CLINICAL LINGUISTICS & PHONETICS, 2003, VOL. 17, NO. 4–5, 411–420 Taylor & Francis healthsciences

Speech evaluation and swallowing ability after intra-oral cancer

K. MÁDY†, R. SADER†, P.H. HOOLE‡, A. ZIMMERMANN† and H-H. HORCH†

† Department of Oral and Maxillofacial Surgery, University of Technology, Munich
‡ Department of Phonetics and Linguistic Communication, Ludwig Maximilian University, Munich

Clin Linguist Phon Downloaded from informahealthcare.com by UB der LMU Muenchen on 04/05/11
For personal use only.

Abstract

In this study, four methods of postoperative speech evaluation are compared for 19 persons with oral cancers who have undergone oral surgery and/or radiotherapy. The Munich Intelligibility Profile was used for intelligibility testing and semiquantitative scoring by novice listeners. Expert ratings were done on the Therapy Outcome Measure (TOM) Phonological Disability form. For self-evaluation, the EORTC QLQ-C30 and the Head and Neck module was used. Swallowing function was scored on the TOM Dysphagia form. There was a high intercorrelation between the results of subjective speech evaluation by experts and non-experts and the intelligibility test, but no correlation with any of these methods could be shown for the self-evaluation by the participants. Voice quality seemed to have an influence on non-expert scores. Swallowing was a more severe problem for our group than speech impairment.

Keywords: Speech evaluation, swallowing, glossectomy.

Introduction

The quality of life (QoL) of patients with oral cancers is often decreased. Problems arise especially regarding speech ability, nutrition and appearance (Deleyiannis, Weymuller and Coltrera, 1997, Gellrich, Schramm, Böckmann and Kugler, 2002). As all three factors are crucial for the social life of every individual and are easily noticed, patients with orofacial cancers have considerable difficulties in maintaining their social role and social activities. Therefore, investigations regarding this group

Address correspondence to: Katalin Mády, Department of Oral and Maxillofacial Surgery, Klinikum rechts der Isar der TU München, Ismaninger Str. 22, D-81675 München, Germany. e-mail: mady@mkg.med.tu-muenchen.de

Clinical Linguistics & Phonetics
ISSN 0269-9206 print/ISSN 1464-5076 online © 2003 Taylor & Francis Ltd
http://www.tandf.co.uk/journals
DOI: 10.1080/0269920031000079921

RIGHTSLINK

of patients often focus on speech and nutrition/swallowing. These abilities are often concerned when the functional outcome of different treatment methods is to be compared (Schliephake, Schmelziesen, Schönweiler, Schneller and Altenbernd, 1998; Pauloski, Logemann and Rademaker, 1993; Pauloski, Rademaker, Logemann and Colangelo, 1998). In other studies, postoperative speech ability has been related to the extent of surgery (Rentschler, Arbor and Mann, 1980; Mackenzie Beck, Wrench, Jackson, Soutar, Robertson and Laver, 1998).

The methods used for speech assessment are heterogeneous: they range from evaluation of single phonemes (Mackenzie Beck *et al*., 1998, articulation test in Pauloski *et al*., 1993) or words (Schliephake *et al*., 1998) on semiqualitative scales and identification of the spoken word to the testing of communicative intelligibility in questions, descriptions (Rentschler *et al*., 1980) or longer text passages (Pauloski *et al*., 1993, 1998). Rentschler *et al*. noted that intelligibility scores were largely dependent on the task type: in their setting, single word identification lead to an intelligibility percentage of 46.6%, whereas questions read aloud by the glossectomized patients were correctly understood in 82.3% of the cases.

We felt it necessary to run a parallel assessment of the speech ability from several angles: (1) how a glossectomized person sounds to his/her communication partners, (2) how he/she is evaluated by a phonetician/speech therapist, and (3) to what percentage she/he is intelligible under context-free testing. Our survey also included data collection on the self-assessment of these patients of their own speech.

Clin Linguist Phon Downloaded from informahealthcare.com by UB der LMU Muenchen on 04/05/11
For personal use only.

Materials and methods

Subjects

Nineteen persons with tumours of the tongue, floor of mouth and/or the mandible were investigated postoperatively. They were aged between 47 and 78 years (mean 58.5, SD 7.8), 17 were males and two females. The largest group underwent only surgery ($n = 16$), two had surgical therapy with adjuvant radiotherapy, one participant received only radiotherapy. Table 1 shows the distribution of subjects regarding tumour localization, tumour stage, and age.

For comparison purposes, ten persons with oral cancer (four of whom also took part in the postoperative testing) were recorded preoperatively. In spite of the presence of a malign change in their oral cavity, none of them had any speech impairment at the time of the recordings.

Speech assessment

Intelligibility measurement

For objective intelligibility, the computer-based Munich Intelligibility Profile (Ziegler, Hartmann and Wiesner, 1992, Ziegler and Hartmann, 1993) was used. A quantitative intelligibility measurement is provided by a multiple choice test where the listener (a linguistically untrained person who is not familiar with distorted speech) is presented with 72 items in isolation or embedded in context-free carrier sentences. After listening to each item, she/he is shown a list of 12 items: the target, plus five items differing from the target by one phonological feature, plus six words differing from the target in two phonological features. The oppositions are based on features like place of articulation (labial vs. apical vs. dorsal), manner of

RIGHTSLINK

Table 1. *List of patients including tumour localization, tumour stage, tumour therapy (surgery and/or radiation therapy), and age (T: tongue, FM: floor of mouth, MB: mandible)*

p#/sex	tumour stage	tumour site	therapy	age
1/m	2	T	surg+rad	47
2/m	2	T+FM	surg	79
3/m	1	FM	surg	54
4/m	1	FM	surg	46
5/m	4	T+FM	rad	54
6/m	1	FM	surg	51
7/m	2	T	surg	67
8/m	2	FM	surg	66
9/m	2	T	surg	51
10/m	4	T	surg	57
11/m	4	MB+FM	surg	60
12/m	2	T	surg	56
13/m	1	FM	surg	54
14/m	1	T+FM	surg	63
15/f	1	T	surg+rad	62
16/m	2	FM	surg	61
17/m	1	MB+FM	surg	60
18/f	1	MB+FM	surg	59
19/m	2	FM	surg	64

articulation (stop vs. fricative vs. nasal), etc., so that all erroneous identifications indicate problematic articulatory features. As reported by Ziegler and Hartmann, all reliability coefficients (r) for MIP are higher than 0.9.

Despite the careful randomization of targets and carrier sentences (from a set of more than 2800 items, respectively), a slight learning effect has been observed in a test-retest setting (Ziegler and Hartmann, 1993), i.e., in the second testing, false identifications occurred less frequently ($Diff_{mean} = 4.4\%$). Therefore, half of our listeners (2 of 4) evaluated the recordings in a reversed order.

Non-expert evaluation

The MIP also includes a semiquantitative scoring for speech impairment (SI), clearness of articulation (CA) and for voice on a scale from 0 to 6 (0 meaning 'no impairment', 6 'severe impairment'), based on a sample of ten sentences. Unlike the items of the intelligibility test, the sentences are identical for each session. The evaluation was conducted by the four listeners who participated in the intelligibility test. No detailed information was given regarding the recordings, i.e., the listeners did not know whether they were listening to a pre- or postoperative recording.

Self-assessment

Participants were requested to complete the EORTC QLQ-C30 survey (European Organization for Research and Treatment of Cancer, Quality of Life Questionnaire, Version. 3.0) and the Head and Neck Module (Fayers, Aaronson, Bjordai and Sullivan, 1997). In this survey, questions like 'Did you have trouble talking on the telephone during the last week' are to be answered on a four-item-scale ('not at all' to 'very much'). Beside the speech module (three questions), the test covers several aspects of nutrition (swallowing, chewing, etc.), social functioning, pain and also

Clin Linguist Phon Downloaded from informahealthcare.com by UB der LMU Muenchen on 04/05/11
For personal use only.

provides a general score for global health and QoL. All calculated reliability coefficients of EORTC QLQ-C30 are higher than 0.70 (Bjordal, de Graeff, Fayers, *et al*., 2000).

Regarding the QoL testing, the compliance of our subjects was low. Some of the participants reported that they felt the questionnaire to be too direct. While some of them chose to ignore questions regarding their sexual life or financial situation, others refused to fill in any part of the form, leaving us with only six participants in this part of the study.

Expert ratings

Expert ratings of the articulatory disability were done by a trained phonetician on the Therapy Outcome Measure (TOM) questionnaire for 'Phonological Disorder' (Enderby and John, 1997). The scoring from 0 ('very poor') to 5 ('no impairment'), also allowing 0.5 scores, is based on functional speech ability, i.e., to what extent the patient is intelligible to familiar and unfamiliar listeners. As reported by John and Enderby (2000), inter- and intra-personal reliability of this test for speech and language therapists is higher than $r = 0.80$ in nearly all settings (and never lower than $r = 0.71$).

Swallowing

Participants were scored on the TOM Dysphagia (Disability) questionnaire. The figures relied on information collected from the patients or given by themselves in the EORTC H&N Module. These results were crosschecked by videofluoroscopy diagnoses.

Results

Speech assessment

In table 2, error rates, non-expert evaluation and expert ratings for speech and swallowing are summarized. The semiqualitative expert and non-expert ratings ranged widely between the minimum and maximum pole of the scale. On the other hand, error rates in the intelligibility test clustered at the low end of the scale: only two subjects, 9 and 10, had an error rate above 20%, despite the fact that no contextual information was provided. Voice scores were better than SI and CA for nearly all participants.

Table 3 shows the correlation coefficients between our measures (Spearman Rho for nonparametric data). There is an overall strong correlation between error rates, expert and non-expert ratings for speech and articulation. In figures 1, 2 and 3, intelligibility, non-expert and expert ratings are shown. CA was chosen as a counterpart of the expert ratings because of its high correlation with the expert scores. Figure 1 shows a strong correlation between the subjective scores. Note that the same error rate may be coupled with varying CA and expert ratings (figures 2 and 3), especially for error rates under 10%.

Intelligibility in subjects without speech impairment lies in the range of 96–100% (Ziegler *et al*., 1992), i.e., error rates less than 4.0% do not necessarily reflect impaired intelligibility. According to this, seven of our subjects were normally

Clin Linguist Phon Downloaded from informahealthcare.com by UB der LMU Muenchen on 04/05/11
For personal use only.

RIGHTSLINK

Table 2. *Expert and non-expert scores for speech and swallowing assessment (Problematic points of articulation: labial, apical and dorsal)*

	Therapy Outcome Measure		Munich Intelligibility Profile				
p#	phonol. disability	dysphagia	errors %	speech impairment	clearness	voice	problematic PoA
1	3	3.0	10	3.75	4	1	dor
2	4	4.0	15	3	2.25	2	ap+dor
3	4.5	5.0	6	1	0	0.24	lab+ap
4	5	3.0	2	1	0.75	1	ap
5	4	4.0	13	2.75	2.25	1	lab+dor
6	5	3.0	3	3.25	0.5	1.25	ap
7	3.5	2.5	15	3	2.75	0.5	lab+ap+dor
8	3	3.0	3	2.75	2.75	2.5	
9	2.5	3.0	22	4.25	4.25	1.25	ap+dor
10	1	0.0	46	5.67	5.67	1	ap+dor
11	3	3.0	12	3.25	2.75	1.5	lab+ap+dor
12	4	2.5	2	2.5	1.75	0.5	ap
13	4	4.0	2	1.67	1	0.33	dor
14	3	3.5	8	2	2	1	ap
15	5	3.0	1	0.33	0.33	0.33	lab+dor
16	5	4.0	5	0.75	0.75	0.25	
17	4	4.0	2	2.25	2.25	1.25	dor
18	4	3.0	7	2.5	2	0.25	dor
19	4	4.0	13	2.33	2.67	0.67	ap+dor
Mean	**3.76**	**3.24**	**10**	**2.53**	**2.14**	**0.94**	
SD	**1.02**	**1.02**	**11**	**1.28**	**1.45**	**0.62**	

Table 3. *Correlation (Spearman Rho) between error rate (amount of incorrectly identified items in MIP), non-expert ratings in MIP and expert ratings (TOM Phonological Disability). **significant if* $p<0.01$*, *significant if* $p<0.05$*. Negative correlation coefficients for the TOM scores are due to the reversed scale in the questionnaire*

Correlation (Spearman Rho)	MIP errors	MIP speech impairment	MIP clearness	MIP voice
MIP speech impairment	0.680**			
MIP clearness	0.693**	0.781**		
MIP voice	0.236	0.591**	0.525*	
TOM Phon. dis.	−0.615**	−0.699**	−0.890**	−0.440

intelligible. However, the means of their subjective scores in MIP did show some impairment ($mean_{SI}=1.96$, $mean_{CA}=1.33$). As no norm values for the subjective scales are given in Ziegler *et al.* (1992), we computed mean values for persons who were available for the preoperative testing. The mean error rate was 4.0% for this group of participants, $mean_{SI}=0.56$, $mean_{CA}=0.59$, $mean_{voice}=0.37$. On the TOM scale, all participants were scored 5.0.

Turning now to speech self-assessment, the EORTC values are summarized in table 4 for the six patients who had completed the questionnaire. The table also shows their MIP and TOM scores. Correlation coefficients show that none of the speech evaluation methods used in this survey reflects the self-assessment of our

Clin Linguist Phon Downloaded from informahealthcare.com by UB der LMU Muenchen on 04/05/11
For personal use only.

RIGHTSLINK

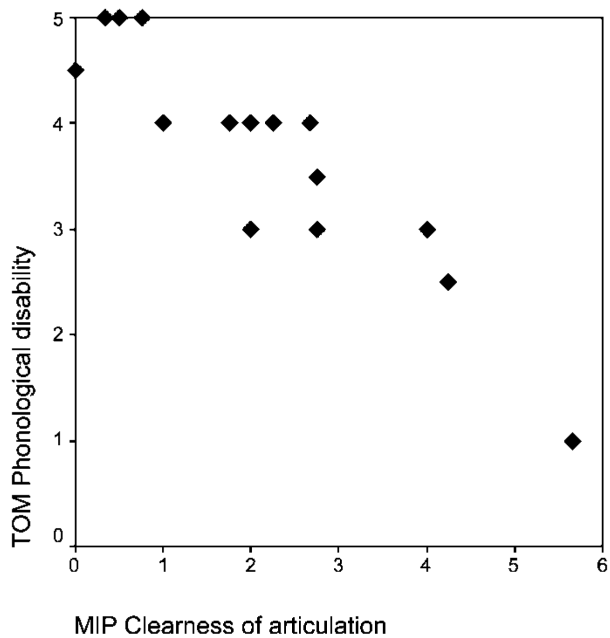

Figure 1. *Correlation between error rate, non-expert ratings for clearness of articulation, and expert ratings.*

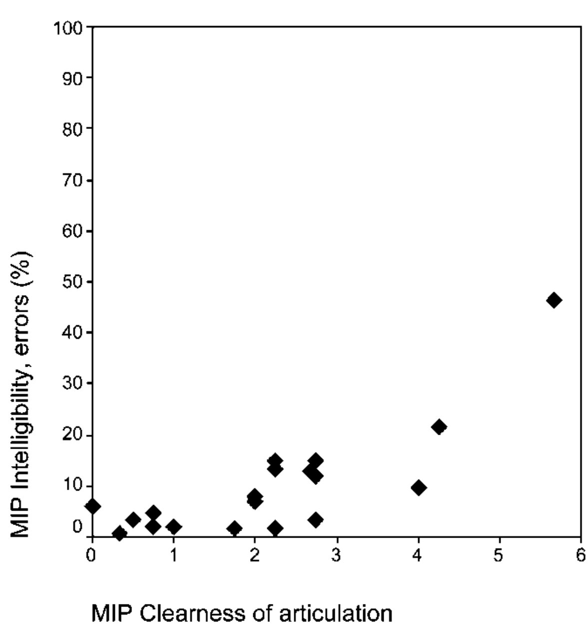

Figure 2. *Correlation between error rate, non-expert ratings for clearness of articulation, and expert ratings.*

Clin Linguist Phon Downloaded from informahealthcare.com by UB der LMU Muenchen on 04/05/11
For personal use only.

RIGHTSLINK

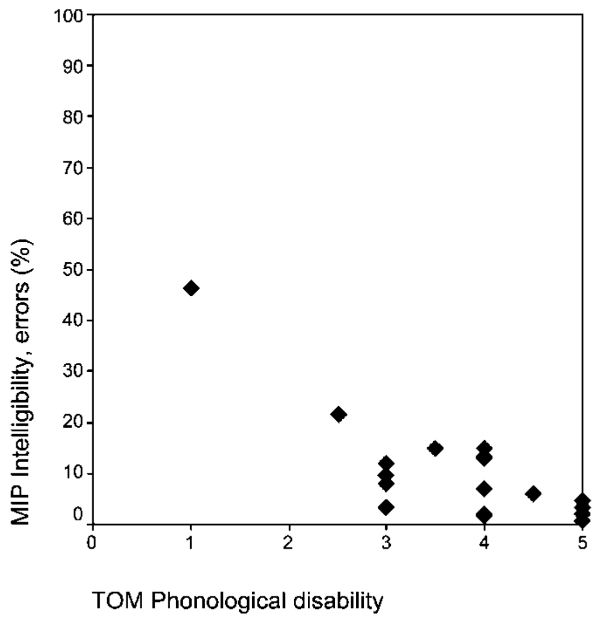

Figure 3. *Correlation between error rate, non-expert ratings for clearness of articulation, and expert ratings.*

participants: Spearman Rho for self-assessment and (a) errors: 0.015, (b) SI: 0.279, (c) CA: 0.368, (d) voice: –0.441, and (e) expert ratings: –0.250.

Swallowing

According to the TOM scales, swallowing was a more severe problem than the articulatory impairment ($mean_{PhonL\ imp} = 3.76$, $mean_{Swall\ imp} = 3.24$). Although the scales cannot be compared directly, it is considerable that while only three persons scored better on the phonological scale, eight persons received lower scores for swallowing. Scores for phonological and swallowing disability correlated only weakly (Spearman Rho = 0.521), i.e., dysphagia did not necessarily go hand in hand with the articulatory disability. On the other hand, the seven participants who were normally intelligible scored no better on the TOM dysphagia scale than the group mean ($mean_7 = 3.21$).

Table 4. *MIP. TOM. and EORTC scores for six patients. Speech self-evaluation ranges from 0.00 ('no impairment') to 1.00 ('maximal impairment')*

p#	Speech self	errors	Speech impairment	clearness	voice	phonol. disability
1	0.33	0.10	3.75	4	1	3
2	0.00	0.15	3	2.25	2	4
4	0.67	0.02	1	0.75	1	5
7	0.22	0.15	3	2.75	0.5	3.5
9	0.78	0.22	4.25	4.25	1.25	2.5
11	0.00	0.12	3.25	2.75	1.5	3

Clin Linguist Phon Downloaded from informahealthcare.com by UB der LMU Muenchen on 04/05/11
For personal use only.

RIGHTSLINK

In the present study, medical therapy was restricted to surgery or irradiation of the oral cavity, therefore swallowing impairment most often emerged from poor oral control or insufficient bolus transport. There is a possible connection between articulatory problems in the dorsal area and inappropriate bolus transport, as both activities require tongue body elevation. Therefore, separate mean values were calculated for patients with and without dorsal mispronunciations in MIP. The TOM dysphagia means differed only slightly: for patients who did not have difficulties with dorsal articulation $mean_{nd} = 3.38$ ($n = 8$), and for patients with dorsal difficulties $mean_d = 3.14$ ($n = 11$). On the whole, patients with larger tumours suffered from a more severe loss of speech and swallowing ability (figure 4).

Discussion

Although we found an overall correlation between speech assessment measures, some details need further explanation. First, the relatively small percentage of incorrectly identified items in the intelligibility test might appear surprising, given that the mean semiquantitative scores for SI and CA ranged between 2 and 3, and even for participants with a normal intelligibility, the scores ranged between 1 and

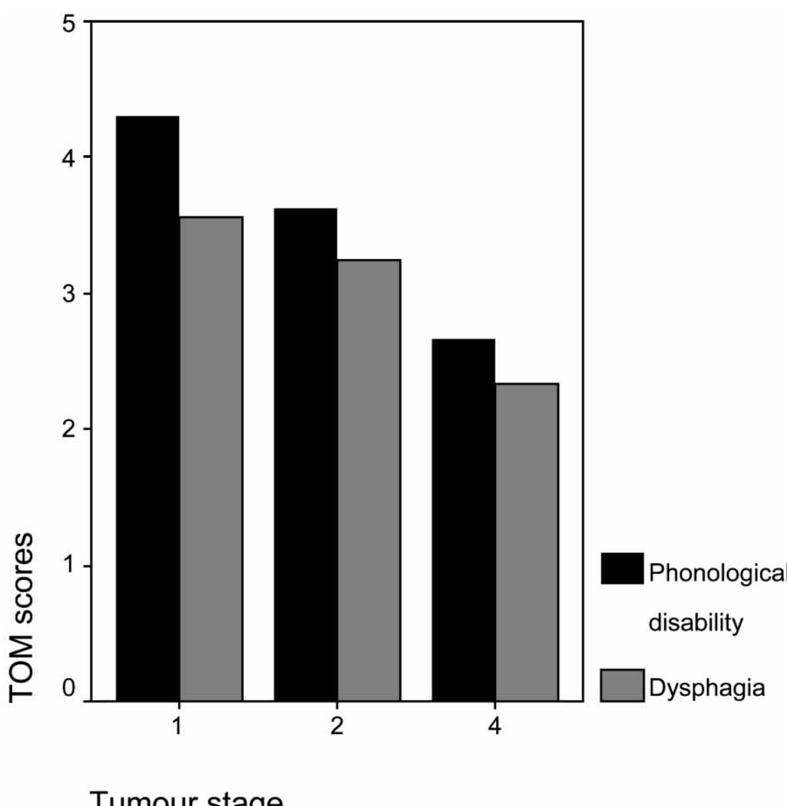

Figure 4. *Speech and swallowing impairment scores in the Therapy Outcome Measure (TOM) test, according to tumour size, '5' meaning no impairment. No T3 tumours occurred in our study.*

Clin Linguist Phon Downloaded from informahealthcare.com by UB der LMU Muenchen on 04/05/11
For personal use only.

Clin Linguist Phon Downloaded from informahealthcare.com by UB der LMU Muenchen on 04/05/11
For personal use only.

2, which is slightly above the unimpaired level. A possible explanation is the scoring technique of the listeners: as they knew they were to listen to speech of persons with oral tumours, they tended to be extremely sensitive to any deviation from what they assumed to be 'normal'. However, the fact that preoperative judgements for SI and CA were much better does not support this interpretation.

It seems more plausible that subjective judgements on these scales were not clearly based on supraglottal features. It is clear from table 3 that SI and CA scores correlated highly with voice quality scores while error rates and expert ratings did not. The correlation between SI/CA and voice reveals a tendency of the listeners to judge these factors somewhat globally rather than in isolation. It seems that speech evaluation by non-experts tends to be unspecific, whereas experts consider various aspects of spoken language separately. However, we cannot exclude the possibility that voice quality was also impaired in our participants. This was in fact shown in a paper by Zimmermann, Sader, Hoole, Bressmann, Mády and Horch (in press). They found that oral tumour surgery often leads to changes in voice quality, especially to a higher fundamental frequency.

Our findings on the divergence of self-evaluation and rating by others lead us to a very general question. As noted by Fairclough (2002), objective and subjective QoL measurements often diverge, and there is no reason to assume the expert ratings to be 'more valid', even if they contain exact measurements or rely on validated methods. On the other hand, the fact that the objective assessment of the patients' speech is often better than their own judgement, points the therapist to the necessity of an objective feedback for each patient on his/her own speech.

It is not obvious why swallowing was a more severe problem for many patients than speaking. Deleyiannis, Weymuller and Coltrera report similar results in their study. They suggest that QoL measurements should be complemented by importance scores, as a moderate loss of an ability that is highly important for a person might influence QoL scores more strongly than a severe loss in a less important field. Swallowing ability is vital for anybody and might rank higher in importance than speaking. On the other hand, it is also possible that postsurgical compensatory strategies for speech are more easily found than for swallowing.

In summary, the comparison of different speech assessment methods did not turn out to lead to basically different results. However, non-experts appeared to form a rather unspecific impression of impaired speech, for example by transferring supraglottal features to voice evaluation whereas experts do not. Finally, our participants considered their own swallowing disability to be more severe than their speech and communication disorders, which might be connected with the primary, more vital function of swallowing.

Our results show that the intelligibility measure provided by MIP cannot be replaced by any of the subjective scales we used. As a next step, we will have to specify those features that have an impact on the subjective scoring but not on intelligibility. Therefore, particular attention will be paid to cases with a remarkable divergence between error rates and subjective scores. This work will include the acoustic analysis of segmental and suprasegmental units of speech.

Acknowledgement

This study was supported by the Deutsche Forschungsgemeinschaft (DFG), Germany, No. HO1972/1.

RIGHTSLINK

References

BJORDAL, K., DE GRAEFF, A., FAYERS, P. M., HAMMERLID, E., VAN POTTELSBERGHE, C., CVRRAN, D., AHLNER-ELMQVIST, M., MAHER, E. F., METZA, F. W., BRÉDART, A., SÖDERHOLM, A. L., ARRARAS, F. F., FEINE, F. S., ABENDSTEIN, H., MORTON, R. P., PIGNON, T., HUGUENIN, P., BOTTOMLY, A. and KAASA, S., 2000, A 12 country field study of the EORTC QLQ-C30 (version 3.0) and the head and neck cancer specific module (The EORTC QLQ-H&N35). *European Journal of Cancer*, **36**, 213–219.

DELEYIANNIS, F. W.-B., WEYMULLER, E. A. Jr. and COLTRERA, M. D., 1997, Quality of life of disease-free survivors of advanced (stage III or IV) oropharyngeal cancer. *Head and Neck*, **19**, 466–473.

ENDERBY, P. and JOHN, A., 1997, Therapy outcome measures (TOM): speech and language therapy (San Diego, CA: Singular), p. 30, p. 42.

FAIRCLOUGH, D. L., 2002, *Design and analysis of quality of life studies in clinical trials* (Boca Raton, CA: Chapman & Hall/CRC Press).

FAYERS, P., AARONSON, N., BJORDAL, K. and SULLIVAN, M., 1997, EORTC QLQ–C30. EORTC Brussels.

GELLRICH, N.-C., SCHRAMM, A., BÖCKMANN, R. and KUGLER, J., 2002, Follow-up in patients with oral cancer. *Journal of Oral Maxillofac Surgery*, **60**, 380–386.

JOHN, A. and ENDERBY, P., 2000, Reliability of speech and language therapists using therapy outcome measures. *International Journal of Language and Communication Disorders*, **35**, 287–302.

MACKENZIE BECK, J., WRENCH, A., JACKSON, M., SOUTAR, D., ROBERTSON, A. G. and LAVER, J., 1998, Surgical mapping and phonetic analysis in intra-oral cancer. In W. Ziegler and K. Deger (Eds.) *Clinical Phonetics and Linguistics* (London: Whurr), pp. 481–492.

PAULOSKI, B. R., LOGEMANN, F. A., RADEMAKER, A. W., MCCONNEL, F. M. S., HEISER, M. A., CARDINALE, S., SHEDD, D., LEWIN, F., BAKER, S. R., GRANER, D., COOK, B., MILIANTI, F., COLLINS, S. and BAKER, T., 1993, Speech and swallowing function after anterior tongue and floor of mouth resection with distal flap reconstruction. *Journal of Speech and Hearing Research*, **36**, 267–276.

PAULOSKI, B. R., RADEMAKER, A. W., LOGEMANN, J. A. and COLANGELO, L. A., 1998, Speech and swallowing in irradiated and nonirradiated postsurgical oral cancer patients. *Otolaryngology Head and Neck Surgery*, **118**, 616–624.

RENTSCHLER, G. J., ARBOR, A. and MANN, M. B., 1980, The effects of glossectomy on intelligibility of speech and oral perceptual discrimination. *Journal of Oral Surgery*, **38**, 348–354.

SCHLIEPHAKE, H., SCHMELZEISEN, R., SCHÖNWEILER, R., SCHNELLER, T. and ALTENBERND, C., 1998, Speech, deglutition and life quality after intraoral tumour resection: a prospective study. *International Journal of Oral and Maxillofacial Surgery*, **27**, 99–105.

ZIEGLER, W., HARTMANN, E. and WIESNER, I., 1992, Dysarthriediagnostik mit dem "Münchner Verständlichkeits-Profil" (MVP) – Konstruktion des Verfahrens und Anwendungen. *Nervenarzt*, **63**, 602–608.

ZIEGLER, W. and HARTMANN, E., 1993, Das Münchner Verständlichkeitsprofil (MVP): Untersuchungen zur Reliabilität und Validität. *Nervenarzt*, **64**, 653–658.

ZIMMERMANN, A., SADER, R., HOOLE, P., BRESSMANN, T., MÁDY, K. and HORCH, H.-H., in press, Oral cavity tumour surgery and voice quality/F_0. *Clinical Linguistics and Phonetics* (accepted for publication).

Clin Linguist Phon Downloaded from informahealthcare.com by UB der LMU Muenchen on 04/05/11
For personal use only.