

## Objective Computer-Based Assessment of Valleculae Residue – Is It Useful?

Jill C. Dyer, MSc, BSc(Hons),<sup>1</sup> Paula Leslie, PhD,<sup>2</sup> and Michael J. Drinnan, PhD<sup>3</sup>

<sup>1</sup>Department of Speech and Language Therapy, University Hospital North Durham; <sup>2</sup>School of Surgical and Reproductive Sciences, The Medical School, University of Newcastle; and <sup>3</sup>Regional Medical Physics, Freeman Hospital, Freeman Road, Newcastle upon Tyne, NE7 7DN, UK

**Abstract.** There have been questions about the reliability of subjective rating scales used to assess valleculae residue from fluoroscopic images. The aim of this study was to assess interrater agreement on one such scale, and compare it with agreement using a new objective measurement scale. Five speech and language therapists rated 100 valleculae residue still images from 20 consecutive patients using standard clinical practice (i.e., subjective visual grading of the videofluoroscopy still and rating as none, mild, moderate, or severe). The images were rerated by the same clinicians using Picture Archiving Communication System measurement tools. The valleculae residue ratio relates the residue size to the size of an individual's valleculae. A valleculae residue ratio scale was devised using a linear classifier, which defines the cutoff between grades of valleculae residue (none, mild, moderate, and severe). The new method proved at least as reliable as the traditional method; for interrater reliability, kappa = 0.73 vs. 0.73; for intrarater reliability, kappa = 0.87 vs. 0.85. The valleculae residue ratio is proposed as a new quick reliable method of quantifying residue where the Picture Archiving Communication System is available. We now wish to test the impact of this method where poor inter- and intrarater reliability exists.

**Key words:** Dysphagia — Picture Archiving Communication Systems — Valleculae residue ratio — Pharyngeal retention — Measurement tool — Deglutition — Deglutition disorders.

---

Up to 30% of adults over 65 years old who are admitted to acute care hospitals have dysphagia [1, 2]. Dysphagic patients require comprehensive clinical evaluation to prevent malnutrition, dehydration, and aspiration. Accurate instrumental assessment and interpretation of each symptom of dysphagia is crucial to diagnose and evaluate the severity of a swallowing disorder and to develop effective treatment plans.

Valleculae residue is an important indicator of swallow efficiency and is a significant part of a patient's swallowing record [3]. Videofluoroscopic swallow studies (VFSS) have demonstrated that when there is insufficient contact between base of tongue and posterior pharyngeal wall, residue remains in the valleculae [4]. Where large quantities of valleculae residue exist, the patient is at increased risk of aspirating residue during respiration after swallowing [5]. Postswallow aspiration is a frequent finding in patients with pharyngeal retention [6]. Dysphagic patients who aspirate food and/or liquids regularly are at risk of developing chest infections or pneumonia. Pneumonia is the most frequent infectious cause of death and the sixth leading cause of death in the United States [7]. The cost for treating patients over 65 years old with community-acquired pneumonia in the United States has been estimated at \$4.8 billion per annum [8]. It is vital that patients at risk of aspiration are accurately assessed and monitored. Yet, clinicians and researchers are unclear how to

---

This work was supported by Durham and Chester-le-Street Primary Care Trust (now assimilated into County Durham Primary Care Trust). Work was performed at University Hospital North Durham, North Road, Durham, DH1 5TW, UK.

*Correspondence to:* Jill C. Dyer, MSc, BSc(Hons), Department of Speech and Language Therapy, University Hospital North Durham, North Road, Durham, DH1 5TW, UK; E-mail: Jill.Dyer@cddah.nhs.uk

objectively grade severity of valleculae residue, what quantity of valleculae residue is within normal limits, and what level of residue in proportion to valleculae size contributes to overflow aspiration.

Dysphagic patients unable to clear large amounts of valleculae residue are frequently prescribed therapeutic treatments, e.g., tongue-base exercises [4] and recommended compensatory strategies [5]. Currently, there are no objective measurement tools or standardized methods capable of accurately monitoring a patient's long-term response to this treatment.

VFSS is the procedure of choice when an understanding of pharyngeal physiology in relation to symptoms is required [5]. All published VFSS protocols identify valleculae residue as a standard diagnostic parameter [5, 9]. For VFSS to be considered a valid assessment tool, consistent, reliable interpretation is required (i.e., test-retest interpretation stability) regardless of whether it is an individual reporting on different occasions or a group of individuals reporting at the same time. Studies of VFSS show high levels of disagreement in judging many parameters of swallowing abnormalities [10–13]. Kuhlemeier et al. [14] showed higher reliability for normal VFSS or where aspiration was absent than for abnormalities such as severity of valleculae residue. Methodologic flaws existed: poor image quality of videotapes, no clinical information available for judges, no slow-motion replay or joint agreement between judges was allowed [15]. In Stoeckli et al.'s study [13] nine experienced observers were asked to estimate percentage of valleculae residue; the resulting range was so "extremely" large that the authors decided it could not be analyzed in any more detail than absent or present ( $\kappa = 0.52$ ). In contrast, 100% agreement level was reached in Dejaeger's [16] study where pharyngeal residue was rated as present or absent using four clearly defined categories: no retention, retention in the valleculae, retention in the pyriform sinuses, and diffuse pharyngeal retention. However, no attempt was made to differentiate between degrees of pharyngeal residue. In both the McCullough et al. [12] and the Stoeckli et al. [13] studies high interrater scores were achieved using the aspiration/penetration scale [17]. This demonstrates that where well-defined rating scales are employed reliability scores increase. No robust valleculae residue severity rating scales currently exist.

Poor rater reliability is unacceptable because the same rater could make different judgments on the same patient over time, the same rater could vary his/her judgment between patients, or different raters could judge the same patient differently. For example, one clinician may grade valleculae residue level as

mild and another clinician may grade it as moderate. This discrepancy may mean the difference between receiving no treatment or being prescribed a range of exercises and compensatory strategies.

Current methods of quantifying pharyngeal residue are limited. Scintigraphy has long been advocated as the only technique for measurement of pharyngeal retention [18]. This nuclear medicine test allows a bolus mixed with a prescribed amount of radioactive material to be traced and residue to be measured. Only one swallow per subject can be analyzed to avoid ingestion of multiple radioactive boluses. Representative sampling of an individual's average valleculae residue level is impossible. Scintigraphic measures have good test-retest reliability [18]. However, it is not a standard diagnostic test nor is it widely available to clinicians. This method gives only overall readings of pharyngeal residue [19] because it does not allow visualization of oropharyngeal structures and as such we cannot measure vallecula residue. The only current alternative is to use a qualitative percentage or graded pharyngeal residue scale, based on 2-dimensional videofluoroscopy images, and/or subjective estimation of valleculae residue level.

Hind et al. [20] described a three-point rating scale [0 = no residue, 1 = coating (line of barium), 2 = pooling of barium] that measures residue at the valleculae, posterior pharyngeal wall, pyriform sinuses, oral cavity, and upper esophageal sphincter. This was based on normal swallows of middle-aged and older adults. The scale does not differentiate between moderate and severe quantities of residue and no reliability measures were completed for the data.

Eisenhuber et al. [6] developed a scale using both anterior and lateral plane stills to investigate residue as a predictor of aspiration: mild = <25% height of valleculae/pyriform sinuses, moderate = 25%–50% height of valleculae/pyriform sinuses, severe = >50% height of valleculae/pyriform sinuses. This scale is more sensitive than the previous one but does not include width measurements. Because no instructions on how to replicate this grading method were given, one can only assume grades were visually estimated and hence unreliable.

In Han et al.'s study [21] a functional dysphagia scale was devised that allowed systematic rating of 11 parameters, including valleculae residue based on videofluoroscopic studies: grade 0 = no residue, grade 1 = <10% of width of valleculae, grade 2 = >10% to <50% of width of valleculae, grade 3 = >50% of width of valleculae. Height of residue was not considered when grading valleculae residue.

**Table 1.** Demographics of raters

|          | Years qualified | Hours VFSS training | Frequency in VFSS clinic | Reading glasses? |
|----------|-----------------|---------------------|--------------------------|------------------|
| Rater 1  | 9               | 30                  | Bimonthly                | No               |
| Rater 2  | 4               | 26                  | Six-monthly              | No               |
| Rater 3* | 4               | 35                  | Monthly                  | Yes              |
| Rater 4  | 10              | 15                  | Monthly                  | No               |
| Rater 5  | 5               | 10                  | Weekly                   | No               |

\*Last eye examination 2 months prior to rating valleculae residue levels

None of these studies refers to taking actual measurements of both height and width of residue or valleculae. We presume that these studies rely on a clinician's visual acuity to grade a cutoff point for each category. In the absence of defined boundaries and appropriate measurement tools, poor interrater reliability will flourish as clinicians struggle to achieve consensus on what constitutes none, mild, moderate, and severe valleculae residue.

Picture Archiving and Communication Systems (PACS) were first introduced in 1982 [22]. PACS is a computer-based digital method for capturing radiologic images [23]. PACS workstations allow clinicians to capture good anatomical image details during transient events such as swallowing [23] but also has facilities for taking measurements of structures.

While custom video acquisition and analysis packages would in principle allow similar measurements, PACS facilities may offer a quick and objective method of grading valleculae residue that is not dependent on experience or training. Using PACS measurement tools, this study was designed (1) to devise an objective valleculae residue rating scale based on valleculae residue ratio (this ratio is based on objective measurements and expresses the size of residue as a proportion of an individual's valleculae size), and (2) to investigate interrater and intrarater reliability among five speech and language pathologists (SLPs) reporting on valleculae residue levels by comparing the traditional visual grading method of estimating residue level with use of computer-based objective measurements.

## Participants and Methods

From December 2003 to February 2004, 168 patients with symptoms of dysphagia were referred to the Department of Speech and Language Therapy, University Hospital North Durham. Of these, 20 patients were consecutively referred for videofluoroscopy examination according to standard departmental criteria. There were no exclusions and patients were entered into the study regardless of underlying disorder: 14 stroke, 2 head and neck

cancer, 2 other neurologic diseases, 1 multiple sclerosis, and 1 Parkinson's disease. There were 8 females and 12 males with a mean age of 72 years (range = 36–92 years).

Five SLPs (this excluded the researcher) were blinded to the purpose of the study. They had at least one year of dysphagia management experience (including VFSS interpretation experience) and a post-basic dysphagia qualification (see Table 1 for rater demographics). To avoid bias, all rating SLPs signed a confidentiality form agreeing not to discuss any part of the study until it was complete.

County Durham Local Research Ethics Committee granted ethical approval for the study. Consent/assent was obtained for all participants following a verbal and written (information leaflet) explanation of the research. Where a participant was unable to give informed consent, the researcher explained as far as possible the nature of the study with gesture and modified language and consent was obtained from the next of kin. If at any time the participant looked uneasy or indicated they wanted to withdraw, the study was terminated for that participant.

## Image Acquisition System

The image acquisition system determines the quality of the images stored. The fluoroscopy unit used for this study was the Siemens Polystar Polydorus SX-65 (Siemens, Erlangen, Germany). All VFSS studies were performed using identical magnification setting 1, field size 28, and "VideoSwallow" modality.

Our Picture Archiving Communication System (PACS, G.E. Medical Systems, Berks, UK) generates lossless DICOM clips and stills. For this study all still images were stored in an identical manner: DICOM file format at a standard size resolution of 1200 × 1600 pixels. For image review, the two-screen diagnostic monitor size was 410 mm × 310 mm, with a resolution of 1200 × 1600 pixels.

All images were presented at the same size, scaled so that 1 mm on the screen corresponded to approximately 1 mm in the real world. Because magnification was fixed at standard setting 1 (i.e., a given real-world distance always occupied the same distance in the image), we could not adjust at the same time for variations in head and valleculae sizes. However, the PACS system provides internally calibrated measurement tools that automatically convert on-screen coordinates into real-world measurements based on image acquisition parameters stored in the DICOM header.

## Videofluoroscopic Examination Technique – Collection of Stills

All VFSS were performed in lateral and anterior planes by five videofluoroscopy-trained radiographers and two speech and language therapists. Each patient was positioned in the lateral plane (head placed in the midline) in a Mangar Porter X-ray chair, with the head and body aligned using markers on the radiography

equipment to ensure the same positioning for all images. The fluoroscopy field of view centered on the oral cavity and pharynx.

The patient was instructed to swallow two individually presented teaspoon boluses each of 5 ml “Varibar” nectar, 5 ml “Varibar” honey, teaspoon quantities of banana and bread soaked in 98% w/w density barium sulfate suspension (E-Z-EM Inc, Bicester, UK), which forms part of the department’s standard VFSS procedure [3]. Where a dysphagic patient demonstrated aspiration or choking on any consistency, appropriate safe modifications were made.

The radiographer stored stills of the valleculae after the swallow, indicated by the hyoid bone returning to its resting position following the test bolus. Only lateral plane stills were stored. Following each examination both radiographer and researching speech and language therapist reviewed all stills. Twenty-eight stills were excluded because of blurring or oblique images or where the image had been stored before the swallow was complete. Four stills per patient were archived into a folder accessible to the researching therapist only.

### Radiation Exposure

The Speech and Language Pathology departmental videofluoroscopy procedure [3] was followed to ensure patients received minimal radiation exposure. Stills were collected using “frame-grabbing,” whereby a still is captured from a moving image. The radiation protection advisor therefore concluded that there was no additional radiation risk to participants over and above the routine diagnostic procedure.

### Ordering of Stills

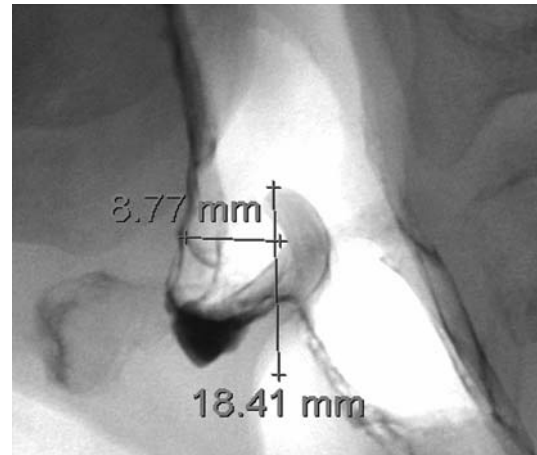
The nonrating researcher printed copies of all 80 stored stills and numbered them from 1 to 80, as far as possible according to increasing size of residue (from no residue to largest amount of residue). The ordering was used simply to allow better block randomization. For assessment of test-retest agreement, 20 stills were selected at random. Four from stills 1–20 (smallest residues), four from stills 21–40, four from stills 41–60, and four from stills 61–80 (largest residues). All 100 stills (80 original plus the 20 repeated for test-retest agreement) were then block-randomized into 5 blocks of 20, prior to first and second presentations.

### First Presentation of Stills to Raters

Raters were instructed to examine all 100 stills (80 original, plus 20 repeated) at the workstation and “document a grade (*none, mild, moderate, or severe*) for each still in their usual way.” Objective measurement tools were NOT used, and the raters were unaware that 20 of the 100 stills were presented twice. No other instructions were given to raters so as not to influence the raters’ traditional method of estimating valleculae residue.

### Second Presentation of Stills to Raters

One week later the same five raters were given 30 min of instruction by the hospital PACS trainer on how to take measurements using the on-screen objective measurement tools. The 30-min training time was judged by the trainer to be the minimum time required to become proficient. Training included written guidelines and gave raters hand-eye coordination and measurement practice on five stills (these practice stills were not included in the analyzed data).



**Fig. 1.** Valleculae height (distance from tip of epiglottis perpendicular to base of valleculae) and valleculae width (widest horizontal portion of the valleculae).



**Fig. 2.** Residue height (distance from the base of valleculae residue perpendicular to residue top) and residue width (widest horizontal portion of valleculae residue).

Raters were then asked to measure the height and width of valleculae and its residue (Figs. 1 and 2) for each of the 100 rerandomized stills. Raters were not informed that the same stills were used in the first and second presentations, albeit in a different order. Moreover, it is unlikely that they would have been able to distinguish a specific individual’s vallecula or remembered how they had rated it.

### Data Analysis

#### Valleculae Residue Ratio

First, all objective measurements were converted to a valleculae residue ratio (VRR). This expresses the size of residue (residue height  $\times$  residue width) as a proportion of an individual’s valleculae size (valleculae height  $\times$  valleculae width). For example a VRR of 0.5 would indicate that residue half-filled the valleculae.

**Table 2.** Mean PACS measurements for each parameter per subjective visual grading

|                          | Subjective visual grading |       |          |        | Overall |
|--------------------------|---------------------------|-------|----------|--------|---------|
|                          | None                      | Mild  | Moderate | Severe |         |
| Valleculeae height (mm)  | 12.60                     | 14.50 | 13.90    | 16.90  | 14.00   |
| Valleculeae width (mm)   | 7.34                      | 8.23  | 9.23     | 9.84   | 8.39    |
| Residue height (mm)      | 0.59                      | 3.33  | 6.24     | 11.20  | 4.20    |
| Residue width (mm)       | 0.88                      | 4.45  | 7.07     | 8.53   | 4.37    |
| Residue ratio (no units) | 0.03                      | 0.17  | 0.40     | 0.61   | 0.24    |

**Table 3.** Landis and Koch guidelines for kappa

| Kappa value | Level of agreement |
|-------------|--------------------|
| <0.20       | Poor               |
| 0.21–0.40   | Fair               |
| 0.41–0.60   | Moderate           |
| 0.61–0.80   | Good               |
| 0.81–1.00   | Very Good          |

### Development of the Valleculeae Residue Ratio Scale

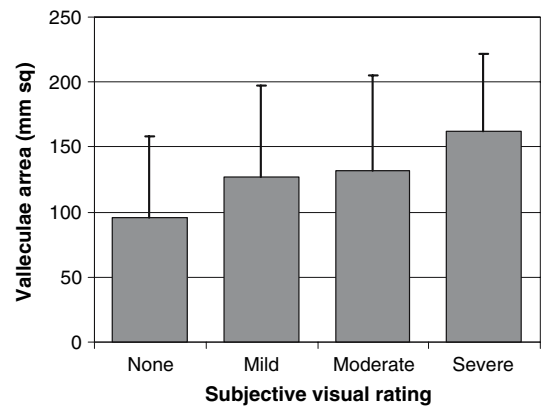
We wished to make direct comparisons of rater agreement for the subjective visual grading and objective measurements. It was therefore necessary to convert the numeric VRR measurements to four corresponding categories—none, mild, moderate, severe—for comparison with subjective visual rating. We applied linear discriminant analysis [24] to data from both first (traditional subjective visual grading) and second (PACS measurement) presentations to devise the valleculeae residue ratio scale (VRRS). Using this classification, assuming a broadly normal distribution, each category (none, mild, moderate, and severe) will include the same overall proportion of measurements in each group. For example, if 20% of subjective visual ratings were “none,” then 20% of the VRRS should also have the “none” rating. By categorizing ratio scores in this way, one can make a direct comparison between an objective ratio score, e.g., 0.5, and a subjective category label, e.g., moderate (Table 2).

### Measurement of Reliability

Inter- and intrarater reliability were measured using weighted kappa, which corrects for the effect of chance and bias. By definition, kappa ranges from 0 (chance agreement) to 1 (complete agreement). While subjective descriptions of agreement have been criticized, kappa is sometimes interpreted as defined by Landis and Koch guidelines [25] (Table 3). The weighted version of kappa gives credit for partial agreement (e.g., ratings of moderate versus severe would gain more credit than ratings of none versus severe).

### Interrater Reliability

Interrater reliability among the five raters was measured for 80 stills (excluding the 20 repeats). Kappa was calculated for the subjective visual ratings (none/mild/moderate/severe) and the residue ratio after the numeric measurements had been categorized as none, mild, moderate, or severe using the VRRS.



**Fig. 3.** The relation of valleculeae area (height  $\times$  width) with subjective visual grading of residue. The figure illustrates that larger valleculeae are associated with subjectively larger residues. Error bars are standard deviations.

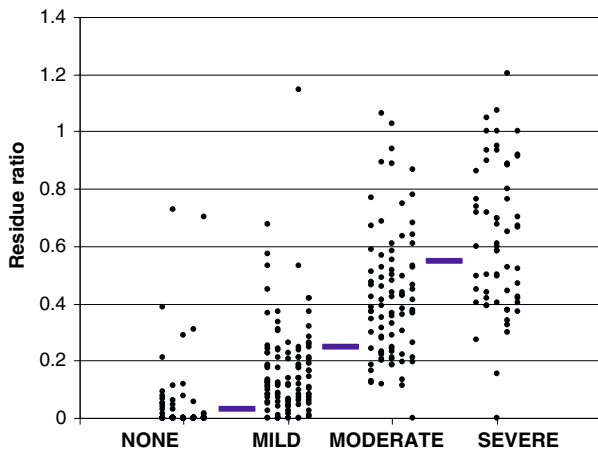
### Intrarater Reliability

Individual and mean intrarater reliability scores were calculated using the 20 repeated stills, first presentation (subjective visual grading). Kappa was calculated separately for the 20 repeated stills, second presentation (objective measurement tools).

### Results

#### Summary of Subjective Grading and Objective Measurements

Table 2 summarizes the mean objective measurements taken from 80 still images (i.e., excluding the 20 repeats). The overall mean values across all subjects and raters were 14.00 mm (valleculeae height), 8.39 mm (width), 4.20 mm (residue height), 4.37 mm (residue width), and 0.24 (residue ratio). Overall, 179 (36%) of subjective visual grading were none, 137 (27%) were mild, 101 (20%) were moderate, and 83 (17%) were severe. In Table 2 the objective measurements are summarized according to the perceived subjective visual grading. Interestingly, mean valleculeae height and width measurements increased with severity of residue, i.e., the larger the valleculeae



**Fig. 4.** The relationship of residue ratio with subjective visual category for 80 still images. The blue lines indicate where the analysis (classification) divides the categories.

surface area, the larger the amount of residue. This relationship is shown in Figure 3.

#### *Development of the Valleculae Residue Ratio Scale*

The VRRS expresses the size of residue as a proportion of size of valleculae. It is independent of any scaling applied to the image regardless of whether the image is enlarged or reduced. This would not be the case if height, width, or area alone were used as the classifier.

Figure 4 shows all the individual objective measurements of residue ratio that correspond to each subjective grade; the VRR scale was based on these data. Cutoffs were set at residue ratios of 0.03 (none/mild), 0.25 (mild/moderate), and 0.55 (moderate/severe); for pragmatic reasons these were rounded from the exact values suggested by the discriminant analysis. Table 4 shows these cutoff points expressed both as a score and a percentage, so a comparison can be made with other percentage valleculae residue scales in the literature. Using these cutoff points to categorize the VRRS measurements, the VRRS demonstrated good agreement ( $\kappa = 0.69$ ) between subjective visual ratings and objective residue ratio measurements.

#### Interrater Reliability

Weighted interrater kappa scores for five raters were identical for both methods: 0.73 (95% CI 0.67–0.79), indicating good agreement whether subjective visual rating or VRR was used.

#### Intrarater Reliability

Mean intrarater reliability scores among five raters were higher than for interrater reliability. That is,

**Table 4.** Valleculae residue ratio scale: devised using classification of subjective visual rating and PACS measurement results from five raters

| Subjective visual grading | Valleculae residue ratio score | Graded percentage (%) |
|---------------------------|--------------------------------|-----------------------|
| None                      | $< 0.03$                       | $< 3$                 |
| Mild                      | $\geq 0.03$ to $< 0.25$        | $\geq 3$ to $< 25$    |
| Moderate                  | $\geq 0.25$ to $< 0.55$        | $\geq 25$ to $< 55$   |
| Severe                    | $\geq 0.55$                    | $\geq 55$             |

individuals were more consistent at rating with themselves than with the team (Table 5). This is inevitable, because individual inconsistencies (poor intrarater agreement) will also be reflected in poor interrater agreement.

Figure 5 shows individual intrarater reliability over 20 stills, comparing residue ratio with subjective visual rating. Four of the five raters scored equal to or marginally higher for self-agreement when using residue ratio rating than subjective visual rating. Overall, there was very good agreement for both subjective visual gradings ( $\kappa = 0.85$ ) and PACS ratings ( $\kappa = 0.87$ ).

It was noteworthy that raters' mean subjective visual category was related to their mean objective measurement (Fig. 6). Raters who rated high (over-estimated) the amount of valleculae residue by subjective visual rating also tended to overestimate the amount of valleculae residue on PACS measurements ( $r = 0.94$ ,  $p < 0.05$ ).

#### Discussion

Up to 97% of clinicians in the United States are reported to have local access to VFSS compared to 53% with access to fiberoptic endoscopic evaluation of swallowing and 1.6% scintigraphy [26]. It is vital that the potential of any technology complimentary to VFSS, which may improve reporting precision, be thoroughly tested.

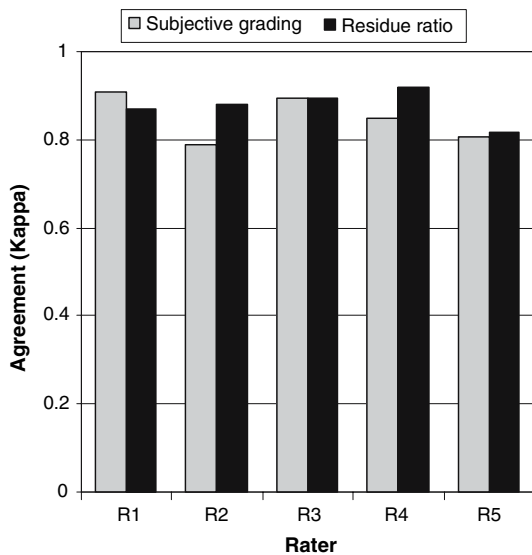
#### *Development of the Valleculae Residue Ratio Scale*

The VRRS differs from previously proposed scales because it has been based on the consensus of five reviewers using visual estimation of valleculae residue severity and objective measurements of the same 80 stills. The scale has isolated a precise cutoff point for each grade of valleculae residue.

It was noteworthy that subjectively larger residues were generally associated with larger valleculae (Fig. 3). Though this might be expected for severe residues (presumably only large valleculae can

**Table 5.** Weighted kappa (95% confidence interval) for inter- and intrarater reliability among five raters for subjective visual grading and PACS residue ratio

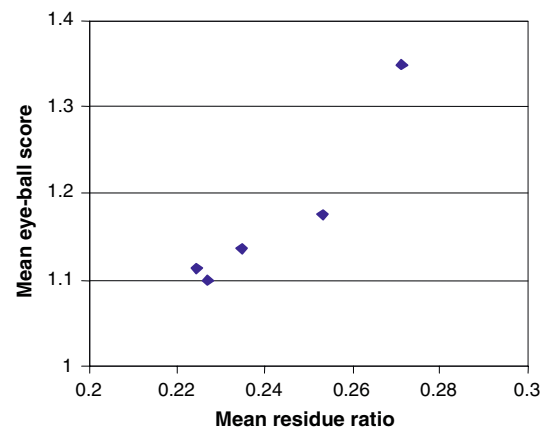
|                           | Interrater agreement<br>( $n = 80$ stills) | Intrarater agreement<br>( $n = 20$ repeated stills) |
|---------------------------|--|---|
| Subjective visual grading | 0.73 (0.67–0.79)                           | 0.85 (0.61–1.0)                                     |
| PACS residue ratio        | 0.73 (0.67–0.79)                           | 0.87 (0.70–1.0)                                     |

**Fig. 5.** Individual intrarater reliability results over 20 stills: comparison of residue ratio and subjective visual rating using kappa coefficient.

hold the biggest residues), it is not clear why it should be the case for small residues that could fit equally in any valleculae. One possibility is that smaller valleculae form a smaller target and therefore present less risk of retaining residue of any size.

#### Inter- and Intrarater Reliability

Raters scored consistently high in both inter- and intrarater agreement, regardless of the rating method. One possible reason is that all raters had been trained in interpreting VFSS signs of dysphagia by the same clinician. The VRR allows clinicians to take account of the variation in valleculae size and is at least as reliable a rating method as visual estimation of valleculae residue levels among clinicians trained in VFSS interpretation. Interrater and intrarater comparison of the subjective visual and PACS rating methods showed minimal differences in kappa scores for this rater group. Residue height and residue width were also investigated as independent classifiers to establish how they impact reliability in comparison to VRR. However, there was minimal difference

**Fig. 6.** Relationship of subjective visual score (0 = none, 1, 2, 3 = severe) with residue ratio for 5 subjects ( $r = 0.94$ ,  $p < 0.05$ ).

between interrater reliability scores for residue height or residue width (both 0.74) and VRR scores (0.73).

One unexpected finding was that raters who rated high/overestimated the amount of valleculae residue by subjective visual rating also tended to measure high/overestimate the amount of valleculae residue on PACS measurements (Fig. 6). This may suggest that raters either measure an internal visual representation of what they think they see rather than what is actually presented, or more likely that where visual discrimination skills are required, subjectivity cannot be removed with use of a measurement tool. Therefore, it is not possible to conclude that the use of PACS measurement tools enhances objectivity of grading valleculae residue in this group of raters.

#### Clinical Relevance in the Assessment of Dysphagia

There was a significant discrepancy between the amount of training raters received in using PACS measurement tools and the amount of training in interpretation of VFSS. Although it is not clear what the impact of this difference in experience was, it is possible that results were unintentionally biased toward the subjective visual grading method. It would be interesting to repeat the study using novice raters with little training in either method.

A combination of measurements from lateral and anterior plane stills was considered as a study

design option but would have been difficult to do. Patients would have had to hold their head position for an extended time while equipment rotated. Several adjustments to account for magnification would be needed, and patients would not have been allowed to swallow between stills. This would limit its applicability in the clinical setting.

Further multicenter reliability studies are required using both qualified and unqualified raters to ascertain the true impact of VRR in grading valleculae residue. If reliability proved to be higher among novice raters for the VRR, it might suggest that there is an application for using PACS as a teaching tool for VFSS reporting.

On the basis of this study alone it is unclear whether PACS gives clinicians a reliable method of monitoring valleculae residue. However, unlike scintigraphy, it allows clinicians to study a sequence of VFSS stills thereby giving a realistic evaluation of swallowing performance (average VRR). With further appraisal and development of the VRRS and VRR, it may no longer be important what constitutes mild, moderate, or severe residue (subjective and inaccurate descriptors) and where a cutoff point lies because a numeric score could be recorded in a videofluoroscopy report.

There may be potential for the VRRS and VRR to be used as a research tool. Researchers and clinicians have traditionally been deterred from attempting to take quantitative measurements of pharyngeal residue in conjunction with VFSS because it is impossible to determine exact volumes based on two-dimensional images [20, 21]. However, the most pertinent question is not the exact volume of residue in the valleculae, but the ratio or qualitative level of residue in proportion to the size of this individual's valleculae. The risk of aspirating from a vallecula that has a high proportion of residue is greater than where a low proportion of residue exists. Because the size of the population that is 65 years old and over continues to expand, it will become increasingly important that clinicians have a clear understanding of the differences between severity of pharyngeal residue in healthy and dysphagic swallows. Up to 20% of elderly asymptomatic individuals have residue in the valleculae and pyriform sinuses [28]. More normative data are required on pharyngeal retention in the elderly [29]. The VRR has a potential role in helping clinicians be more objective in what constitutes a normal level of valleculae residue according to age, bolus size, and consistency. Limited retention in the valleculae and/or pyriform sinuses in the elderly is considered normal and is not associated with aspiration [16]. Important factors that increase the risk of elderly dysphagic patients acquiring pneu-

monia include volume of aspiration and colonization of the oropharynx with gram-negative pathogens and *Streptococcus aureus* [30].

#### *Uptake of PACS in Healthcare*

While our study could in principle have been carried out without the assistance of a computer, PACS is fast becoming the standard framework of filmless radiology. All United Kingdom acute care hospital trusts should have acquired the system by 2006 [27]. This is the first study to investigate the use of objective measurements as provided by PACS in grading valleculae residue.

The minimum cost of a G.E. Basic PACS unit, which includes one workstation, a database, and a small storage unit, is approximately \$193,220. It is therefore clear that justification for purchasing this expensive equipment could not be made based solely on its ability to measure valleculae residue. It is likely that the addition of basic PACS units to all hospitals will improve accessibility, storage, and communication between professionals regarding a patient's radiologic record, which will contribute to better patient care.

On the basis of this study the authors have chosen to continue to use the traditional method of visually estimating valleculae residue based on the high level of accuracy that the experienced raters in the department achieved. If, however, the group of raters changed to include staff with less VFSS experience, we would use the VRR scale. Further studies are necessary to establish whether this system reduces the training time of novice raters with no loss of accuracy or repeatability.

#### *Limitations of the Study*

This study concentrated solely on pharyngeal retention in the valleculae and has not attempted to measure diffuse pharyngeal residue or pyriform sinus residue, both of which contribute to overflow aspiration [6]. Recent studies in the orthopedic field [31] have used PACS interfaced with a digitizing software program that prompts the user to select several anatomical landmarks. It then uses these landmarks to create on-screen coordinates and calculates reliable radiographic measurements using trigonometric and algebraic algorithms. A similarly complex method of calculation would be required to give a total measure of pharyngeal retention specific to pharyngeal regions, which was beyond the scope of this study.



## Conclusion

In the absence of reliable and compatible measurement tools, research will continue to challenge the premise that videofluoroscopy is the “gold standard” of swallowing assessment [12, 13]. Successful dysphagia management requires the development and scientific validation of measures of critical variables. In order that dysphagia research and evidence-based treatments continue to evolve, it is essential that potential new methods of grading the function of individual components of the swallow be developed and thoroughly tested. We believe the VRR scale is such a method.

*Acknowledgments.* The authors thank the 20 patients who volunteered to participate in this study; the five Speech and Language Pathologists who rated the stills; the five designated radiographers who captured the videofluoroscopy images; Dr. DeSai, Consultant Radiologist, Val Gleason, Superintendent Radiographer, and Mark Brennan, PACS technician, for their support throughout this challenging study; and Linda Templey, Director of Primary and Community Care, Durham and Chester-le-Street PCT (now assimilated into County Durham PCT) for agreeing to sponsor this research.

## References

- Barczy SR, Sullivan P, Robbins J: How should dysphagia care of older adults differ? Establishing optimal practice patterns. *Semin Speech Lang 21(4)*:347–361, 2000
- Lin L, Wu S, Chen H, Wang T, Chen M: Prevalence of impaired swallowing in institutionalized older people in Taiwan. *J Am Geriatr Soc 50*:1118–1123, 2002
- Logemann JA: *Manual for the videofluorographic study of swallowing*. Austin, TX: Pro-Ed, 1993
- Veis S, Logemann JA, Colangelo MS: Effects of three techniques on maximum posterior movement of the tongue base. *Dysphagia 15*:142–145, 2000
- Logemann JA: *Evaluation and treatment of swallowing disorders*. Austin, TX: Pro-Ed, 1998
- Eisenhuber E, Schima W, Schober E, Pokieser P, Stadler A, Scharitzer M, Oschatz E: Videofluoroscopic assessment of patients with dysphagia: pharyngeal retention is a predictive factor for aspiration. *AJR Am J Roentgenol 178*:393–398, 2002
- Centers for Disease Control and Prevention: Pneumonia and influenza death rates – United States. *MMWR Morb Mortal Wkly Rep 44*:535–537, 1995
- Niederman MS, McCombs JS, Unger AN, Kumar A, Popovian R: The cost of treating community-acquired pneumonia. *Clin Therap 20*:820–837, 1998
- Palmer JB, Kuhlemeier KV, Tippet DC, Lynch C: A protocol for the videofluorographic swallowing study. *Dysphagia 8*:209–214, 1993
- Ekberg O, Nylander G, Frans-Thomas F, Sjoberg S, Birch-Jensen M, Hillarp B: Interobserver variability in cineradiographic assessment of pharyngeal function during swallow. *Dysphagia 3*:46–48, 1988
- Wilcox F, Liss JM, Siegel GM: Interjudge agreement in videofluoroscopic studies of swallowing. *J Speech Hear Res 39*:144–152, 1996
- McCullough GH, Wertz R, Rosenbek JC, Mills RH, Webb WG, Ross KB: Inter- and intrajudge reliability for videofluoroscopic swallowing evaluation measures. *Dysphagia 16*:110–118, 2001
- Stoeckli SJ, Thierry AGM, Seifert B, Martin-Harris BJW: Inter-rater reliability of videofluoroscopic swallow evaluation. *Dysphagia 18*:53–57, 2003
- Kuhlemeier KV, Yates P, Palmer J: Intra- and inter-rater variations in the evaluation of videofluorographic swallowing studies. *Dysphagia 13*:142–147, 1998
- Ott D: Observer variation in evaluation of videofluoroscopic swallowing studies: a continuing problem [editorial]. *Dysphagia 13*:148–150, 1998
- Dejaeger E: Mechanisms involved in postdeglutition retention in the elderly. *Dysphagia 12*:63–67, 1997
- Rosenbek JC, Robbins JA, Roecker EB, Coyle JL, Wood JL: A penetration-aspiration scale. *Dysphagia 11*:93–98, 1996
- Galli J, Valenzano V, Ottaviani F, D’Alatri L, Almadori G, D’Errico G, Serra F, De Rossi G, Paludetti G: Scintigraphic assessment of swallowing after horizontal supraglottic laryngectomy. *J Otolaryngol 28*:190–196, 1999
- Shaw DW, Williams RBH, Cook IJ, Wallace KL, Weltman MB, Collins PJ, McKay E, Smart R, Simula ME: Oropharyngeal scintigraphy: a reliable technique for the quantitative evaluation of oral-pharyngeal swallowing. *Dysphagia 19*:36–42, 2004
- Hind JA, Nicosia MA, Roecker EB, Carnes ML, Robbins J: Comparison of effortful and non-effortful swallows in healthy middle-aged and older adults. *Arch Phys Med Rehabil 82*:1661–1665, 2001
- Han T, Nam-Jong P, Park JW: Quantifying swallowing function after stroke: A functional dysphagia scale based on videofluoroscopic studies. *Arch Phys Med Rehabil 82*:677–682, 2001
- International Society of Optical Engineering: Proceeding of the first international conference and workshop on picture archiving communication systems. January 18th–21st (1982), Newport Beach, California, Published by Bellingham, Washington
- Huang HK: *PACS: basic principles and applications*. New York: Wiley-Liss, 1999
- Friedman JH: Regularized discriminant analysis. *J Am Stat Assoc 84*:165–175, 1989
- Landis J, Koch G: The measurement of observer agreement for categorical data. *Biometrics 33*:159–174, 1977
- Mathers-Schmidt BA, Kurlinski M: Dysphagia evaluation practices: inconsistencies in clinical assessment and instrumental examination decision-making. *Dysphagia 18*:114–125, 2003
- NHS National Programme for Information Technology: Making IT work. 1:15, 2004 (available at: <http://www.dh.gov.uk/assetRoot/04/07/71/57/04077157.pdf>)
- Ekberg O, Feinberg M: Altered swallowing function in elderly patients without dysphagia: radiologic findings in 56 cases. *AJR Am J Roentgenol 156*:1181–1184, 1991
- Cook IJ, Weltman MD, Wallace K, Shaw DW, McKay E, Smart RC, Butler SP: Influence of ageing on oral-pharyngeal bolus transit and clearance during swallowing: scintigraphic study. *Am J Physiol 266*:G972–G977, 1994
- Marik PE, Kaplan D: Aspiration pneumonia and dysphagia in the elderly. *Chest 124(1)*:328–336, 2003
- Pedersen DR, Lamb CA, Dolan L, Ralston H, Weinstein S, Morcuende J: Radiographic measurements in development dysplasia of the hip. *J Pediatr Orthop 24*:156–160, 2004