administer and interpret the 3-ounce water swallow screen. It is interesting to note that although the initial goal of the training was that it be reviewed annually, the average time between completion of the web-based training program and demonstration of their knowledge via this study was 26.8 months. While this finding may allow needed action to be taken regarding reinforcement of the review of the web-based training, it is also evidence that despite the time lapse in training, nurses were able to successfully carry out the administration and interpretation of the screen. The most likely contributing factor to their success is the widespread use of this screen within this institution as it has become part of their standard clinical practice. It should be noted that due to this solid integration into clinical practice at the time of the study, these outcomes may not be typical in all hospital settings.

Limitations

Limitations of this study include the fact that although nurses were blinded to the specific purpose of the study, they were chosen by the SLP to perform the 3-ounce challenge with their patients. While every effort was made by the observing speech-language pathologist to remain out of view when observing the administration of the screen, it could be argued that the knowledge of the need to perform the screen itself could inadvertently cause the nurse to perform the screen differently, to wit, in a more vigilant fashion. Unfortunately, there was no reasonable way to avoid this request on the part of the nurse as observing the protocol as part of daily practice would be extremely difficult from a data collection perspective.

Additionally, it would have been optimal from an intra-rater reliability perspective to have the speech-language pathologists blinded to the initial results of their own screen. Unfortunately, this would have made research design difficult from both a personnel staffing and a data collection perspective as it would have required two different speech-language pathologists to be available at each point of patient contact.

Another potential limitation concerns the nursing training module. While the training module was developed to be similar to the training for other hospital based protocols, it could be argued that other institutions may have difficulty replicating the precise training that was undertaken given the potential for the variable nature of some aspects of the training program. For example, the appendices found in this dissertation could be easily reported to interested recipients, but for replication purposes, the web-based training module is a rather extensive document and unlikely to be shared or published in its entirety. Likewise, the in-servicing that the nurses experience could potentially vary based on the presenting clinician. Also, this study did not address the minimum amount of training needed or if variation in delivery of training material could have had an impact on nurse's ability to perform the screen.

Lastly, and not unique to this study, is the caveat of swallow stability over time. As with any swallow test, this screen provides only a snapshot-in-time of a patient's swallowing abilities and cannot guarantee continued swallowing success in the future. This is an inherent limitation of any swallow screen or dysphagia evaluation. Therefore, caregivers must remain vigilant to signs of aspiration risk, e.g., coughing at meal-times, febrile status, or signs/symptoms of upper respiratory infection, and use this information to recommend a formal swallowing evaluation.

Conclusions

The results of this study confirm both the reliability and accuracy of registered nursing administration of the 3-ounce water swallow challenge protocol. Correct use of this screening tool by nurses allows for early identification of potential swallowing problems. Passing results in timely ingestion of food, fluids, and medication. Failing results in nil-per-os status and triggers a referral to speech-language pathology for further evaluation.

Preventing aspiration of food, fluids, and medications as a cause of prandial pulmonary infection is a key goal in the care of hospitalized patients. The results of the present study support the use of a RN-administered 3-ounce water swallow challenge screening protocol. The protocol was shown to be both a highly accurate and reliable tool that can be used for early identification of aspiration risk with the majority of hospitalized patients.

APPENDIX ONE: 3-Ounce Water Swallow Challenge Screening Protocol

Step 1: Exclusion Criteria

Swallow Screening Protocol Deferred due to NO risk factors for aspiration.

Any YES answer to the following risk factors will also defer administration of protocol:

Yes No

 Unable to remain alert for testing

 Eating a modified diet (thickened liquids) due to pre-existing dysphagia

 Existing enteral tube feeding via stomach or nose

 Head-of-bed restrictions < 30°

 Tracheotomy tube present

 Nil per os by physician order

If the patient's clinical status changes resulting in a new risk for aspiration, the protocol must be re-administered before oral alimentation or medications are ordered.

APPENDIX TWO: 3-Ounce Water Swallow Challenge Screening Protocol**Step 2: Administration Instructions**

If patient is deemed at-risk for aspiration and all exclusion criteria in Step 1 are checked “NO”, perform the 3-ounce water swallow challenge protocol.

- Sit patient upright at 80-90° (or as high as tolerated > 30°)
- Ask patient to drink the entire 3-ounces (90cc) of water from a cup or with a straw, in sequential swallows, and slow and steady but without stopping (Note: Cup or straw can be held by RN or patient)
- Assess patient for coughing or choking during or immediately after completion of drinking
- Brief Cognitive Screen: ___ What is your name?
___ Where are you right now?
___ What year is it?
- Oral-Mechanism Examination: ___ Labial closure
___ Lingual range of motion
___ Facial symmetry (smile/pucker)

APPENDIX THREE: 3-Ounce Water Swallow Challenge Screening Protocol**Step 3: Pass/Fail Criteria****Results and Recommendations**

- **PASS:** Complete and uninterrupted drinking of all 3-ounces of water without overt signs of aspiration, i.e., coughing or choking, either during or immediately after completion.
- If patient passes, collaborate with MD/PA/LIP to order appropriate oral diet. If dentate, order a soft solid consistency or regular consistency diet. If edentulous, order a liquid and puree diet. Consult with speech-language pathologist for other diet modifications.
- **FAIL:** Inability to drink the entire 3-ounces in sequential swallows due to stopping/starting or patient exhibits overt signs of aspiration, i.e., coughing or choking, either during or immediately after completion.
- If patient fails, keep nil per os (including medications) and request the MD/LIP to order a consult for an objective swallowing evaluation by speech-language pathology.
 - Re-administer the protocol in 24 hours if patient shows clinical improvement.

APPENDIX FOUR: DATA COLLECTION SHEET

RN Screening: 3-Ounce Water Swallow Challenge

Subject #

Completed Web Based Training: Yes No

Performed Screening Correctly: Yes No

3-ounce Water Swallow Challenge: Yes

No – reason:

Capable: Yes No

Drank: Complete Partial

Results: Passed/no response Y N

Cough Y N

Delayed cough Y N

Wet voice Y N

Throat clear Y N

Rec: If no-----NPO and refer to speech language pathology

If yes -----PO

Dentate: Regular soft ground/chopped

Edentulous: Puree clear liquids

TABLE 1: SWALLOW SCREENS

Study	Population	n	Criterion Standard	Screening Modality	Sensitivity%/Specificity%	Aspiration Risk/Dysphagia	Other Statistics	Reliability
Linden et al (1983)	H	15	VFSS	CBE: subj, OME, PO trials (liquids, semisolids)	N/A	N/A	ratios - INADEQUATE	no
Splaingard et al(1988)	H	107	VFSS	CBE: hx, subj, OME, PO trials(liquid:sips, puree/ solids: tsp)	omitted /91%	A	PPV 0.75, NPV 0.70	no
DePippo et al(1992)	stroke	44	VFSS	WST (3 ounce)	76%/59% (wet voice)94%/26% (cough)94%/30% (cough/vocal qual)	AA >10% bolus - thick liq	N/A	no
DePippo et al(1994)	stroke	139	No	BDST (screening questions, WST (3 ounce))	N/A	pna, UAO death	Relative Risk	no
Gottlieb et al (1996)	stroke	180	No	50 ml WST (non-sequent), OME, cough, voice	N/A	pna	Relative Risk	no
Daniels et al (1997)	stroke	59	VFSS	SUBJ*, WST (5,10,20 ml)	92.3%/66.7% (2/6 clinical features)	Dysphagia severity	N/A	Yes (inter, intra)
Daniels et al (1998)	stroke	55	VFSS	OME, BST (semisolid, liquid 5, 10, 20 ml, solid)	69.9%/84.4%	AR	1. Logistic regression2. 90% of pts w/asp 2/6 clin. features	Yes (inter, intra)
Logemann et al (1999)	H	200	MBS	HX, SUBJ, BST (puree 1ml, thin 1ml, solid)	78%/58% (throat clear/cough)72%/67% (decr. laryng. elevation)	ARD	N/A	blinded no reliability
Westergren et al (1999)	stroke	160	No	Quest, ADL rating, water & sour milk (30 ml)	N/A	D	Chi square Fisher's exact test	Yes (inter, limited)
Daniels et al (2000)	stroke	56	VFSS	SUBJ, WST (5,10,20 ml)	N/A	AR/dys.severity	chi square (2 or more)x2=19.895, p = .00001	no
Teramoto et al (2000)	H w/asp pna	26	NO	SPT	76.4-100%/83.8-100% (SPT 0.4/2ml)72%/70.3% (WST)	asp. Pna	N/A	no
Mann et al (2000)	stroke	128	VFSS	CBE: hx, OME, WST (5, 20 ml), thickened prn	73%/89%/93%/63%	DA	N/A	Yes (inter)
Smith et al (2000)	stroke	53	VFSS	CBE: poorly defined	73%/76% (CBE & 02 desat)87%/39% (02 desat > 2%)	A	N/A	blinded no reliability
Lim et al (2000)	stroke	50	FEES	50 ml WST (10 ml aliquots) and pulse ox	76.9/83.3 (02 desat > 2%)84.6/75(50ml WST)100/70.8 (combined)	A	N/A	no
Hammond et al (2001)	stroke	43	VFSS or FEES	voluntary cough	N/A	AR	Logistic regression	blinded no reliability
McCullough et al (2001)	stroke	60	VFSS	CBE: hx, subj, OME, PO trials (thin/thick liquid:5ml, solid)	68%/82% (cough)77%/63% (overall asp rating)	A	N/A	yes
Leder et al (2002)	stroke	49	FEES	CBE: subj (6 clinical identifiers)*, PO (liquid:sips)	86%/30%	AR	N/A	blinded no reliability
Massey et al (2002)	H	25	VFSS	subj, OME, WST (5, 60ml)	N/A	D	Chi square	Yes (inter)
Wu et al (2004)	stroke	59	VFSS	WST (100ml)	85.5%/50% (swallow speed)47.8%/91.7% (choking/wet voice)	D A/P	N/A	no
Rosenbeck et al (2004)	stroke	60	VFSS	CBE: hx, OME, voice, PO trials (thin, thick, puree: 5ml, 10ml, 3 oz)	86%/50% (3 ounce swallow)91%/47% (overall)	AD	N/A	Yes (inter,intra)
Nishiwaki et al (2005)	stroke	61	VFSS	SUBJ, SST, MWST (20ml)	72%/67% (cough/voice chng : WST)	A	Chi square	no
McCullough et al (2005)	stroke	165	VFSS	HX, SUBJ, WST (5ml, 10ml, 3 oz)	48%/95% (3 ounce only)	A	N/A	Yes (inter, intra)
Ramsey et al (2006)	stroke	189	VFSS	mBSA (OME, 5ml, 75 ml of contrast, pulse ox)	33%/62% (pulse ox, >2%)13%/95% (pulse ox, >5%)	A	N/A	blinded no reliability
Trapl et al (2007) Gugging	stroke	50	FEES	SST, BST (puree, thin 5, 10, 20, 50 ml, solids)	100%/50% (S1*) 100%/69% (S2*)	AR and D	N/A	Yes (inter)
Suiter, Leder (2008)	H	3000	FEES	WST (3 ounce)	96.5%/48.7%96.4%/46.4%	Asafe for PO	N/A	no
Clave et al (2008)	H	85	VFSS	V-VST (liq,nectar,pudding: 5, 10, 20ml) SpO2, cough, voice	88.2%/64.7%100%/28.8%*	A & PA	N/A	blinded no reliability
Weinhardt et al (2008)	stroke	83	No	NIHSS, PO trials (lemon ice, puree, water: 1tsp)	N/A	safe for PO	Reliability	Yes
Turner-Lawrence et al (2009)	stroke	84	NO	Quest, WST (10ml)	96%/56%	D	N/A	Yes (inter)
Suiter et al (2009)	pedi	56	FEES	WST (3 ounce)	100%/51.2%100%/44.1%	Asafe for PO	N/A	no
Bravata (2009)	stroke	101	No	RN questionnaire (subj, NIHSS severity)	29%/84% (RN screen)	D	N/A	Yes chart review only
Martino et al (2009)	stroke	311	VFSS	TOR-BSST: Voice, lingual movement, WST (5ml x 10, sip)	91.3%/omitted	D	N/A	Yes (inter)
Edmiaston et al (2010)	stroke	300	NO	ASDS (3 ounce WST and screen questions)	95%/68%91%/74%	ARD	N/A	yes
Schultheiss et al (2011)	H	62	FEES	SST, WST(5,10,20ml), BST (.3.,.5 and 1 tsp jelly)	89.6%/72.7% (BST + SST)70.7%/82.5% (WST)	AR	N/A	Yes (inter)
Patterson et al (2011)	H&N Ca	110	FEES	WST (100ml)	>67%/>46%*	AR	N/A	Yes (inter, intra)
Mandysova et al (2011)	neuro, ENT pts	108	FEES	SUBJ, BST (puree 20 ml, thin 20 ml, 60ml)	87%/30.4% (overall)95.2%/27.5% (neuro)60%/60% (ENT)	AR	N/A	no
Steele et al (2011)	H	40	VFSS	SUBJ, WST (5ml, cup/sips)	64%/16%71%/13%	ARD	N/A	Yes (intra, inter)
Yeh et al (2011)	stroke	176	No	3 SSS (Quest, WST (3ml, 100ml))	N/A	SAP	Chi square, Fisher's exact test, Logistic Regression	no

KEY: SST= saliva swallow test; BST= bolus swallow test; WST= water swallow test; 3-SSS = 3-step swallow screen; ASDS= acute stroke dysphagia screen; mBSA= modified bedside swallow assessment; H=heterogeneous; SAP= stroke acquired pneumonia; UAO= upper airway obstruction Color: blue = WST; purple = BST; pink= other assessments

Table 2

STUDY	Tongue	Lips	Facial Weakness	Jaw	Soft Palate	Volitional Cough	Spontaneous Cough	Reduced Laryngeal Elevation	Gag	Oral Apraxia	Motor Speech	Dysarthria	Intelligibility	Voice Quality/Dysphonia	Secretions	Wet Voice	Breathy Voice	Pneumonia	Poor nutrition	Feeding Tube	Overt signs on trial swallows
Daniels et al (1997)	n/a	n/a	n/a	n/a	n/a	30/94 (d.severity)	n/a	n/a	53/66 (d.severity)	n/a	n/a	76/80 (d.severity)	n/a	73/75 (d.severity)	n/a	n/a	n/a	n/a	n/a	n/a	61/78 (cough: d.severity) 30/87 (voice change: d.severity)
Daniels et al (1998)	n/a	n/a	n/a	n/a	n/a	48/94 (abn:aj)	n/a	n/a	61/82 (a)	n/a	n/a	76/53(a)	n/a	76/67 (a)	n/a	n/a	n/a	n/a	n/a	n/a	57/65 (cough:a) 38/85 (voice change:a)
Logemann et al (1999)	n/a	n/a	30/52 (d)	n/a	n/a	n/a	n/a	66/57 (a)	33/81 (a)	5/80 (d)	n/a	n/a	n/a	n/a	n/a	42/76 (a)	n/a	n/a	n/a	n/a	78/58 (a) 72/53 (d)
McCullough * (2001)	50/73 (strength:a) 36/71 (strength:a) 50/74 (strength:a)	27/76 (strength:a) 48/71 (ROM)(pucker:a) 76/41 (ROM)(retract:a)	n/a	38/74 (ROM:a)	41/71 (strength:a)	70/45 (strength:a)	55/68 (quality:a)	68/82	41/65(a)	90/38 (pharyngeal)	41/68 (a)	77/85 (a)	73/58 (a)	n/a	50/84(a)	50/63 (a) (chance)	n/a	n/a	76/71 (a)	36/95 (a)	n/a
Rosenbeck et al (2004)	36/71 (ROM:a)	84/41 (ROM:a)	n/a	38/74 (ROM:a)	50/53 (symmetry:a)	55/68 (quality:a)	68/82 (a)	n/a	91/18 (a)	41/68 (a)	n/a	77/85 (a)	73/58 (a)	91/22 (a)	50/84 (a)	50/78 (a)	82/20 (a)	32/92 (a)	50/76 (a)	36/95 (a)	n/a
Nishiwaki et al (2005)	72/47 (ROM:a)	67/49 (ROM:a)	n/a	n/a	67/49 (elevation:a)	n/a	n/a	n/a	88/36 (a)	n/a	78/44	n/a	n/a	83/40 (a)	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Steele et al (2011)	14/72 (ROM:aj)	21/73 (ROM: dys)	n/a	n/a	n/a	42/67 (abn:dys)	n/a	n/a	n/a	n/a	n/a	n/a	n/a	57/56 (abn:p/a)	n/a	n/a	n/a	n/a	n/a	n/a	21/52 (p/a) 38/60 (dys)

Table 3**Intra and inter-rater agreements for speech-language pathologists and nurses**

Intra-rater agreement	
Speech-language pathologist #1 (based on 26 screens)	100%
Speech-language pathologist #2 (based on 75 screens)	100%
Inter-rater agreement	
Speech-language pathologists #1 and #2 vs registered nurses (101 screens)	98.01%

Table 4

RN pass/fail results of the 3-ounce water swallow challenge protocol dependent upon diagnosis (n=101)

Diagnosis	n	Pass	Fail
Cardiothoracic surgery	6	4	2
Medical	10	8	2
Pulmonary	15	10	5
Cancer	3	2	1
Left Stroke	10	4*	6
Right Stroke	8	4	4
Neurosurgery	10	8	2
Other Neurological	31	24*	7
Total	101	69	32

*nurse incorrectly passed patient when interrupted drinking occurred

Table 5

A. Cohen's kappa calculation: Agreement

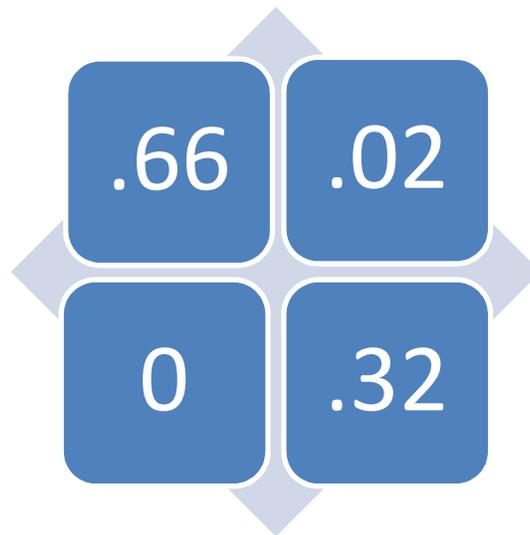
Speech-Language Pathologist

		Pass	Fail	
Nurse	Pass	67	2	n=101
	Fail	0	32	

Table 5

B. Cohen's kappa calculation

Calculation of observed agreement via conversion to frequencies and addition of agreed circumstances



$$\text{Kappa} = \frac{\text{observed} - \text{expected}}{1 - \text{expected}}$$

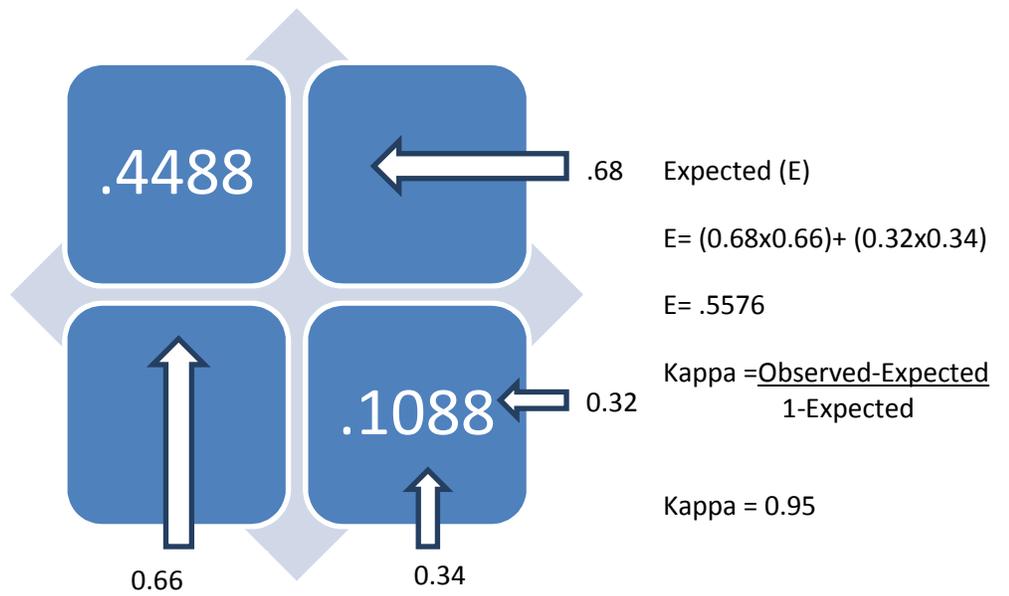
Observed

$$O = 0.66 + 0.32 = 0.98$$

Table 5

C. Cohen's kappa calculation

Calculation of expected agreement via marginal and cross products



References

Abe, S., Ishihara, K., Adachi, M., & Okuda, K. (2006). Oral hygiene evaluation for effective oral care in preventing pneumonia in dentate elderly. *Archives of gerontology and geriatrics*, 43(1), 53-64.

American Speech Language Hearing Association (2004). ASHA preferred practice pattern for speech pathology. Rockville, MD. p. 3-10.

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.

Barnard, S. L. (2011). Nursing dysphagia screening for acute stroke patients in the emergency department. *Journal of Emergency Nursing*, 37(1), 64–67.

Bours GJJ, Speyer R, Lemmens J, Limburg M, de Wit R (2008) Bedside screening tests vs. videofluoroscopy or fiberoptic endoscopic evaluation of swallowing to detect dysphagia in patients with neurological disorders: Systematic review. *Journal of Advanced Nursing* 65, 477-493.

Bravata, D. M., Daggett, V. S., Woodward-Hagg, H., Damush, T., Plue, L., Russell, S., Allen, G., et al. (2009). Comparison of two approaches to screen for dysphagia among acute ischemic stroke patients: Nursing admission screening tool versus National Institutes of Health Stroke Scale. *Journal of rehabilitation research and development*, 46(9), 1127–1134.

Cichero, J. A., Heaton, S., & Bassett, L. (2009). Triaging dysphagia: nurse screening for dysphagia in an acute hospital. *Journal of clinical nursing*, 18(11), 1649-1659.

Clavé, P., Arreola, V., Romea, M., Medina, L., Palomera, E., & Serra-Prat, M. (2008). Accuracy of the volume-viscosity swallow test for clinical screening of oropharyngeal dysphagia and aspiration. *Clinical Nutrition*, 27(6), 806–815.

Cochrane AL., Holland WW. (1971) Validation of screening procedures. *British Medical Bulletin* 27, 3-8.

Daniels, S. K., Ballo, L. A., Mahoney, M.-C., & Foundas, A. L. (2000a). Clinical predictors of dysphagia and aspiration risk: Outcome measures in acute stroke patients. *Archives of Physical Medicine and Rehabilitation*, 81(8), 1030–1033. doi:10.1053/apmr.2000.6301

Daniels, S. K., Ballo, L. A., Mahoney, M.-C., & Foundas, A. L. (2000b). Clinical predictors of dysphagia and aspiration risk: outcome measures in acute stroke patients. *Archives of physical medicine and rehabilitation*, 81(8), 1030–1033.

Daniels, S. K., Brailey, K., Priestly, D. H., Herrington, L. R., Weisberg, L. A., & Foundas, A. L. (1998). Aspiration in patients with acute stroke. *Archives of physical medicine and rehabilitation*, 79(1), 14–19.

Daniels, S. K., McAdam, C. P., Brailey, K., & Foundas, A. L. (1997). Clinical assessment of swallowing and prediction of dysphagia severity. *American Journal of Speech-Language Pathology*, 6(4), 17.

DePippo, K. L., Holas, M. A., & Reding, M. J. (1992). Validation of the 3-oz water swallow test for aspiration following stroke. *Archives of Neurology*, 49(12), 1259.

DePippo, K. L., Holas, M. A., & Reding, M. J. (1994). The Burke dysphagia screening test: validation of its use in patients with stroke. *Archives of physical medicine and rehabilitation*, 75(12), 1284.

Donzelli, J., Brady, S., Wesling, M., & Craney, M. (2003). Predictive value of accumulated oropharyngeal secretions for aspiration during video nasal endoscopic evaluation of the swallow. *The Annals of otology, rhinology, and laryngology*, 112(5), 469.

Edmiaston, J., Connor, L. T., Loehr, L., & Nassief, A. (2010). Validation of a dysphagia screening tool in acute stroke patients. *American Journal of Critical Care*, 19(4), 357–364.

Edmiaston, J., Connor, L., Steger-May, K., Ford, A. (2013) A Simple Bedside Stroke Dysphagia Screen Validated against Videofluoroscopy, Detects Dysphagia and Aspiration with High Sensitivity. *Journal of Stroke and Cerebrovascular Diseases*. (In press)

Go, A. (1998). *Refining probability: an introduction to the use of diagnostic tests*. Evidence-Based Medicine, New York, NY: McGraw-Hil, 12–33.

Gottlieb, D., Kipnis, M., Sister, E., Vardi, Y., Brill, S. (1996) Validation of the 50ml drinking test for evaluation of post-stroke dysphagia. *Disability and Rehabilitation*, 18(10), 529-532.

Hammond, C. S., Goldstein, L., Zajac, D., Gray, L., Davenport, P., & Bolser, D. (2001). Assessment of aspiration risk in stroke patients with quantification of voluntary cough. *Neurology*, 56(4), 502–506.

Hinchey, J. A., Shephard, T., Furie, K., Smith, D., Wang, D., & Tonn, S. (2005). Formal dysphagia screening protocols prevent pneumonia. *Stroke*, 36(9), 1972–1976.

Hines, S., Wallace, K., Crowe, L., Finlayson, K., Chang, A., & Pattie, M. (2011). Identification and nursing management of dysphagia in individuals with acute neurological impairment (update). *International Journal of Evidence-Based Healthcare*, 9(2), 148–150.

Holas, M. A., DePippo, K. L., & Reding, M. J. (1994). Aspiration and relative risk of medical complications following stroke. *Archives of neurology*, 51(10), 1051.

Kelly, J., D’Cruz, G., & Wright, D. (2010). Patients with dysphagia: experiences of taking medication. *Journal of advanced nursing*, 66(1), 82-91.

Leder, S. B. (2000). Use of arterial oxygen saturation, heart rate, and blood pressure as indirect objective physiologic markers to predict aspiration. *Dysphagia*, 15(4), 201–205.

Leder, S. B., & Ross, D. A. (2000). Investigation of the causal relationship between tracheotomy and aspiration in the acute care setting. *The Laryngoscope*, 110(4), 641-644.

Leder, S. B., & Espinosa, J. F. (2002). Aspiration risk after acute stroke: comparison of clinical examination and fiberoptic endoscopic evaluation of swallowing. *Dysphagia*, 17(3), 214–218.

Leder, S. B., Suiter, D. M., & Warner, H. L. (2009). Answering orientation questions and following single-step verbal commands: effect on aspiration status. *Dysphagia*, 24(3), 290–295.

Leder, S. B., & Ross, D. A. (2010). Confirmation of no causal relationship between tracheotomy and aspiration status: a direct replication study. *Dysphagia*, 25(1), 35-39.

Leder, S. B., Suiter, D. M., & Green, B. G. (2011). Silent aspiration risk is volume-dependent. *Dysphagia*, 26(3), 304–309.

Leder, S. B., Suiter, D. M., Warner, H. L., & Kaplan, L. J. (2011). Initiating safe oral feeding in critically ill intensive care and step-down unit patients based on passing a 3-ounce (90 milliliters) water swallow challenge. *The Journal of Trauma and Acute Care Surgery*, 70(5), 1203.

Leder, S. B., & Lerner, M. Z. (2013). Nil per os except medications order in the dysphagic patient. *QJM*, 106(1), 71-75.

Leder, S. B., Suiter, D. M., Warner, H. L., Acton, L. M., & Swainson, B. A. (2012). Success of Recommending Oral Diets in Acute Stroke Patients Based on Passing a 90-cc Water Swallow Challenge Protocol. *Topics in Stroke Rehabilitation*, 19(1), 40–44.

Leder, S., Suiter, D., Warner, H., Acton, L., & Siegel, M. (2012). Safe initiation of oral diets in hospitalized patients based on passing a 3-ounce (90 cc) water swallow challenge protocol. *QJM*, 105(3), 257–263.

Leder, S. B., Suiter, D. M., Murray, J., & Rademaker, A. W. (2013). Can an Oral Mechanism Examination Contribute to the Assessment of Odds of Aspiration?. *Dysphagia*, 1-5.

Leder, S. B., Judson, B. L., Sliwinski, E., & Madson, L. (2013). Promoting Safe Swallowing When Puree is Swallowed Without Aspiration but Thin Liquid is Aspirated: Nectar is Enough. *Dysphagia*, 1-5.

Lim, S. H., Lieu, P., Phua, S., Seshadri, R., Venketasubramanian, N., Lee, S., & Choo, P. W. (2001). Accuracy of bedside clinical methods compared with fiberoptic endoscopic examination of swallowing (FEES) in determining the risk of aspiration in acute stroke patients. *Dysphagia*, 16(1), 1–6.

Linden, P., & Siebens, A. (1983). Dysphagia: predicting laryngeal penetration. *Archives of physical medicine and rehabilitation*, 64(6), 281.

Link, D., Willging, J. P., Miller, C. (2000). Pediatric laryngopharyngeal sensory testing during flexible endoscopic evaluation of swallowing: feasible and correlative. *The Annals of otology, rhinology & laryngology*, 109(10), 899-905.

Logemann, J. A., Veis, S., & Colangelo, L. (1999). A screening procedure for oropharyngeal dysphagia. *Dysphagia*, 14(1), 44–51.

Mandysova, P., Škvrňáková, J., Ehler, E., & Černý, M. (2011). Development of the Brief Bedside Dysphagia Screening Test in the Czech Republic. *Nursing & Health Sciences*, 13(4), 388–395.

Mann, G., Hankey, G. J., & Cameron, D. (2000). Swallowing disorders following acute stroke: prevalence and diagnostic accuracy. *Cerebrovascular Diseases*, 10(5), 380–386.

Mann, G. (2002). MASA: the Mann assessment of swallowing ability. Singular Publishing Group.

Martino, R., Silver, F., Teasell, R., Bayley, M., Nicholson, G., Streiner, D. L., & Diamant, N. E. (2009). The toronto bedside swallowing screening test (TOR-BSST) development and validation of a dysphagia screening tool for patients with stroke. *Stroke*, 40(2), 555–561.

Massey, R., & Jedlicka, D. (2002). The Massey bedside swallowing screen. *Journal of Neuroscience Nursing*, 34(5), 252–260.

McCullough, G. H., Wertz, R. T., Rosenbek, J. C., & Dinneen, C. (1999). Clinicians' preferences and practices in conducting clinical/bedside and videofluoroscopic swallowing examinations in an adult, neurogenic population. *American journal of speech-language pathology*, 8(2), 149.

McCullough, G. H., Wertz, R. T., Rosenbek, J. C., Mills, R. H., Webb, W. G., & Ross, K. B. (2001). Inter-and intrajudge reliability for videofluoroscopic swallowing evaluation measures. *Dysphagia*, 16(2), 110–118.

McCullough, G., Rosenbek, J., Wertz, R., McCoy, S., Mann, G., & McCullough, K. (2005). Utility of clinical swallowing examination measures for detecting aspiration post-stroke. *Journal of Speech, Language and Hearing Research*, 48(6), 1280.

McCullough, G., Wertz, R., & Rosenbek, J. (2001). Sensitivity and specificity of clinical/bedside examination signs for detecting aspiration in adults subsequent to stroke. *Journal of communication disorders*, 34(1), 55–72.

Nishiwaki, K., Tsuji, T., Liu, M., Hase, K., Tanaka, N., & Fujiwara, T. (2005). Identification of a simple screening tool for dysphagia in patients with stroke using factor analysis of multiple dysphagia variables. *Journal of Rehabilitation Medicine*, 37(4), 247–251.

Patterson, J. M., Hildreth, A., McColl, E., Carding, P. N., Hamilton, D., & Wilson, J. A. (2011). The clinical application of the 100mL water swallow test in head and neck cancer. *Oral oncology*, 47(3), 180–184.

Perry, L. (2001). Screening swallowing function of patients with acute stroke. Part one: Identification, implementation and initial evaluation of a screening tool for use by nurses. *Journal of Clinical Nursing*, 10(4), 463–473.

Petersen, N. J. (2013). Implementation of Stroke Dysphagia Screening in the Emergency Department. *Nursing Research and Practice*, 2013.

Ramsey, D. J., Smithard, D. G., & Kalra, L. (2006). Can pulse oximetry or a bedside swallowing assessment be used to detect aspiration after stroke? *Stroke*, 37(12), 2984–2988.

Rosenbek, J. C., McCullough, G. H., & Wertz, R. T. (2004). Is the information about a test important? Applying the methods of evidence-based medicine to the clinical examination of swallowing. *Journal of communication disorders*, 37(5), 437–450.

Schultheiss, C., Nusser-Müller-Busch, R., & Seidl, R. O. (2011). The semisolid bolus swallow test for clinical diagnosis of oropharyngeal dysphagia: a prospective randomised study. *European Archives of Oto-Rhino-Laryngology*, 268(12), 1837–1844.

Smith, H. A., Lee, S. H., O'Neill, P. A., & Connolly, M. J. (2000). The combination of bedside swallowing assessment and oxygen saturation monitoring of swallowing in acute stroke: a safe and humane screening tool. *Age and Ageing*, 29(6), 495–499.

Smithard, D. G., O'Neill, P. A., Park, C., England, R., Renwick, D. S., Wyatt, R., Morris, J., et al. (1998). Can bedside assessment reliably exclude aspiration following acute stroke? *Age and Ageing*, 27(2), 99–106.

Splaingard, M., Hutchins, B., Sulton, L., & Chaudhuri, G. (1988). Aspiration in rehabilitation patients: videofluoroscopy vs bedside clinical assessment. *Archives of physical medicine and rehabilitation*, 69(8), 637.

Steele, C., Molfenter, S., Bailey, G., Cliffe Polacco, R., Waito, A., Zoratto, D., & Chau, T. (2011). Exploration of the utility of a brief swallow screening protocol with comparison to concurrent videofluoroscopy. *Can J Speech Lang Pathol Audiol*, 35(3), 228–42.

Suiter, D., & Leder, S. (2008). Clinical Utility of the 3-ounce Water Swallow Test. *Dysphagia*, 23(3), 244–250. doi:10.1007/s00455-007-9127-y

Suiter, D. M., Leder, S. B., & Karas, D. E. (2009). The 3-ounce (90-cc) water swallow challenge: a screening test for children with suspected oropharyngeal dysphagia. *Otolaryngology--Head and Neck Surgery*, 140(2), 187–190.

Suiter DM, Sloggy J, Leder SB. Validation of the Yale Swallow Protocol: A prospective double-blinded videofluoroscopic study. *Dysphagia*. In Press, doi: 10.1007/s00455-013-9488-3.

- Teramoto, S., & Fukuchi, Y. (2000). Detection of aspiration and swallowing disorder in older stroke patients: simple swallowing provocation test versus water swallowing test. *Archives of physical medicine and rehabilitation*, 81(11), 1517–1519.
- Trapl, M., Enderle, P., Nowotny, M., Teuschl, Y., Matz, K., Dachenhausen, A., & Brainin, M. (2007). Dysphagia bedside screening for acute-stroke patients The Gugging Swallowing Screen. *Stroke*, 38(11), 2948–2952.
- Turner-Lawrence, D. E., Peebles, M., Price, M. F., Singh, S. J., & Asimos, A. W. (2009). A feasibility study of the sensitivity of emergency physician Dysphagia screening in acute stroke patients. *Annals of emergency medicine*, 54(3), 344.
- Weinhardt, J., Hazelett, S., Barrett, D., Lada, R., Enos, T., & Keleman, R. (2008). Accuracy of a bedside dysphagia screening: a comparison of registered nurses and speech therapists. *Rehabilitation Nursing*, 33(6), 247–252.
- Westergren, A., Hallberg, I. R., & Ohlsson, O. (1999). Nursing assessment of dysphagia among patients with stroke. *Scandinavian Journal of Caring Sciences*, 13(4), 274.
- Wood, J.M. (2007). Understanding and Computing Cohen's Kappa: A tutorial. WebPsychEmpiricist. Retrieved from http://wpe.info/papers_table.html
- Wu, M.-C., Chang, Y.-C., Wang, T.-G., & Lin, L.-C. (2004). Evaluating swallowing dysfunction using a 100-ml water swallowing test. *Dysphagia*, 19(1), 43–47.

Yeh, S.-J., Huang, K.-Y., Wang, T.-G., Chen, Y.-C., Chen, C.-H., Tang, S.-C., Tsai, L.-K., et al. (2011). Dysphagia screening decreases pneumonia in acute stroke patients admitted to the stroke intensive care unit. *Journal of the neurological sciences*, 306(1), 38–41.