Swallowing disorders are common in patients after surgery for head and neck cancer. The clinical assessment of oral and laryngopharyngeal abilities is widely used as a dysphagia assessment tool in this patient group, despite a lack of research. The goal of this study was to assess the predictability of clinical parameters for aspiration and limitation of oral intake. A swallowing disorder with the need for further intervention was identified by fiberoptic endoscopic evaluation of swallowing (FEES) in 65%, with aspiration in 49%, silently in 21%, and limited oral intake with tube dependency in 56% of studied patients. Four clinical parameters (dysglossia, wet voice, tongue motility, and tongue strength) correlated significantly with aspiration and limitation of oral intake. However, none of these clinical parameters was able to predict one of our two reference criteria, due to low positive likelihood ratios, mostly less than two. Clinical assessment is therefore inappropriate for early detection of swallowing disorders in such patients.

Swallowing disorders are still a common sequelae of tumor resection in patients with head and neck cancer (HNC) (1), despite improved modern resection and reconstruction procedures (2-5). The prevalence rate is reported to be up to 88% (6-8) depending on localization, tumor size, and chosen treatment. Due to the frequent occurrence of post-surgical biomechanical dysfunctions, the examination of the swallowing-related oral and laryngopharyngeal abilities in this patient group is widely used as a dysphagia assessment tool, irrespective of the lack of research concerning its clinical utility.

Today, research is largely focused on the examination of non-swallowing oral and laryngeal functions as clinical parameters due to the possibility of predicting swallowing problems, predominantly in patients after stroke. The most frequently reported parameters are dysarthria (9-15), wet voice (10, 11, 12, 16), limited tongue motility (17, 18) and strength (19, 20), abnormal volitional cough (10, 11, 12, 21-23), and absent gag reflex (10-12, 22, 24, 25) as more or less sensitive signs of aspiration, are the most critical aspect of swallowing disorders in patients after stroke. Limitation of oral intake as a typical result of impaired bolus propulsion is not rare post-surgically in patients with HNC as a sign of severe oropharyngeal dysphagia even without signs of aspiration. It is of interest whether these clinical parameters, together with limitations in mouth opening, often occurring in patients with HNC (26), can be predictive of swallowing disorders with the need for further intervention in patients after surgery for HNC.

Therefore, the goal of the present study was to identify in such patients: (a) the predictability of the clinical parameters dysglossia, wet voice, abnormal volitional cough, reduced mouth opening, limited tongue motility, limited tongue strength, and gag reflex for the reference criteria aspiration and limitation of oral intake defined by fiberoptic endoscopic evaluation of swallowing (FEES); and (b) whether the combination of significantly correlating parameters enhances the predictability of these disorders.

Patients and Methods

Patients. Between November 2010 and December 2012, 98 patients were recruited for the presented prospective study after surgery for HNC. Inclusion criteria were defined as Union internationale contre
le cancer (UICC) stage II–IV, age between 18 and 99 years, and written informed consent. Exclusion criteria were neurological diseases and pre-existing oropharyngeal dysphagia. Out of the 98 patients, five were excluded due to neurological diseases, eight due to pre-existing oropharyngeal dysphagia, and five due to refused consent (Figure 1).

A total of 80 patients were thus included in our study, predominantly men (58 male, 22 female), with a mean age of 60.96 years (SD=12.93 years; range=18.92 to 87.75 years). Refer to Table I for patient characteristics.

The study was approved by the Institutional Review Board of the University Hospital Frankfurt/Main, Germany (approval # 240/10).

Procedure. After tumor resection and surgeon’s approval for oral intake, all included patients firstly underwent clinical swallowing assessment performed by one of two speech and language pathologists (SLP), both with over five years’ experience in dysphagia evaluation and treatment. As a reference standard, fiberoptic endoscopic evaluation of swallowing (FEES, Langmore standard) (27) was performed with a time frame of one hour after the clinical swallowing assessment. The examiner and rater of the FEES was a phoniatrician with over 10 years’ FEES experience, blinded to information regarding the results of the two SLPs.

Clinical swallowing assessment. The clinical swallowing assessment, including dysglossia, wet voice, volitional cough, range of mouth opening, tongue motility, tongue strength, and gag reflex, was standardized and binarily scored with a pass/fail judgment (see Table II). Both SLPs were trained in administering and interpreting the seven clinical parameters on five healthy subjects and 20 patients with HNC who were not included in the study to optimize test reliability. Confirming inter- and intra-rater reliability tests was performed on 24 additional patients with HNC. The entire clinical swallowing assessment was administered and scored in less than 10 minutes.

Reliability. The overall intra-rater reliability (Cohen’s \( \kappa =1.00, p<0.001 \)) and the overall inter-rater reliability (Cohen’s \( \kappa =0.98; p<0.001 \)) were both excellent.

FEES. The FEES examinations were performed by the first author Christina Hey according to the FEES protocol by Langmore et al. (27). The recordings were taken with the ENT video endoscopy system EndoStrob-DX (Xion medical GmbH, Berlin, Germany) (video system: PAL; output: S-Video, DVI-I, VGA IEEE/1394 (DV); input DVI-I, S-Video; 3 lux; sensor: CCD matrix with micro lenses and mosaic filter, 1/3” 752 (h) x 582 (v) pixels; digital interfaces: IEEE/1394 FireWire DV; light source: power 50 W; color temperature 5.700 K; camera head with integrated microphone, videoadapter: focal length 22 mm; clip coupling) and the transnasal flexible endoscope 11101 RP2 (Karl Storz GmbH, Tuttlingen, Germany).

The examination was initiated with 2 ml water and followed accordingly with 5, 10, and 20 ml. In addition, pure and solid consistencies with progressive volume were tested. If a patient aspirated, without the possibility of improving swallowing ability by therapeutic maneuvers, the FEES examination was discontinued.

The level of aspiration was scored by the eight-point penetration-aspiration scale (PAS) by Rosenbek et al. (28), the level of oral intake by the seven-point functional oral intake scale (FOIS) (29). Dysphagia with the need for further intervention was defined by a PAS ≥4 (4=material enters the airway, contacts the vocal folds, and is ejected from the airway), or FOIS ≤4 (4=total oral diet of a single consistency), aspiration by a PAS ≥6 (6=material enters the airway, passes below the vocal folds, and is ejected into the larynx or out of the airway), and therapy-relevant limitation of oral intake by FOIS ≥4.

FEES reliability. Assessments were rated by the first author Christiane Hey, who also re-rated 20% of the digital clips at least two months after the initial examination in order to assess the intra-rater reliability. Ten percent of the recordings were rated by the second person with experience in FEES, who was blinded to the

Table I. Patients’ characteristics.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>n</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤40</td>
<td>4</td>
<td>5.0</td>
</tr>
<tr>
<td>41-60</td>
<td>34</td>
<td>42.4</td>
</tr>
<tr>
<td>&gt;60</td>
<td>42</td>
<td>52.5</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>58</td>
<td>72.5</td>
</tr>
<tr>
<td>Female</td>
<td>22</td>
<td>27.5</td>
</tr>
<tr>
<td>Tumour stage</td>
<td></td>
<td></td>
</tr>
<tr>
<td>II</td>
<td>23</td>
<td>28.8</td>
</tr>
<tr>
<td>III</td>
<td>9</td>
<td>11.3</td>
</tr>
<tr>
<td>IV</td>
<td>48</td>
<td>60.0</td>
</tr>
<tr>
<td>Tumour site</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cavitas oris</td>
<td>30</td>
<td>37.5</td>
</tr>
<tr>
<td>Oropharynx</td>
<td>34</td>
<td>42.5</td>
</tr>
<tr>
<td>Hypopharynx/larynx</td>
<td>16</td>
<td>20.0</td>
</tr>
</tbody>
</table>

Figure 1. Study inclusion flow diagram.
patients’ details, to assess the inter-rater reliability. The overall intra-rater reliability (Cohen’s $\kappa=0.89$, $p<0.001$) and the overall inter-rater reliability (Cohen’s $\kappa=0.92$, $p<0.001$) were both excellent.

Statistical analyses. The prevalence of dysphagia as well as both reference criteria aspiration and limitation of oral intake in our study population were determined by descriptive statistics. Direction and strength of relationships with both reference criteria were tested using Pearson’s correlation.

An analysis of variance (ANOVA) was calculated to test the effects of tumor site and stage.

For the significantly correlating clinical parameters, the accuracy was determined by sensitivity, specificity, and the positive likelihood ratio (LR).

All statistical analyses were performed using SPSS (Version 20, International Business Machines Corp., Armonk, New York, USA).

Results

FEES results. A swallowing disorder with the need for further intervention was identified by FEES in 65.0% (52/80), with aspiration in 48.8% (39/80), silently in 21.3% (17/80), and limitation of oral intake with tube dependency in 56.3% (45/80) of all patients with HNC. A level of 4 in the FOIS did not occur in any of the patients.

The more advanced the tumor stage, the higher the rate of aspiration and tube dependency. Refer to Table III for prevalence, as well as tumor stage and site-related details.

Regarding the tumor site, the patients with oropharyngeal carcinoma presented therapy-relevant dysphagia in 73.5% (25/34) aspiration in 61.8% (21/34), and tube dependency in 67.7% (23/34) of cases. However, as revealed by ANOVA, the tumor site had no effect on the two reference criteria ($p>0.4$; $\eta^2_p<0.03$), whereas tumor stage influenced both aspiration ($p=0.005$; $\eta^2_p=0.138$) and limitation of oral intake ($p=0.010$; $\eta^2_p=0.122$).

Correlation of the clinical parameters with aspiration, oral intake, and therapy-relevant dysphagia. Three out of seven clinical parameters, namely mouth opening, volitional cough, and gag reflex, showed no noteworthy correlation with either reference criterion ($r\leq0.2$; $p>0.05$) and were thus considered negligible (see Table IV).

The remaining four parameters, namely dysglossia, wet voice, tongue motility, and tongue strength, were identified to be significantly related to limitation of oral intake and, except for dysglossia, with aspiration (Table IV).

The raw data, sensitivity, specificity, and the positive LR are listed for these four clinical parameters and their combinations in Table V.

The highest sensitivity for aspiration and limitation of oral intake was identified for tongue strength (aspiration=79.49%, limitation of oral intake=75.56%) and tongue motility (aspiration=74.36%, limitation of oral intake=71.11%) but
Aspiration and limitation of oral intake.

For the clinical assessment before commencement of the study, we standardized the examination and intake in patients after surgery for HNC. To enhance the laryngeal functions for aspiration and limitation of oral intake in patients following (chemo-)radiation (30-32).

While tumor site did not influence the prevalence rate, patients with HNC have to be defined as a high-risk group for developing dysphagia and aspiration. The high prevalence of aspiration (49%), limitation of oral intake (56%), and swallowing disorder with need for further intervention (65%) in our study underlines that post-surgically patients with HNC have to be defined as a high-risk group for developing dysphagia and aspiration. Noteworthy is the rate of silent aspiration, over 20% of our patients, which is as high as that reported normally for German patients with HNC (38), four clinical parameters, namely dysglossia, wet voice, tongue motility, and tongue strength, correlated significantly with the FEES-detected signs of dysphagia; the three others, namely reduced mouth-opening, abnormal volitional cough, and absent gag reflex, did not. The two latter are also discussed controversially for patients after stroke, and categorically denied among others (23, 39, 40-42).

In our study, tongue motility and tongue strength presented the highest sensitivity, with values from 71% to 80% in all target conditions, therefore these two parameters could be interpreted as possible predictors. However, ruling a disease for subsequent management or not, has to result in a comparison of pre- and post-test probability instead of considering sensitivity and specificity in isolation as is widely done (43). The LR is a statistic quality criterion that is more useful clinically than other statistical criteria (44). It converts the pre-test into post-test probability by combining sensitivity and specificity. In contrast to predictive values it is prevalence-independent and indicates how more or less probable patients with aspiration or limitation of oral intake are to have a positive test result within our clinical parameters than patients without the target condition. A positive LR value of 1 demonstrates that the probability of having a specific test result is equal for patients both with and without the target condition, which points out that such a test result is not able to differentiate between patients with and without the target condition. LR of 1 to 2 alters probability to a small and rarely important degree; LR of 2 to 5 generates small changes in probability; LR of 5 to 10 generates moderate shifts in pre- and post-test probability (45). Given this, none of our four significantly correlating clinical parameters was able to predict aspiration or limitation of oral intake sufficiently, with values of LR mostly being less than two. Even if wet voice demonstrates a remarkably high specificity (90% to 94%) and good LR (4 to 7), because specificity influences the positive LR (46), wet voice does not reflect a valid predictor due to its low sensitivity.

Combining reliable clinical parameters is proposed in order to enhance validity over single items for the better predictability of aspiration in patients after stroke (10-12, 17, 23). Daniels et al. revealed that the presence of two out of six clinical parameters (dysarthria, dysphonia, abnormal volitional cough, abnormal gag reflex, cough or throat clearing and voice change after swallow) validly predicts aspiration in patients.

Discussion

The high prevalence of aspiration (49%), limitation of oral intake (56%), and swallowing disorder with need for further intervention (65%) in our study underlines that post-surgically patients with HNC have to be defined as a high-risk group for developing dysphagia and aspiration. Noteworthy is the rate of silent aspiration, over 20% of our patients, which is as high as that reported normally for patients with HNC following (chemo-)radiation (30-32). While tumor site did not influence the prevalence rate, increasing tumor stage extended the risk of aspiration and limitation of oral intake. This corresponds to the results of previous research showing that the larger the tumor size and resection, the greater the signs of dysphagia (30, 33-36).

The goal of our study was to examine the predictability of seven clinical parameters of non-swallowing oral and laryngeal functions for aspiration and limitation of oral intake in patients after surgery for HNC. To enhance the reliability of testing, we standardized the examination and simplified interpretation with binary pass/fail decisions. We also trained two SLPs in the administration and judging of the clinical assessment before commencement of the study.

As a result, the intra- and inter-rated reliability was excellent in our study when compared to studies in which prior training was not applied (23), which underscores the necessity for standardization and training in order to achieve unified and valid examinations of clinical parameters as recommended in several articles (17, 18, 37).

In our study population, which is representative insofar as it matches the prevalence rate of German patients with HNC (38), four clinical parameters, namely dysglossia, wet voice, tongue motility, and tongue strength, correlated significantly with the FEES-detected signs of dysphagia; the three others, namely reduced mouth-opening, abnormal volitional cough, and absent gag reflex, did not. The two latter are also discussed controversially for patients after stroke, and categorically denied among others (23, 39, 40-42).

In our study, tongue motility and tongue strength presented the highest sensitivity, with values from 71% to 80% in all target conditions, therefore these two parameters could be interpreted as possible predictors. However, ruling a disease for subsequent management or not, has to result in a comparison of pre- and post-test probability instead of considering sensitivity and specificity in isolation as is widely done (43). The LR is a statistic quality criterion that is more useful clinically than other statistical criteria (44). It converts the pre-test into post-test probability by combining sensitivity and specificity. In contrast to predictive values it is prevalence-independent and indicates how more or less probable patients with aspiration or limitation of oral intake are to have a positive test result within our clinical parameters than patients without the target condition. A positive LR value of 1 demonstrates that the probability of having a specific test result is equal for patients both with and without the target condition, which points out that such a test result is not able to differentiate between patients with and without the target condition. LR of 1 to 2 alters probability to a small and rarely important degree; LR of 2 to 5 generates small changes in probability; LR of 5 to 10 generates moderate shifts in pre- and post-test probability (45). Given this, none of our four significantly correlating clinical parameters was able to predict aspiration or limitation of oral intake sufficiently, with values of LR mostly being less than two. Even if wet voice demonstrates a remarkably high specificity (90% to 94%) and good LR (4 to 7), because specificity influences the positive LR (46), wet voice does not reflect a valid predictor due to its low sensitivity.

Combining reliable clinical parameters is proposed in order to enhance validity over single items for the better predictability of aspiration in patients after stroke (10-12, 17, 23). Daniels et al. revealed that the presence of two out of six clinical parameters (dysarthria, dysphonia, abnormal volitional cough, abnormal gag reflex, cough or throat clearing and voice change after swallow) validly predicts aspiration in patients.

Table IV. Pearson’s correlation of all clinical parameters with aspiration and limitation of oral intake.

<table>
<thead>
<tr>
<th>Clinical Parameter</th>
<th>Aspiration (PAS ≥6)</th>
<th>Limitation of oral intake (FOIS ≥4)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(r)</td>
<td>(r)</td>
</tr>
<tr>
<td>Dysglossia</td>
<td>0.20</td>
<td>0.37**</td>
</tr>
<tr>
<td>Wet voice</td>
<td>0.38**</td>
<td>0.45**</td>
</tr>
<tr>
<td>Volitional cough</td>
<td>0.01</td>
<td>0.09</td>
</tr>
<tr>
<td>Mouth opening</td>
<td>0.11</td>
<td>0.20</td>
</tr>
<tr>
<td>Tongue motility</td>
<td>0.28*</td>
<td>0.35**</td>
</tr>
<tr>
<td>Tongue strength</td>
<td>0.28*</td>
<td>0.33**</td>
</tr>
<tr>
<td>Gag reflex</td>
<td>0.13</td>
<td>0.12</td>
</tr>
</tbody>
</table>

df=79 for all correlations; *p<0.05; **p<0.01.
after stroke (Sensitivity=69.6%, Specificity=84.4%, LR=4.46) (11). McCullough et al. detected a better validity with the presence of four rather than two of the same clinical parameters (Sensitivity=77.8%, Specificity=71.9%, LR=2.76) (23). Leder et al., however, denied that these six clinical parameters are able to predict aspiration in patients after stroke (Sensitivity=86.4%, Specificity=29.6%, LR=1.23) (40). In contrast to the former studies, Leder et al. used FEES instead of the Videofluoroscopic Swallowing Study (VFSS) as the reference gold standard. In our study, the combination of the four correlating parameters (dysglossia, wet voice, tongue motility, and tongue strength), as a combination of two or more failed clinical parameters (aspiration: Sensitivity=71.0%, Specificity=61.2%, LR=1.83; limitation of oral intake: Sensitivity=67.7%, Specificity=71.4%, LR=2.37), and as a combination of three or more failed clinical parameters (aspiration: Sensitivity=63.6%, Specificity=63.9%, LR=1.76; limitation of oral intake: Sensitivity=59.1%, Specificity=75.0%, LR=2.36), did in fact improve the test properties when compared to the validity of each single parameter but was still not convincing.

However, beside the obvious fact that these four clinical parameters are really not able to predict aspiration and limitation of oral intake, a possible reason for this result, compared to the studies by Daniels et al. and McCullough et al., is that we had only four rather than six clinical parameters to combine, and used FEES as the gold standard like Leder et al. Furthermore, we used only seven clinical parameters, although the ones most reported, and even if our patient group is representative of the prevalence rate for German patients with HNC, the patient group is too small to carry out calculations for different tumor sites.
43 Boyko EJ: Ruling out or ruling in disease with the most sensitive or specific diagnostic test: Short cut or wrong turn? Med Decis Making 14: 175-179, 1994.

Received May 29, 2013
Revised June 27, 2013
Accepted July 2, 2013