Available online at www.sciencedirect.com
ScienceDirect
British Journal of Oral and Maxillofacial Surgery 57 (2019) 866–872
BRITISH Journal of Oral and Maxillofacial Surgery
www.bjoms.com

Case-mix adjustment in audit of length of hospital stay in patients operated on for cancer of the head and neck

D. Tighe[a,*], I. Sassoon[b], A. Hills[a], R. Quadros[a]

[a] *William Harvey Hospital, EKHUFT, Ashford, KENT TN24 0LZ*
[b] *Dept. of Informatics, King's College London, London*

Available online 25 July 2019

Abstract

Patients treated surgically for squamous cell carcinoma (SCC) of the head and neck form a heterogeneous group, and it is difficult to take this variation into account when measuring the quality of care. We have tested the feasibility of mathematical models that allow for the adjustment for case mix when auditing the length of hospital stay as a proxy indicator of the quality of care. We completed a case-note audit of 733 surgical episodes of care for SCC of the head and neck in five cancer networks, and used logistic regression and decision tree analysis to adjust for case mix using pertinent preoperative variables. Risk adjustment models of length of stay included age, alcohol, T classification, performance status, tracheostomy, high-risk status, and complexity of operation. The risk-adjusted length of stay differed significantly between the cancer networks studied ($p < 0.001$). The models performed acceptably for the purpose of audit when this was under 15 days. Length of stay is a measurable outcome that can be used as a benchmark of surgical care. Audits of this after operations for cancer of the head and neck, if reported in national clinical audits, should take case mix into account.

Keywords: Outcomes; Length of stay HNSCC; Audit

Introduction

As survival after cancer improves, the focus on cure as the all-important marker of effective treatment seems to be inadequate, particularly when complex multimodal treatments can result in so many complications that affect a patient's quality of life. In the UK surgical community and in our specialty in particular, the focus is now turning towards outcomes. Datasets of the surgical outcomes of consultants routinely report unplanned returns to theatre and deaths within 30 days, usually with specialty-specific complications. All specialties, except that of head and neck oncology, adjust for case mix when reporting early complications. The complexity of the intervention and adjustments for case mix currently do not feature in reports of length of stay.

Patients who have operations with curative intent for SCC of the head and neck form a particularly heterogeneous group and many are frail with multiple coexisting conditions. The risk factors for developing the disease are also implicated in cardiac, respiratory, and liver disorders. Notwithstanding the anatomical disruption of the operation, the effort to return patients to form and function with immediate reconstruction, can leave them with considerable adjustments to make in swallowing and speaking. These can make it necessary to consider a transfer to a residential or nursing home, which can often take time after a patient has been deemed fit enough to be discharged from hospital.

Our primary outcome was to model the length of hospital stay in this group of patients to enable effective audit of the quality of care. To test length of stay as a proxy marker for the quality of care, we did not include postoperative com-

* Corresponding author.
E-mail addresses: David.tighe@nhs.net (D. Tighe), sassoon@kcl.ac.uk (I. Sassoon), Ajh502@gmail.com (A. Hills), r.quadros@nhs.net (R. Quadros).

https://doi.org/10.1016/j.bjoms.2019.07.007
0266-4356/

plications in the analysis, though we expect that this would add to the accuracy of the model. A secondary outcome was to define "outlier status" (patients whose stay in hospital is extremely long). Finally, we wanted to find out if markers of socioeconomic deprivation contributed to a longer stay.

Methods

To build models for the adjustment of case mix, we analysed a combined dataset of four case-note audits from four cancer networks that treated patients for SCC of the head and neck (n = 638 care episodes) between 2009 and 2015. A total of 112 care episodes from a fifth cancer network were tested as an (external) validation set. The data from each centre comprised a combination of retrospective (12 months) and prospective (6-12 months) data. All patients had histopathologically-confirmed SCC of the head and neck, and operations were done under general anaesthesia using conventional surgical techniques (excluding laser cases) by maxillofacial or ear, nose, and throat (ENT) surgeons, or both. All the datasets were registered with the clinical audit departments at the respective hospital Trusts, and data were collected by the lead author with permission from the treating consultants. Ethics approval was sought for the validation phase of the audit, as application of the mathematical models could be generalised. Data on socioeconomic status was sought only in the validation group, and were not included in the model's development.

For consistency throughout the audit period, the BUPA severity of surgery index (minor = less than 1 hour; intermediate = less than 6 hours; and major or major complex = 6 hours or requirement for free tissue transfer) was used to grade the complexity of the operation. Patients' demographics; comorbidity using the Adult Comorbidity Evaluation-27 index (ACE-27); World Health Organization (WHO) performance status; American Joint Commission on Cancer (AJCC version 7) tumour size and nodal classification; and operative and anaesthetic treatment, were included. Office of Population Censuses and Surveys Classification of Surgical Operations and Procedures (OPC4) codes were used for a derived variable, the "high-risk" field, which included any operation requiring mucosal closure in association with a neck dissection that could lead to an escape of saliva (such as mandibulectomy, glossectomy, excisions of the floor of the mouth, and pharyngectomy with or without laryngectomy). Any case for which pertinent data points were missing (as selected by each predictive model) was excluded.

Demographic and socioeconomic data were captured from the East Kent Patient Centre. Independent variables included cohabitation with next of kin, and distance from home to hospital (using postcodes). Postcodes were also used to derive the level of deprivation.[1] Student's *t* tests and correlation coefficients were used to test continuous and discreet data, and linear regression and decision tree analyses to build the models. These were based on a training set (60% random selection of cases) and tested on the remainder (40%, the test set). They were then tested again on the validation set (the dataset from the fifth unit).

Linear regression is a traditional means of modelling a continuous dependent variable. Decision tree analysis is a more modern computer-intensive method that is used in the context of live business decisions. It is a hierarchical multivariate technique with a graphical structure that shows the importance of, and the inter-relations between, pertinent variables. The outputs and structure can be updated to reflect new knowledge.[2] The main challenge when building the tree is to decide which attribute to split to have the "best" data split at each step, and the concept of "information gain" informs this decision. Information gain is the difference between the amount of uncertainty before and after a decision is made. The aim, which is to achieve a perfect classification with a minimal number of decisions, is not always possible because of noise or inconsistencies in the data. The variables listed in Table 1 were all potential inputs.

The length of hospital stay, defined as the date of operation to the date of discharge or death, was the output.

Results

A total of 733 patient-care episodes were investigated. Of them, 61 patients with incomplete timeline data were excluded (Site 1: 29/160; Site 2: 23/203; Site 3: 1/175; Site 4: 0/83; and Site 5: 8/112). The mean age of the patients was similar in all the centres but there was a difference in sex distribution (73.5% men at Site 4, and 58% at Site 5). There were obvious differences in case mix – for example, between the incidences for heavy drinking (24% at Site 1, 11% at Site 4), free tissue transfer (24% at Site 1 compared with 60% at Site 3), proportion of patients treated with a tracheostomy (18% at Site 5 and 57% at Site 4), and proportion of patients with a performance status of less than 2 (2.5% at Site 4 compared with 8% at Site 3).

The initial modelling of the length of stay was done with a linear model and a regression tree. A stay of more than 50 days was rare, and accounted for less than 5% of the data. This was chosen to define outlier status, a SD of approximately 1 from the mean, and 24 patients who stayed for more than 50 days were removed (Site 1: 5/131; Site 2: 6/180; Site 3: 2/174; Site 4: 4/83; and Site 5: 4/104) (Table 2).

Linear regression modelling of the independent factors showed age, intake of alcohol, T classification, performance status, tracheostomy, high-risk operation, and complexity of surgery, as independent predictors for an increased duration of stay. However, the model suggested a poor fit, as the residuals were related to duration of stay (heteroskedasticity), and the mean standard error (mean standard error 55.9, adjusted R^2 0.42) was 55.9 (Fig. 1).

Attempts to improve this by transforming the duration of stay with Poisson function were unhelpful (mse 158, multiple R^2 not available).

Table 1
Variables by hospital.

	Hospital					Total
	Site 1	Site 2	Site 3	Site 4	Site 5	
Mean (range) age (years)	66 (63.6–68.1)	67 (64.9–68.6)	66 (64.4–68.1)	62(59.7–65.2)	69 (66.6– 71.4)	
Sex:						
Male	109	147	122	61	65	
Female	51	56	53	22	47	
Total	160	203	175	83	112	733
Alcohol:						
1	69	99	56	28	28	
2	31	54	48	20	46	
3	7	14	38	23	17	
4	31	32	20	5	8	
5	5	10	13	7	6	
Total	143	209	175	83	105	715
Smoking:						
Current	56	83	66	16	46	
Ex-current or non-smoker	88	117	109	67	59	
Total	144	200	175	83	105	707
ACE 27:						
0	62	7	39	35	29	
1	56	123	97	35	51	
2	35	67	32	12	20	
3	1	5	7	1	7	
Total	154	202	175	83	107	721
Performance status:						
0	25	14	102	28	47	
1	90	119	36	47	35	
2	29	54	23	4	18	
3	8	13	14	2	5	
Total	152	200	175	81	105	713
Flap:						
0	118	156	69	51	79	
1	39	47	106	32	31	
Total	157	203	175	83	110	728
Tracheostomy:						
0	124	145	128	48	87	
1	32	56	47	35	20	
Total	156	201	175	83	107	722
Scale of operation:						
1	41	35	27	3	36	
2	65	96	32	28	31	
3	51	72	116	52	45	
Total	157	203	175	83	112	730
High risk:						
0	96	132	101	29	64	
1	61	71	74	54	48	
Total	157	203	175	83	112	730
T classification:						
0	26	50	30	14	13	
1	57	55	35	16	36	
2	32	33	35	19	29	
3	9	12	10	7	2	
4	30	47	63	27	29	
Total	154	197	173	83	109	716
N classification:						
0	88	108	98	42	71	
1	19	24	20	16	7	
2a	14	14	6	0	1	
2b	27	30	37	18	26	
2c	5	9	6	4	1	
3	1	7	5	0	0	
Total	154	192	172	80	106	704

Table 1 (*Continued*)

	Hospital					Total
	Site 1	Site 2	Site 3	Site 4	Site 5	
Previous radiotherapy:						
0	147	159	148	50	99	
1	13	44	27	33	13	
Total	160	203	175	83	112	733
Previous operation:						
0	134	166	138	52	79	
1	26	37	37	31	33	
Total	160	203	175	83	112	733

Table 2
Frequency of extended hospital stay (more than 50 days) by hospital.

Site	No. of patients	Maximum No. of patients	No. of patients with a stay of over 50 days	Mean (SD)	95% CI	Median (95%CI)
1	131	161	5	11.5 (22.6)	7.6 to 15.4	3.0 (3.0 to 4.0)
2	180	144	6	13.5 (20.3)	10.5 to 16.5	5.5 (3.0 to 4.0)
3	174	111	2	11.0 (11.4)	9.3 to 12.7	9.0 (8.0 to 10.0)
4	83	79	4	16.8 (17.3)	13.0 to 20.6	11.0 (8.0 to 14.9)
5	104	84	4	11.0 (15.4)	8.0 to 14.0	6.0 (5.0 to 8.0)

Review of Fig. 1 showed that discrimination of the model degraded after a stay of 10-15 days so a new approach was tested and we resisted the inclusion of complication data, as they would obscure our testing of the duration of stay as a proxy indicator for the quality of surgical care. We set a cut off of hospital stay at less than 15 days, and applied decision-tree methods to model for a short compared with a long stay. The data were again split into a train set (60%) and test set (40%), and the following attributes included: age, T size, performance status, tracheostomy indicator, high-risk indicator, scale of operation, and alcohol. A total of 607 had complete datasets for these variables; 350 had stays of less than 15 days, and 257 had stays of more than 15 days. The decision tree correctly assigned a short compared with a long stay in 79% (sensitivity = 0.8, specificity = 0.78, positive predictive value = 0.73, and negative predictive value = 0.84) (Fig. 2). The model performed well on the (external) validation set (sensitivity 0.8, and specificity 0.74).

Using this model we graphically illustrated the proportion of the hospital's case mix that would stay less than 15 days when the model predicted they would stay longer, the expected length of stay, and durations that were longer than expected (Fig. 3).

The linear regression method was used on the subset of patients who were predicted to be in the short-stay group (as defined by the decision tree), split into a train set (60%) and test set (40%), and the model again identified age, alcohol, T classification, performance status, tracheostomy, high-risk status, and complexity of operation. We then re-tested the linear regression for patients with stays of less than 15 days. The performance of the model improved greatly (mse 4.99, adjusted R^2 0.64) on the test set and was deemed reliable enough for the risk adjustment of data on the external validation dataset. Fig. 4 shows the risk-adjusted length of stay for each hospital. Lines of best fit suggest that this was shortest at Site 1 (test statistic t 4.18, p = 0.0001) and longest at Site 3 (test statistic t -5.0, p = 0.0001) (adjusted for case mix).

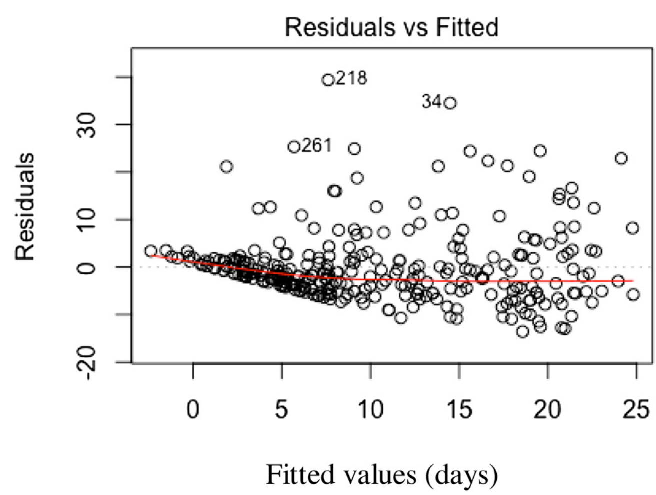

Fig. 1. Residual plot.

The exploration of socioeconomic and demographic factors focused on 104 patients in the external validation set. A total of 62 (60%) had next of kin staying at their address. There was no significant difference in mean length of stay when the group that stayed less than 15 days was tested (test statistic t 1.04, p = 3), but it approached significance when a stay of up to 50 days was investigated (without next of kin: mean 9.2 days, with next of kin 6.27 days, test statistic t -1.76, p = 0.08). In the group that stayed less than 50 days, the distance from home to the hospital did not correlate with an increased duration of stay (r = 0.004, p = 0.9). Finally, Index of Multiple Deprivation deciles were not associated with a longer stay (r = -0.05, p = 0.63).

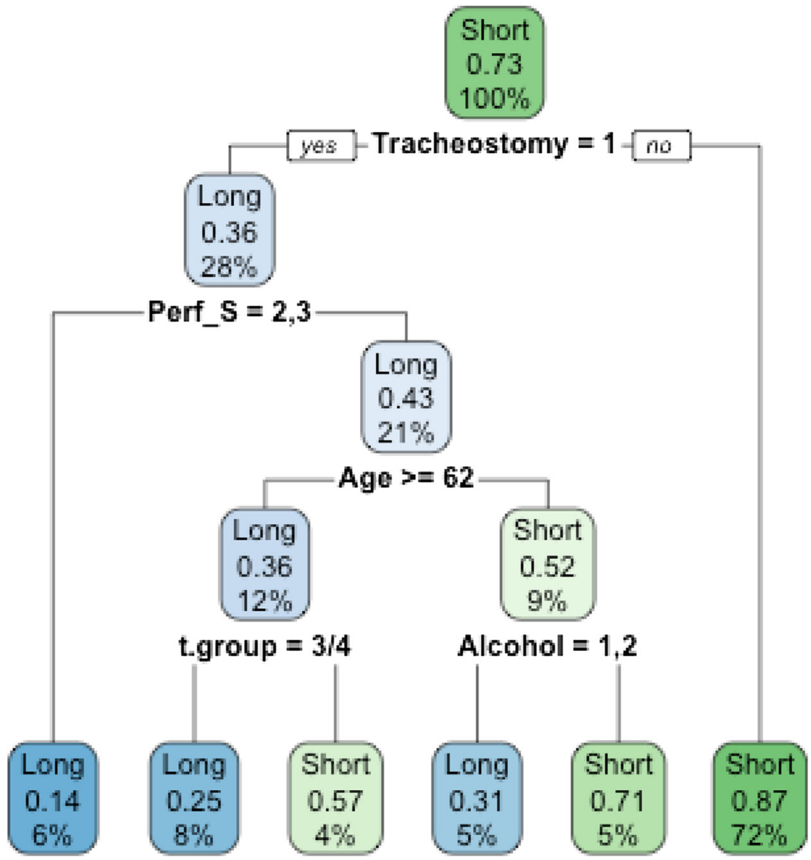

Fig. 2. Decision tree to identify long compared with short hospital stays.

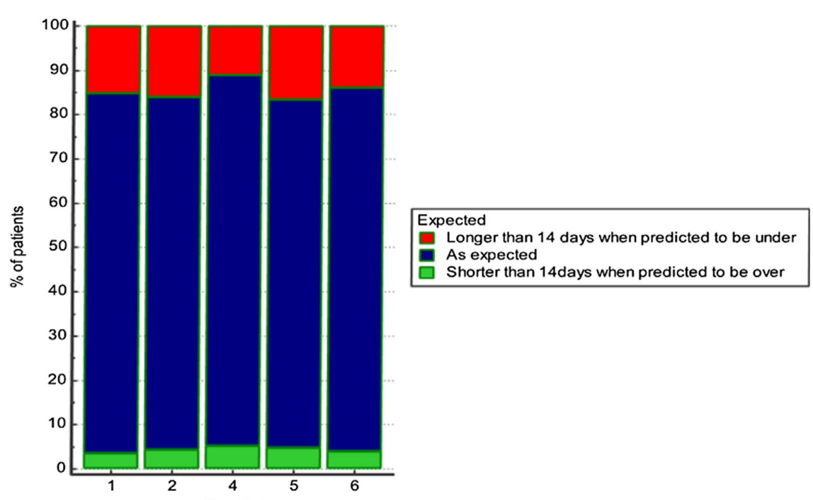

Fig. 3. Histogram to compare frequency of observed with predicted length of stay by type.

Discussion

Modelling for the complexity of cases allows for a more meaningful interpretation of data on length of stay (and all outcome data). We found significant differences in case mix between units, which underline the importance of risk adjustment. Linear regression and decision tree models that have established important predictors of increased length of stay are age, alcohol intake, performance status, and T stage on presentation, together with free tissue transfer and tracheostomy.

Raw data suggested considerable differences between the units with the shortest (Site 1) and the longest durations of stay (Site 3, Site 4). These persisted when case-mix adjust-

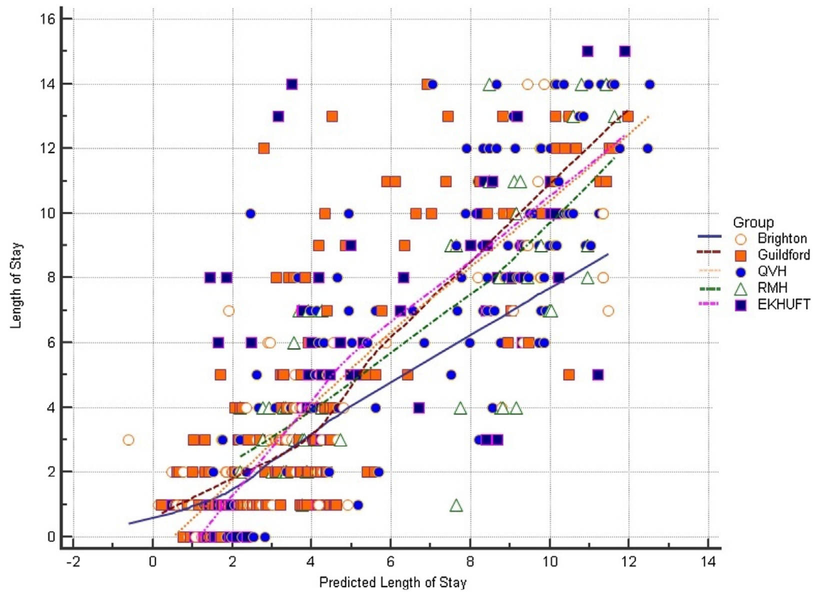

Fig. 4. Scatter diagram to compare length of stay (observed compared with predicted) by treatment centre (QVH = Queen Victoria Hospital; RMH = Royal Marsden Hospital; EKHUFT = East Kent Hospitals University Foundation Trust).

ment was taken into account in those with stays of less than15 days.

The modelling of extreme lengths of hospital stay (or outliers) was not the focus of this paper, though for purposes of allocating resources and hospital finances, it will be useful to explore further.[3,4] This, however, is likely to be difficult if preoperative data are used alone without the incorporation of data on postoperative events, which we deliberately avoided doing in this paper. Further efforts to model for stays of over 15 days, and certainly for those over 50 days, could include data on postoperative complications, because they affect 50% to 70% of patients. Complication rates could be published separately, perhaps to emphasise severity.[2,5]

These "length of stay" models for this group of patients are new in the UK. The American College of Surgeons' National Surgical Quality Improvement Program (NSQIP) risk calculator is an alternative model that predicts length of stay based on preoperative data. A recent audit to validate it using retrospective multicentre data in an Australian group (n = 127) of patients who had glossectomy found that it significantly underestimated length of stay.[6] However, they also cited potential confounding factors that were not taken into account by the risk calculator: age, sex, rural or urban location, TNM classification, previous chemoradiotherapy, neck dissection, free flap reconstruction, and grade of surgical risk. Another study, which used the NSQIP risk calculator for a group of 157 patients who had ablative surgery on the head and neck with microvasculature reconstruction with fibular flaps, reported only slight concordance between the observed and actual lengths of stay.[7] Despite some studies highlighting the role of postoperative complications in longer stays, we think that these data are not helpful when the length of stay is used as a marker of quality of care.[8,9]

To effectively model postoperative outcomes, we assessed the socioeconomic and demographic factors that may be implicated. As we explored this possibility on the last group only, the power of the analysis is limited, but the data suggest that patients who live with a member of the family may need to spend less time in hospital. There was no correlation with a standard index of socioeconomic deprivation (IMD decile) or distance from the hospital. Further work will focus on how often a change in social circumstances is needed postoperatively, as this could function as a proxy marker of the quality of surgical care. This question will be included in the collection of further prospective data.

A limitation of the study was the recording of the discharge date as a single endpoint. We did not record the "medically fit for discharge" date to account for whether discharge was delayed for social reasons. Audits that show stays that are longer than expected for reasons other than surgical performance will potentially support the development of better discharge pathways.

This study has shown that it is possible to base the assessment of a unit's performance on length of stay because it adjusts for case mix and complexity of operation. The decision tree algorithm accurately predicts for a stay of more than 15 days. The logistic regression algorithm allows for the comparison between units of patients who stay in hospital for less than 15 days. Further work needs to be done before these socioeconomic data can be added to the models. It is conceivable that centres that treat patients with SCC of the head and neck can use such models to show safe standards of care in routine audits of outcome.

Conflict of interest

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. The lead author is employed at Site 6, East Kent University Hospitals NHS Foundation Trust.

Ethics statement/confirmation of patients' permission

This audit was registered at each audit department of the respective hospitals. As the culmination of this study could potentially be generalised, ethics approval was obtained from the EKHUFT, Grey Area Project Group, Research and Innovation Department.

Acknowledgements

We thank the consultants whose patients' outcomes form the basis of this report.

References

1. English indices of deprivation 2015. Available from URL: http://imd-by-postcode.opendatacommunities.org (last accessed 14 June 2019).
2. McMahon J, Handley TP, Bobinskas A, et al. Postoperative complications after head and neck operations that require free tissue transfer - prevalent, morbid, and costly. *Br J Oral Maxillofac Surg* 2017;**55**:809–14.
3. Belderrar A, Hazzab A. Hierarchical genetic algorithm and fuzzy radial basis function networks for factors influencing hospital length of stay outliers. *Healthc Inform Res* 2017;**23**:226–32.
4. Freitas A, Silva-Costa T, Lopes F, et al. Factors influencing hospital high length of stay outliers. *BMC Health Serv Res* 2012;**12**:265.
5. Tighe DF, Thomas AJ, Sassoon I, et al. Developing a risk stratification tool for audit of outcome after surgery for head and neck squamous cell carcinoma. *Head Neck* 2017;**39**:1357–63.
6. Kao SS, Frauenfelder C, Wong D, et al. Validity of the American College of Surgeons' National Surgical Quality Improvement Program risk calculator in South Australian glossectomy patients. *J Laryngol Otol* 2018;**132**:173–9.
7. Arce K, Moore EJ, Lohse CM, et al. The American College of Surgeons National Surgical Quality Improvement Program Surgical Risk Calculator does not accurately predict risk of 30-day complications among patients undergoing microvascular head and neck reconstruction. *J Oral Maxillofac Surg* 2016;**74**:1850–8.
8. Penel N, Mallet Y, Roussel-Delvallez M, et al. Factors determining length of the postoperative hospital stay after major head and neck cancer surgery. *Oral Oncol* 2008;**44**:555–62.
9. BuSaba NY, Schaumberg DA. Predictors of prolonged length of stay after major elective head and neck surgery. *Laryngoscope* 2007;**117**:1756–63.