

Reliability and validity of a new method of nurse-led assessment of trismus

J. Hoole^a, G.W. Jenkins^b, A.N. Kanatas^{a,*}, D.A. Mitchell^a

^a *Oral and Facial Specialties Department, Mid-Yorkshire Hospitals NHS Trust, Pinderfields Hospital, Aberford Road, Wakefield WF1 4DG, United Kingdom*

^b *Oral and Maxillofacial Department, Bradford Teaching Hospitals, Bradford Royal Infirmary, BD9 6RJ, United Kingdom*

Accepted 12 July 2010

Available online 6 August 2010

Abstract

Our aim was to develop a way of measuring trismus and evaluating it in terms of reliability and validity for use by nurses who care for patients with head and neck cancer. We developed a method from existing resources and tested it for interoperator reliability and validity against a “gold standard”. We showed that within the variables outlined, nursing staff could use it after minimal training to identify patients broadly as being “at risk”, “low risk”, or having “normal mouth opening”. Interexaminer reliability was poor. There was a trend towards a group of people having reasonable interexaminer reliability, and this same group showed consistency towards the gold standard. Nurses’ accuracy and consistency in using a simple method of assessing trismus depends on more than the method used, and neither validity nor reliability can be assumed. The accurate assessment of trismus for comparison across studies is complex, and future research must define exactly what method was used.

© 2010 British Association of Oral and Maxillofacial Surgeons. Published by Elsevier Ltd. All rights reserved.

Keywords: Head and neck cancer; Trismus; Radiotherapy; Oral oncology

Introduction

Malignancy of the head and neck is the sixth most prevalent cancer worldwide.¹ In the United Kingdom there are about 4650 reported diagnoses of oral cancer every year, equating to roughly 2% of all cancers.² Though there have been improvements in treatment, there are well-recognised complications of treatment or of the disease itself.³ Trismus is a common but poorly documented complication of the treatment, or the disease, or both,⁴ but although it is recognised, there is no published standard definition. Dijkstra et al. described it as “severely restricted mouth opening” and, after a retrospective study, described it as “mouth opening of less than 35 mm”.⁵ It has also been described as “mouth

opening less than 20 mm”,⁶ and “mouth opening less than 40 mm”.⁷ Other authors have attempted more graded objective approaches based on the measured mouth opening. Mild or “light” trismus has been described as mouth opening of more than 30 mm; mouth opening between 15 and 30 mm as moderate trismus; and mouth opening of less than 15 mm as severe trismus.⁸

The reported prevalence of trismus after head and neck malignancy ranges from 5% to 38%,^{8,9} and given the variability of definitions the width of this range could have been expected. There is also an inherent assumption that “trismus” is an easily measured physical state (interincisal opening). It is clearly not, as variables such as overjet and overbite; and degree, state, and presence of dentition have an influence. There is also a subjective component with wide individual variation.

Trismus may result from invasion of the tumour into the muscles of mastication as the disease progresses, or be a consequence of muscular fibrosis in patients free of cancer as

* Corresponding author.

E-mail addresses: Julie.hoole@midyorks.nhs.uk (J. Hoole), gwenjenkins@hotmail.co.uk (G.W. Jenkins), a.kanatas@doctors.org.uk (A.N. Kanatas), David.mitchell@midyorks.nhs.uk (D.A. Mitchell).

a result of otherwise successful treatment. Radiotherapy is recognised as one of the primary risk factors for the development of trismus, dependent upon the orientation, shielding, and total dose given.¹⁰ The formation of scar tissue postoperatively is also a risk factor.

It is clear that trismus is a serious complication in patients who have, or who have survived, treatment for head and neck cancer and, if studies into its development, prevention, and treatment are to be comparable, an attempt must be made at consistency of assessment. In this study we have attempted to evaluate the acceptability, reliability, and validity of a new way of measuring trismus among nurses who routinely care for patients with head and neck cancer.

Patients, material, and methods

Study design

The first stage included the design and development of a way of assessing and measuring trismus. We then tested its reliability among eight nurses by assessing and measuring measurements of mouth opening in the same subjects. That was followed by the testing of the validity of the method by comparing the nurses' measurements against those of a "gold standard". The final stage included a series of descriptive evaluations of ease of use and overall acceptability.

Development of a way of assessing and measuring trismus

There is currently no "gold standard" measurement that can be used to find out whether or not a patient has trismus. We felt that the diagnosis was not only quantitative (the measured distance of mouth opening) but subjective (the patient's perceived oral function).^{11,12} For this reason we combined objective measurements of mouth opening and a subjective description of oral function derived from the patient. It was divided into three sections, and presented in the form of a questionnaire (Appendix A). This study details

the reliability and validity of assessments of the interincisal measurements made using the Therabite motion scale (Platon Medical Ltd., Eastbourne, UK) on the basis that this component of the assessment should prove the most amenable to statistical analysis. We also made a complementary acceptability assessment using a modified Likert scale.

The method was finalised and guidelines issued on its use. The assessment and measurement aspects, together with the guidelines, were independently tested by four specialist nurses in the head and neck unit. No changes were implemented. Healthy volunteers were used for interincisal measurements. The consent of the local Ethics and Research and Development committee was given.

Eight general nurses volunteered to participate in the study and attended a 1-h training session. The training session encompassed all facets of the method, but emphasised the accurate use of the "Therabite" motion scale.

Intraobserver and interobserver reliability

Repeated assessments were made of individual nurses' measurements and the median value across all the nurses measurements was calculated. All participants of the study were asked to measure interincisal mouth opening on two healthy volunteers (Tables 1 and 2) three times, with an intervening rest period between each measurement. A blank "Therabite" motion scale was used for each new measurement, and appropriately marked at maximum mouth opening.

Interexaminer reliability and validity by comparison of nurses' measurements against a gold standard

As no "gold standard" existed, for the purpose of this study a consultant oral and maxillofacial surgeon who had been trained in the method and been validated by repeated testing until measurement in the same (rested) subjects was consistent on consecutive days was used as the gold standard. Sixteen healthy volunteers had their maximum mouth opening measured (Table 3). All eight nurses and the "gold standard" surgeon measured the same test subjects. The ease

Table 1
Three times blind mouth opening measurements (mm) of volunteer 1.

	Nurse							
	1	2	3	4	5	6	7	8
1	30	46	38	50	41	41	42	45
2	40	45	33	51	41	41	41	43
3	28	46	38	45	41	42	50	44
Mean	32.7	45.7	36.3	48.7	41	41.3	44.3	44
Mode	–	46	38	–	41	41	–	–
Median	30	46	38	50	41	41	42	44
Highest	40	46	38	51	41	42	50	45
Lowest	28	45	33	45	41	41	41	43
Range	12	1	5	6	0	1	9	2
Percentage of value	40	2	13	12	0	2	21	5

Table 2
Three times blind mouth opening measurements (mm) of volunteer 2.

	Nurse							
	1	2	3	4	5	6	7	8
1	38	45	40	43	43	43	40	45
2	30	43	47	42	44	43	40	39
3	44	43	47	44	38	42	41	42
Mean	37.3	43.7	44.7	43	41.7	42.7	40.3	42
Mode	–	43	47	–	–	43	40	–
Median	38	43	47	43	43	43	40	42
Highest	44	45	47	44	44	43	41	45
Lowest	30	43	40	42	38	42	40	39
Range	14	2	7	2	6	1	1	6
Percentage of value	37	5	15	5	14	2	3	14

Table 3
Comparison of gold standard measurements of mouth opening of 16 volunteers with measurements taken by nurses of the same volunteers (mm).

	Volunteer															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Gold standard	42	48	51	47	54	53	46	43	42	52	57	49	49	45	44	54
Nurse 1	47	36	52	53	53	51	53	44	46	52	56	54	49	46	54	56
Nurse 2	41	42	53	47	53	51	51	38	49	50	56	52	45	46	55	54
Nurse 3	44	29	51	44	45	46	46	30	37	41	51	52	39	39	52	31
Nurse 4	39	41	50	44	48	49	44	38	37	46	48	51	45	41	40	41
Nurse 5	40	42	43	50	48	45	34	36	43	42	49	50	43	42	47	34
Nurse 6	43	34	50	47	46	46	47	37	47	49	53	52	50	44	49	50
Nurse 7	41	29	51	48	51	50	42	38	49	51	54	52	48	44	51	54
Nurse 8	41	36	50	45	52	45	41	31	40	49	54	51	45	44	46	53

of use and acceptability were assessed on a Likert scale by the users.

The eight nurses who volunteered for the study were asked to complete a questionnaire (Appendix B). A Likert scale was used to aid statistical analysis.

Statistical analysis of reliability

The assessment of reliability was an analysis of intraobserver and interobserver variability in recording of interincisal opening. Intraobserver variability (also known as test-retest reliability) is tested when each nurse makes three blinded measurements on the same volunteer. The variation in, or consistency of, the measurements give a within-observer error for each nurse. The median and mode are calculated, as is the difference between the highest and lowest of the three measurements for each nurse. This is then expressed numerically as the percentage of the median, with the acceptable clinical degree of error being less than or equal to 5%.

To measure and convey the numerical data for between-observer measures against the gold standard, the coefficient of variation (CV) can be considered, but a more informative measure of reliability (as opposed to variability) would be the percentage agreement between the scores and, more importantly, the gold standard. It is therefore more appropriate to use this test for analysis of this subset of data.

Statistical analysis of validity

Though the reliability and accuracy of individual measurements are important factors, the validity of these measurements is paramount. It is this measurement, together with the other components of the measurement and assessment of the method, that forms the basis of a clinical decision. A numerical comparison is no longer possible, as the data have been converted from a continuous to a categorical scale of measurement. Validity is assessed using a simple 2×2 table, in which true positives, true negatives, false positives, and false negatives are recorded. As the assessment tool has a three-point scale, we have devised a 3×3 table. The statistical analysis undertaken is similar to that used in the paper by Jager-Wittenaar et al.¹³

Results

Statistical analysis of reliability

To measure test-retest reliability each individual nurse made three blind measurements on the same two healthy volunteers.

Healthy volunteer 1

Only half the nurses made consistent measurements. The four nurses who were consistently accurate had a range from the

median of 0–4.5% (within the “less or equal to 5% acceptable error” range). The four nurses who lacked consistency ranged between 12% and 40% from the median.

Healthy volunteer 2

Again half the nurses made consistent measurements. The four nurses who were consistently accurate had a range from the median of 2.3–4.6% compared with those nurses who lacked consistency and who ranged between 13.9% and 36.8%.

Only two nurses were consistently accurate over the two healthy volunteers. It is interesting to note that one nurse showed perfect correlation in healthy volunteer 1, but had a 13.9% range from the median in volunteer 2. The nurse with the highest percentage difference from the median in both healthy volunteers was the same nurse. To analyse the results further, the interobserver variability and reliability can be measured by comparing each of the eight nurses' mean and median values. Even though there are variations in the intraobserver reliability, the median comparison of interobserver reliability, with an acceptable clinical error being less than or equal to 5%, shows that five of the eight nurses showed good agreement in their interobserver measurements for healthy volunteer 1, and six of the eight nurses showed good agreement in interobserver measurements for healthy volunteer 2.

Statistical analysis of validity

The validity of a measurement is crucial as it relates the measurement to a clinical decision, in this case low risk, at risk, and high risk. The results establish that 12 of the 16 volunteers were accurately placed by all eight nurses in the correct category ‘no risk’, and therefore a true negative as defined by the gold standard. The percentage median difference between each nurse's 16 measurements and that of the gold standard show that only three nurses were consistently reliable against the gold standard. However, the percentage mean difference means that half the nurses are consistently accurate and reliable in their measurements against the gold standard. Three of the eight nurses were consistently accurate and reliable against the gold standard with both the mean and the median percentage differences. Four nurses were inconsistent in both the mean and median percentage difference with one nurse being accurate against the gold standard if measuring mean but inconsistent if measuring median value as a percentage difference.

Feasibility of the questionnaire

The results indicate that 14 of the 16 nurses “strongly agreed” in answer to all seven questions about the feasibility of measuring and assessment. The nurses are not identifiable on the questionnaire and no comparisons can be made with their measurements. They strongly agreed that it was an appropriate method of assessment and acceptable to them.

Discussion

As clinicians it is difficult for us to categorise those patients who have been treated for an oral malignancy and who require further specialist services for rehabilitation of trismus. In addition, therapists do not know what degree of mouth opening should be aimed for to provide adequate oral function. With no standard definition of trismus and a lack of specific groupings it is difficult to calculate the degree of mouth opening required for normal oral function. Trismus may affect the patient's quality of life enough to reduce the effects of rehabilitation and overall wellbeing.¹⁴

Jager-Wittenaar et al.¹³ stated that “maximal mouth opening can be assessed reliably, regardless of the observer in head and neck cancer patients”. In their study maximal interincisal mouth opening was measured only by two non-specialised people, though training was provided. It should be noted that these measurements were not compared with a gold standard.

We accept that there are limitations to our study (the limited numbers of volunteers and the use of healthy volunteers allow for possible skewing of measurements), but the use of healthy dentate volunteers could be expected to make obtaining measurements much simpler than if they were made in partially dentate or edentulous patients before or after treatment for head and neck cancer. However, while the use of the assessment allowed patients to be classified into broad categories of risk, the reliability and validity of half the test measurements were inadequate. In addition, despite the objective shortcomings, all the users felt themselves to be competent and comfortable with the method. Previous publications have suggested^{4,13} that there needs to be an improvement of more than 5 mm in maximum mouth opening to be of any clinical relevance in assessing whether there has been improvement in a patient's maximum mouth opening. It is for this reason that we have classified the various degrees of risk of trismus in 5 mm bands. Our results have shown that while the technique can be used by nurses to “diagnose” trismus, in that patients can be placed into broad groups as illustrated by the 3 × 3 tables, it cannot be reliably used by minimally trained general nurses to detect small variations (mm of mouth opening) accurately and reliably. The ability to predict the onset of trismus reliably would enable further research to be made into the most effective treatments. It would also enable the effects of different treatments to be delineated accurately, but studies that purport to do this must describe their intraexaminer and interexaminer reliability in the use of their methods of assessment for the work to be credible.

Acknowledgement

The Therabite measuring device was supplied by Platon UK.

Appendix A. Mouth opening assessment method for adults treated for oral cancer

1. Demographic and treatment history

<p>PATIENT STICKER Consultant..... Date.....</p>
--

Male Female

Diagnosis

Date of completed:

Surgery N/A

Radiotherapy N/A

2. Oral examination:

(Please tick appropriate boxes)

Full set own teeth

Own top set of teeth complete partial

Own bottom set teeth complete partial

Edentulous (toothless) wears dentures: Yes No

If they wear dentures are they a:

Full set Top set Bottom set

Other please indicate

3. Please tick the column against each question that best describes your answer to the following questions:

*

		Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree
		5	4	3	2	1
3a	My mouth opening restricts the type of food I eat					
3b	My mouth opening makes chewing food difficult					
3c	When I chew food it is painful.					
3d	When eating my mouth opening restricts me using the following: <ul style="list-style-type: none"> • Teaspoon • Fork • Dessert spoon • Cup or mug 					
3e	My speech is impaired due to my restricted mouth opening					
3f	I am not able to open my mouth wide enough to clean and brush my teeth					
3g	My restricted mouth opening stops me from being able to examine the inside of my own mouth					

4. Mouth opening measurement.

4a. Using Platon “Therabite motion scale” method

Mouth opening measures mm

<i>< 30 mm</i> High risk	<i>30-35 mm</i> At risk	<i>35 mm ></i> Normal (low risk)

Patients identified as high risk or at risk please advise on mouth opening exercises and the technique of using wooden spatulas.

4b: Has advice been given on spatulas and how to use them:

Yes No (referred to speech therapy) N/A

- If appropriate the number of spatulas a patient can tolerate when assessed today

4c: Has advice about how to do mouth opening exercises and a practical demonstration been given:

Yes No (referred to speech therapy) N/A

5 From this assessment, has the patient been referred to:

Dietician Yes No N/A
 Speech therapist Yes No N/A

Signed.....

Appendix B. Questionnaire: ‘The ease and use of the trismus measuring method’ (Please tick the column against each question that best describes your answer)

	Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree
	5	4	3	2	1
1					
2					
3					
4					
5					
6					
7					

References

- Watkinson JC, Gaze MN, Wilson JA, editors. *Stell and Maran's head and neck surgery*. 4th ed. Oxford: Butterworth Heinemann; 2000.
- Cancer Research UK 2009 at: www.cancerresearchuk.org.
- Jemal A, Siegel R, Ward E, Murray T, Xu J, Smigal C, et al. Cancer statistics 2006. *Cancer J Clin* 2006;**56**:106–30.
- Dijkstra PU, Kalk WWI, Roodenburg JLN. Trismus in head and neck oncology: a systematic review. *Oral Oncol* 2004;**40**:879–89.
- Dijkstra PU, Huisman PM, Roodenburg JLN. Criteria for trismus in head and neck oncology. *Int J Oral Maxillofac Surg* 2006;**35**:337–42.
- Jen YM, Lin YS, Su WF, Hsu WL, Hwang JM, Chao HL, et al. Dose escalation using twice-daily radiotherapy for nasopharyngeal carcinoma: does heavier dosing result in a happier ending? *Int J Radiat Oncol Biol Phys* 2002;**54**:14–22.
- Nguyen TD, Panis X, Froissart D, Legros M, Coninx P, Loirette M. Analysis of late complications after rapid hyperfractionated radiotherapy in advanced head and neck cancers. *Int J Radiat Oncol Biol Phys* 1988;**14**:23–5.
- Thomas F, Ozanne F, Mabelle G, Wibault P, Eschwege F. Radiotherapy alone for oropharyngeal carcinomas: the role of fraction size (2 Gy vs 2.5 Gy) on local control and early and late complications. *Int J Radiat Oncol Biol Phys* 1998;**15**:1097–102.
- Stelman R, Sokol J. Quantification of trismus following irradiation of the temporomandibular joint. *Mo Dent J* 1986;**66**:21–3.
- Teguh DN, Levendag PC, Voet P, van der Est H, Noever I, de Kruijff W, et al. Trismus in patients with oropharyngeal cancer: relationship with dose in structures of mastication apparatus. *Head Neck* 2008;**30**:622–30.
- Melchers LJ, Van Weert E, Beurskens CH, Reintsema H, Slagter AP, Roodenburg JL, et al. Exercise adherence in patients with trismus due to head and neck oncology: a qualitative study into the use of the Therabite 1. *Int J Oral Maxillofac Surg* 2009;**38**:947–54.
- Speksnijder CM, van der Glas HW, van der Bilt A, van Es RJ, van der Rijt E, Koole R. Oral function after oncological intervention in the oral cavity: a retrospective study. *J Oral Maxillofac Surg* 2010;**68**:1231–7.
- Jager-Wittenaar H, Dijkstra PU, Vissink A, Van Oort RP, Roodenburg JL. Variation in repeated mouth-opening measurements in head and neck cancer patients with and without trismus. *Int J Oral Maxillofac Surg* 2009;**38**:26–30.
- Scott B, Butterworth C, Lowe D, Rogers SN. Factors associated with restricted mouth opening and its relationship to health-related quality of life in patients attending a Maxillofacial Oncology clinic. *Oral Oncol* 2008;**44**:430–8.