Frozen section analysis of margins for head and neck tumor resections: reduction of sampling errors with a third histologic level

Stephen M Olson, Mohammad Hussaini and James S Lewis Jr

Department of Pathology and Immunology, Washington University in St Louis, St Louis, MO, USA

Frozen section analysis is an essential tool for assessing margins intra-operatively to assure complete resection. Many institutions evaluate surgical defect edge tissue provided by the surgeon after the main lesion has been removed. With the increasing use of transoral laser microsurgery, this method is becoming even more prevalent. We sought to evaluate error rates at our large academic institution and to see if sampling errors could be reduced by the simple method change of taking an additional third section on these specimens. All head and neck tumor resection cases from January 2005 through August 2008 with margins evaluated by frozen section were identified by database search. These cases were analyzed by cutting two levels during frozen section and a third permanent section later. All resection cases from August 2008 through July 2009 were identified as well. These were analyzed by cutting three levels during frozen section (the third a 'much deeper' level) and a fourth permanent section later. Error rates for both of these periods were determined. Errors were separated into sampling and interpretation types. There were 4976 total frozen section specimens from 848 patients. The overall error rate was 2.4% for all frozen sections where just two levels were evaluated and was 2.5% when three levels were evaluated ($P = 0.67$). The sampling error rate was 1.6% for two-level sectioning and 1.2% for three-level sectioning ($P = 0.42$). However, when considering only the frozen section cases where tumor was ultimately identified (either at the time of frozen section or on permanent sections) the sampling error rate for two-level sectioning was 15.3 *versus* 7.4% for three-level sectioning. This difference was statistically significant ($P = 0.006$). Cutting a single additional 'deeper' level at the time of frozen section identifies more tumor-bearing specimens and may reduce the number of sampling errors.
Modern Pathology (2011) **24,** 665–670; doi:10.1038/modpathol.2010.233; published online 7 January 2011

Negative margin status is an important predictor for local control and overall outcome in head and neck cancers.[1–3] Frozen section analysis is a useful tool for assessing margin status intra-operatively, allowing for additional resection during the initial surgery in case of positive margins.[4,5] There is no standard method for frozen section sampling, but the most common method is for the surgeon to, after removal of the gross tumor, take small samples of tissue from the defect cavity and send them to pathology for evaluation.[6,7] Other methods involve the surgeon

sampling the margins directly from the main resection specimen or the pathologist receiving and sampling directly from it. These resection specimen margins may either be shave or radial sections taken from where the tumor is grossly closest to the margin.[6] Shave margins are those in which small pieces of tissue are sampled from the periphery of a resection. This is done when the margins are not thought to harbor tumor and certainly, at least, are not grossly noted to harbor tumor. They are a 'positive or negative' exercise in that the presence of any tumor, regardless of its amount or location in the tissue, indicates a positive margin. Radial margins, on the other hand, are those taken where one is purposely sampling the tumor and the leading edge of a resection perpendicularly in the same histologic section so that the margin can be seen, and the distance from the tumor to the margin observed. The defect sampling method

Correspondence: Dr JS Lewis Jr, MD, Department of Pathology and Immunology, 660 South Euclid Avenue, Box 8118, St Louis, MO 63110, USA. E-mail: jlewis@path.wustl.edu

(which is essentially a form of shave margin) is becoming increasingly common, particularly as a result of transoral laser microsurgery[8,9] in which the tumor is removed in multiple pieces, leaving defect sampling as the major method to clearly assess margins.

There is no consensus on how to handle such specimens once received in the frozen section area. Cutting numerous sections of the tissue likely increases the sensitivity for finding small foci of tumor, but there must be a balance between marginally increased sensitivity and practical issues such as time involved and cost. To our knowledge, most centers simply cut two full hematoxylin and eosin sections per margin specimen with a possible third level cut at the pathologist's discretion.[10,11]

Error rates for frozen section margins are generally low, with reported frozen—permanent correlation rates as low as 96% and as high as 99%.[10–19] Correlation rates have been similarly high in studies that focused exclusively on head and neck cases, ranging from 96 to 98%.[10–13,15] However, there are few studies involving large academic centers. There is also little consensus on the best methodology for evaluating frozen sections or on ways to reduce errors.[20]

Two types of errors occur during frozen section analysis on defect sampling margin specimens. The first is sampling error in which the levels of tissue examined during frozen section contain no tumor, but the permanent section performed later on the remaining tissue does. The second is interpretation error in which the tumor is present on the frozen section slide but is misdiagnosed by the pathologist or, less commonly, where no tumor is present but the pathologist diagnoses it as present.

We retrospectively studied 3.5 years of head and neck defect sampling margin frozen sections at our academic institution, which used two-level sectioning and compared these with one year of three-level sectioning. Our analysis was intended to establish our baseline error rates and to see if three-level sectioning could decrease these rates.

Materials and methods

This study was approved by the Human Research Protection Office of Washington University. The first goal of this study was to determine baseline head and neck margin frozen section error rates at our institution, Barnes–Jewish Hospital. All head and neck tumor resection cases from January 2005 through July 2008 were identified by a database search, which was specified to capture all specimens received from the Otolaryngology Head and Neck Surgery department but only those which had some form of intra-operative consultation. We included patients with any tumor type, whether benign or malignant or squamous or non-squamous for which margins were submitted for frozen

section. The only head and neck anatomic subsites that were excluded were brain and spine. The cases came from 11 different head and neck surgeons and were all evaluated in our routine departmental manner. This included cases reviewed by all departmental faculty who cover the frozen section service (24 different surgical pathologists, most of whom are not specifically head and neck pathology specialists). Two slides (levels) were cut for each margin specimen and examined at the time of frozen section by the staffing resident(s) and fellow(s) and by the attending pathologist(s) (Figure 1). Beginning in August 2008, a third 'deep' level was cut on all frozen section head and neck margins (Figure 2). This was initiated in a joint pathology–clinical endeavor to reduce error rates. For both the two-level and three-level periods, the remaining tissue was handled in the same manner, being fixed in formalin, and a single permanent slide cut following standard overnight tissue processing and paraffin embedding.
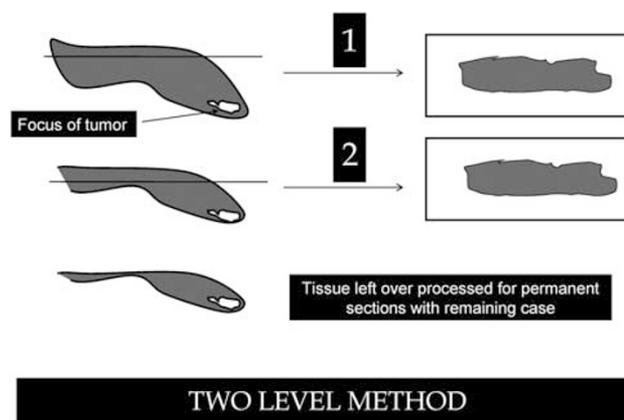

Figure 1 Graphical representation of the two-level method. In this hypothetical example, the tumor is deep in the tissue and was missed when only two levels were taken.
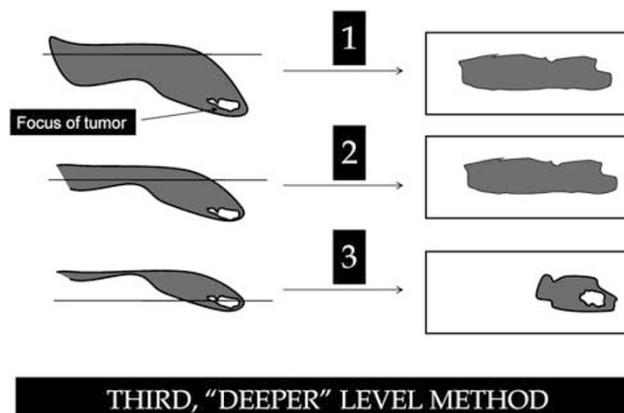

Figure 2 Graphical representation of the three-level method. In this hypothetical example, cutting an additional level during frozen section analysis allows the tiny focus of tumor to be identified intra-operatively where it would not have been if only two levels were taken.

The Copath reports on all resulting cases from this search were reviewed for the study. Sampling and interpretation errors were identified based on the reports noting them as one type of error or the other. If there was a discrepancy in the report between the frozen and permanent sections, but the reason was not specifically noted, the actual slides were reviewed by one study pathologist (JSL) to determine if the error was sampling or interpretation. All discrepancies between the presence or absence of tumor were captured. Discrepancies involving squamous dysplasia were considered errors and included only if they involved moderate or severe dysplasia. We did not consider discrepancies between diagnoses of negative and those of mild dysplasia as errors.

Squamous cell carcinoma cases were divided for analysis into different histologic types. Squamous cell carcinomas of the oropharynx were divided into keratinizing and non-keratinizing types as previously reported.[21] For this study, cases with hybrid features (or 'non-keratinizing squamous cell carcinoma with maturation') were combined with the strictly non-keratinizing type as both are almost always p16 positive and have similar outcomes. All non-oropharyngeal squamous cell carcinomas were of the keratinizing type or were specific variants of squamous cell carcinoma such as basaloid, verrucous, adenosquamous, papillary, or spindle cell carcinoma.

Statistical analysis was performed using two-tailed Fisher's exact tests to compare categorical data.

Results

For the two-level sectioning period (43 months), there were 647 total cases, which included 3758 frozen section margin specimens. There were 90 errors (2.4% of all cases). In all, 60 of the errors (66%) were sampling errors for an overall sampling error rate of 1.6%. There were 30 interpretation errors (0.8%). No errors were made on bone marrow margins. These results are summarized in Table 1.

For the three-level sectioning period (12 months), there were 204 total cases, which included 1218

total frozen sections. There were 31 errors (2.5%). This was almost identical to the two-level sectioning period ($P = 0.67$). There were 15 sampling errors (1.2%). Although this was lower than the 1.6% rate of the two-level sectioning period, this difference was not statistically significant ($P = 0.42$). The interpretation error rate was 1.3%, which was actually slightly higher than for the two-level sectioning period (0.8%), but this difference was not statistically significant ($P = 0.12$). The errors for the entire study period are shown in Figure 3.

The error rates can also be calculated as a percentage of the margins that were eventually shown to be positive (ie, only those where tumor was present on frozen section, permanent section, or both, and excluding those specimens where no tumor or dysplasia was identified). When calculated in this fashion, for the two-level period, there were 90 overall errors among 393 cases (22.9%), and for the three-level period, 31 errors among 204 cases (15.2%). This difference was statistically significant ($P = 0.03$). This included 60 and 15 sampling errors (15.3 and 7.4%; $P = 0.006$), respectively and 30 and 16 interpretation errors (7.6 and 7.8%; $P = 1.0$), respectively.

Interestingly, 16.7% of the total frozen section margins were found to harbor tumor during the
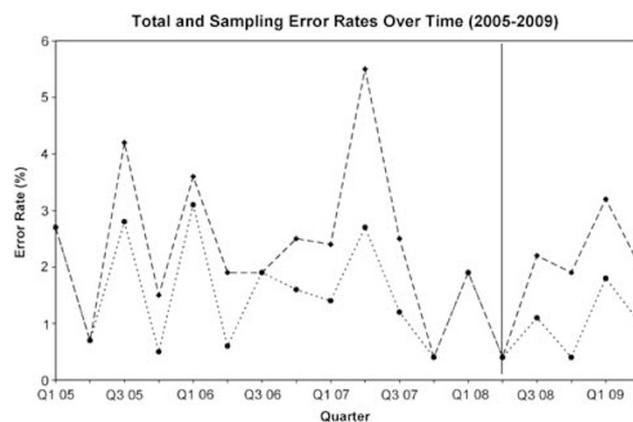

Figure 3 Error rates for the entire study period by annual quarters (Total error rates are represented by the dashed line; sampling error rates represented by the dotted line; vertical solid line indicates time of switch to three-level sectioning).

Table 1 Frozen section analysis for the two-level sectioning period.

	Total margins	Total margins with dysplasia or tumor	Total errors	Sampling errors	Interpretation errors
All margins	3758	393	90 (2.4%, 23%)	60 (1.6%, 15.3%)	30 (0.8%, 7.6%)
Soft tissue margins	3697	386	90 (2.6%, 23%)	60 (1.6%, 15.3%)	30 (0.8%, 7.6%)
Bone margins	61	7	0	0	0
Squamous cell carcinoma, keratinizing type	2511	254	67 (2.7%, 26.4%)	47 (1.9%, 18.5%)	20 (0.8%, 7.9%)
Non-keratinizing squamous cell carcinoma	384	40	8 (2.1%, 20%)	5 (1.3%, 12.5%)	3 (0.8%, 7.5%)
Other tumors	863	99	15 (1.7%, 15%)	8 (0.9%, 8.1%)	7 (0.8%, 7.1%)

During this period, two levels were obtained during frozen section. The first error percentage in parentheses is out of all frozen section margins, the second is out of positive margins only.

Table 2 Frozen section analysis for three-level sectioning period.

	Total margins	Total margins with dysplasia or tumor	Total errors	Sampling errors	Interpretation errors
All margins	1218	204	31 (2.5%, 15.2%)	15 (1.2%, 7.4%)	16 (1.3%, 7.8%)
Soft tissue margins	1208	204	31 (2.6%, 15.2%)	15 (1.2%, 7.4%)	16 (1.3%, 7.8%)
Bone margins	10	0	0	0	0
Squamous cell carcinoma, keratinizing type	701	98	11 (1.6%, 11.2%)	3 (0.4%, 3.1%)	8 (1.1%, 8.2%)
Non-keratinizing squamous cell carcinoma	147	43	8 (5.4%, 18.6%)	6 (4.1%, 14%)	2 (1.4%, 4.7%)
Other tumors	370	63	12 (3.2%, 19%)	6 (1.6%, 9.5%)	6 (1.6%, 9.5%)

During this period, three levels were obtained during frozen section. The first error percentage in parentheses is out of all frozen section margins, the second is out of positive margins only.

three-level period compared with only 10.5% during the two-level period. This difference was statistically significant ($P < 0.0001$).

The results stratified by specific histologic subtypes of tumor are presented in Tables 1 and 2. Total error rates were higher in non-keratinizing squamous cell carcinoma (3.0%) *versus* keratinizing squamous cell carcinoma (2.4%) although the difference was not statistically significant ($P = 0.67$). There was also no statistically significant difference between total error rates in keratinizing squamous cell carcinoma *versus* other tumors (which includes all non-squamous cell tumors) ($P = 0.74$). When examining the difference in error rates by tumor subtype, only keratinizing SCC showed a significant difference between the two-level method (18.5% of positive margins showed sampling errors) and the three-level method (3.1% of positive margins showed sampling errors, $P < 0.0001$). For non-keratinizing squamous cell carcinoma and the other tumors, the total error rate and sampling error rate remained similar and any differences were not statistically significant.

Our surgeons also sample bone marrow or cortex by burring/curettage for frozen section, and these specimens were included. Frozen section is performed as usual, although the material on the slides is typically small because of fragmentation from the calcified material. There were 61 such bone margins during the two-level period and 10 during the three-level period. 9.9% of these had tumor in them. There were no errors among any of these specimens.

Discussion

One would speculate that 'the more you sample, the more you find,' and the goal of frozen section is just that—to find tumor in the specimen if it is present. We did more frequently find tumor in margin specimens with our three-level method, although the number of cases is admittedly modest. The sampling error rate was lower in the three-level period than in the two-level period, but this difference was not statistically significant when the errors were considered as a percentage of all frozen section margins ($P = 0.42$). However, both the overall and sampling error rates were statistically significantly lower with the three-level method when considering them only as a percentage of tumor-bearing specimens ($P = 0.03$ and $P = 0.006$, respectively). Furthermore, a graph of the error rates by quarter over time does show a trend of error reduction even when errors are calculated as a percentage of all frozen sections (Figure 3). The interpretation error rate was, as would be expected, unchanged with this different method.

One could conceivably draw two different, and completely disparate, conclusions from our findings, however. One could conclude, as we do, that adding a third level does reduce sampling error rates. Sampling errors can only occur on specimens bearing tumor. It seems possible that this study did not find a statistically significant lowering of sampling errors when considering all specimens because the large pool of margin specimens that have no tumor in them dilutes out the pool of cases used in calculating the sampling error percentage. If one assumes that the cases in which no tumor was identified by frozen or by the permanent section are either completely (or even just predominantly) ones where no tumor was ever present in the tissue, then these are ones where no tumor could have been found by any method of sectioning. So, when we considered just the cases that had tumor in them, the reduction in sampling errors was statistically significant ($P = 0.0005$) with three-level sectioning.

One could alternatively conclude that the difference in sampling error rates between the two periods is so modest that it justifies and, in a sense, validates, two-level sectioning as an adequate method for evaluating defect sampling-type frozen section specimens. This is not an unreasonable conclusion, either, in our opinion. Clearly, there is no practical way to eliminate sampling errors completely as this would require sectioning in all tissue, an untenable proposition in terms of both time and staff resource utilization.

In an attempt to better analyze the impact of our findings, we considered the following hypothetical analysis. If one considers that in the three-level period, there were one-third the number of the overall cases compared with the two-level period (1250 *versus* 3700), and there were 15 sampling

errors *versus* 60 in the two-level period. So if you multiply the results of the three-level sectioning period by three, you should have 45 total sampling errors over a 3-year period. That means that you would have sectioned and read 3700 extra slides over a 3-year period and would have caught only 15 margins that would have otherwise been missed (would have had sampling errors). Put another way, that is about 15 patients spared an error out of a total of approximately 650 total patients, or approximately 1 in 40–50 patients. We looked at the clinical records of the 14 patients for whom sampling errors actually did occur during the three-level sectioning period, and four of these patients were taken back to the operating room for re-resection. If one also then assumes that 4 of the 15 patients for whom a sampling error was likely avoided in the three-level sectioning period were thus spared a re-operation, sizeable money would have been saved. Considering the surgeon, pathology, anesthesia, and operating room fees, and finally, hospital fees assuming a 23-h admission for the patients, we estimate this to be as much as 60 to 80 000\$ (15 to 20 000\$ per patient in billing). Is this actually cost effective overall, then? We did not aim to specifically evaluate the cost effectiveness of our method change, but one could make a reasonable argument in either direction.

One must also consider other reasons why we may have found more positive margins during the three-level period, although we do not think any of these potential alternative possibilities explain our results. Rather than the method itself resulting in more positive margins, it may rather reflect some change in the surgeons' approaches, the makeup of the different surgeons operating over these periods, or perhaps a change in actual surgical methods. Recently, a greater number of surgeries are being performed using transoral laser microsurgery. It is possible that some aspect of this technique is resulting in a greater number of tumor-bearing margin specimens. For example, surgeons may be performing more focused surgeries such as by transoral laser microsurgery as opposed to wide resections (such as total laryngectomies), thus leading to a higher percentage of tumor-bearing margin specimens. This would not, however, explain the decrease in sampling error rates.

Other consequences of cutting an additional level must also be considered. Increased cost and increased turnaround time are both potential negative aspects of cutting an additional level. The increased cost is small as it includes only an additional slide, a small amount of additional staining reagent and some additional wear on the cryostat. As in our hypothetical cost analysis above, we do not think that the costs of additional slide preparation and reading would be more than those of the costs of surgery for the few patients who need reoperation because of sampling errors on their margins. Increased turnaround time is more difficult to quantify and, unfortunately, over the study period,

we did not have the types of records necessary to accurately measure this. However, in our experience, the increased turnaround time was not significant enough to yield complaints from surgeons or staff, who, to the contrary, have commented in appreciation of the decrease in error rates. Head and neck surgical resections are complicated and time consuming with many needing lengthy post-resection reconstruction so the addition of a small amount of time for a third level, in our opinion, does not impact the process to any significant degree. Our pathologists, initially somewhat hesitant, have found that having a third H&E slide means that there will be another chance at a good, full, less- or un-folded level to evaluate. It has decreased the need the frozen section team to go back to the cryostat to cut additional sections.

Another concern in using this method is that leveling through the tissue during frozen section analysis could result in small foci of tumor being 'left in the cryostat,' resulting in falsely negative margins. The results of this study argue against this phenomenon, however, as the rate of positive margins actually increased (from 10.5 to 16.7%—an approximately 60% increase). With our method, we are sampling the tissue with more total sections (four) than in previous years (three). If tumor foci were being missed, the positive margin rate should have decreased or at least been relatively similar.

Ours is the largest study of frozen section error rates with 4976 total frozen sections included. This is more than twice as many as that included in the previous largest study,[12] which also focused on head and neck cases. Overall, these results show that frozen section error rates at Barnes–Jewish Hospital compare very favorably with those previously published. Overall frozen section accuracy varies between 96 and 99%. In studies examining head and neck cases only, the accuracy varies from 96 to 98%. Some of these studies focused only on specific types of head and neck specimens, including laryngectomies,[10] and oral squamous cell carcinoma.[14] Others were similar to this study in that they included head and neck cases from all regions.[11,12,15] The latter studies included greater numbers of frozen sections (420–2210) than the former and showed error rates from all head and neck subsites. While our study only examined the accuracy of frozen section margins, other studies also included diagnostic specimens.[12,15] Unfortunately, these latter studies did not separate out the margin specimens from the ones for tumor diagnosis so a true comparison between error rates cannot be made. Nevertheless, our overall accuracy of 97.5% compares favorably with any of these studies.

In summary, in a very large cohort of frozen sections, the addition of a third level during frozen section analysis reduced the sampling error rate by a modest degree. Although there may be some minor drawbacks in terms of turnaround time and reagent use, our analysis and experience suggest that the

benefits to the patients and the potential for medical expenditure savings likely outweigh these issues.

Disclosure/conflict of interest

The authors declared no conflict of interest.

References

1 Bradley PJ, MacLennan K, Brakenhoff RH, *et al.* Status of primary tumour surgical margins in squamous head and neck cancer: prognostic implications. Curr Opin Otolaryngol Head Neck Surg 2007;15:74–81.

2 Cook JA, Jones AS, Phillips DE, *et al.* Implications of tumour in resection margins following surgical treatment of squamous cell carcinoma of the head and neck. Clin Otolaryngol 1993;18:37–41.

3 Looser KG, Shah JP, Strong EW. The significance of 'positive' margins in surgically resected epidermoid carcinomas. Head Neck Surg 1978;1:107–111.

4 Ackerman LV, Ramirez GA. The indications for and limitation of frozen section diagnosis. Br J Surg 1959; 46:336–350.

5 Howanitz PJ, Hoffman GG, Zarbo RJ. The accuracy of frozen section diagnosis in 34 hospitals. Arch Pathol Lab Med 1990;114:355–359.

6 Black C, Marotti J, Zarovnaya E, *et al.* Critical evaluation of frozen section margins in head and neck cancer resections. Cancer 2006;107:2792–2800.

7 Yahalom R, Dobriyan A, Vered M, *et al.* A prospective study of surgical margin status in oral squamous cell carcinoma: a preliminary report. J Surg Oncol 2008; 98:572–578.

8 Werner JA, Dunne AA, Folz BJ, *et al.* Transoral laser microsurgery in carcinomas of the oral cavity, pharynx, and larynx. Cancer Control 2002;9:379–386.

9 Upile T, Fisher C, Jerjes W, *et al.* The uncertainty of the surgical margin in the treatment of head and neck cancer. Oral Oncol 2007;43:321–326.

10 Cooley ML, Hoffman HT, Robinson RA. Discrepancies in frozen section mucosal margin tissue in laryngeal squamous cell carcinoma. Head Neck 2002;24: 262–267.

11 Remsen KA, Lucente FE, Biller HF. Reliability of frozen section diagnosis in head and neck neoplasms. Laryngoscope 1984;94:519–524.

12 DiNardo LJ, Lin J, Karageorge LS, *et al.* Accuracy, utility, and cost of frozen section margins in head and neck cancer surgery. Laryngoscope 2000;110: 1773–1776.

13 Gandour-Edwards R, Donald PJ, Wiese D. The accuracy and clinical utility of frozen section diagnosis in head and neck surgery. Experience at a university medical center. Head Neck 1993;15:33–38.

14 Ikemura K, Ohya R. The accuracy and usefulness of frozen section diagnosis. Head Neck 1990;12:298–302.

15 Ord RA, Aisner S. Accuracy of frozen sections in assessing margins in oral cancer resection. J Oral Maxillofac Surg 1997;55:663–669.

16 Ahmad Z, Barakzai MA, Idrees R, *et al.* Correlation of intra-operative frozen section consultation with the final diagnosis at a referral center in Karachi, Pakistan. Indian J Pathol Microbiol 2008;51:469–473.

17 Coffin CM, Spilker K, Zhou H, *et al.* Frozen section diagnosis in pediatric surgical pathology: a decade's experience in a children's hospital. Arch Pathol Lab Med 2005;129:1619–1625.

18 Caya JG. Accuracy of breast frozen section diagnosis in the community hospital setting: a detailed analysis of 628 cases. Wis Med J 1991;90:58–61.

19 Dankwa EK, Davies JD. Frozen section diagnosis: an audit. J Clin Pathol 1985;38:1235–1240.

20 Gandour-Edwards RF, Donald PJ, Lie JT. Clinical utility of intraoperative frozen section diagnosis in head and neck surgery: a quality assurance perspective. Head Neck 1993;15:373–376.

21 Chernock RD, El-Mofty SK, Thorstad WL, *et al.* HPV-related nonkeratinizing squamous cell carcinoma of the oropharynx—utility of microscopic features in predicting patient outcome. Head Neck Pathol 2009; 3:186–194.