Sources of Error in Neuropathology Intraoperative Diagnosis

Matthew Meyer, Julia Keith-Rokosh, Hasini Reddy, Joseph Megyesi, Robert R. Hammond

ABSTRACT: *Objective:* The goal of this study was to optimize intraoperative neuropathology consultations by studying trends and sources of diagnostic error. We hypothesized that errors in intraoperative diagnoses would have sampling, technical, and interpretive sources. The study also audited diagnostic strengths, weaknesses and trends associated with increasing experience. We hypothesized that errors would decline and that the accuracy of "qualified" diagnoses would improve with experience. *Methods:* The pathologist's first 100 cases (P1), second 100 (P2), and most recent 100 (P3, after ten years in practice) formed the data set. Intraoperative diagnoses were scored as correct, minor error or major error using the final diagnosis as the gold-standard. Incorrect diagnoses were re-examined by two reviewers to identify sources of error. *Results:* Among the 300 cases there were 22 errors with 11 in P1, 9 in P2 and 2 in P3. Sampling contributed to 17 errors (77%), technical factors to 7 (32%) and interpretive factors to 16 (73%). Improvement in diagnostic accuracy between P1 and P2 (p=0.8143), or P2 and P3 (p=0.0582) did not reach significance. However, significant improvement was found between P1 and P3 (p=0.0184). *Conclusion:* The present study was a practical and informative audit for the pathologist and trainees. It reaffirmed the accuracy of intraoperative neuropathology diagnoses and informed our understanding of sources of error. Most errors were due to a combination of sampling, technical and interpretive factors. A significant improvement in diagnostic proficiency was observed with increasing experience.

RÉSUMÉ: Sources d'erreurs dans le diagnostic neuropathologique extemporané. *Objectif :* Le but de cette étude était d'optimiser les consultations neuropathologiques extemporanées en étudiant les tendances et les sources d'erreurs de diagnostic. Nous avons émis l'hypothèse que les sources d'erreurs de diagnostic extemporané proviennent de l'échantillonnage, de la technique et de l'interprétation. L'étude a également vérifié les forces diagnostiques, les faiblesses et les tendances associées à des niveaux croissants d'expérience. Nous avons émis l'hypothèse que le nombre d'erreurs diminuerait et que la précision des diagnostics "vérifiés" s'améliorerait avec l'expérience. *Méthodes :* Les données analysées incluent les 100 premiers cas examinés par un même pathologiste (P1), les 100 cas suivants (P2) et les 100 cas les plus récents (P3, après 10 ans de pratique). Les diagnostiques extemporanés ont été classifiés comme étant exacts, comportant une erreur mineure ou une erreur majeure par rapport au diagnostic final utilisé comme référence. Les diagnostics incorrects ont été réexaminés par deux réviseurs pour identifier les sources d'erreurs. *Résultats :* Parmi les 300 cas, il y avait 22 erreurs dont 11 dans le groupe P1, 9 dans le groupe P2 et 2 dans le groupe P3. L'échantillonnage a contribué à 17 erreurs (77%), des facteurs techniques à 7 (32%) et des facteurs d'interprétation à 16 (73%). L'amélioration de l'exactitude du diagnostic entre P1 et P2 (p = 0,8143), ou P2 et P3 (p = 0,0582) n'était pas significative au point de vue statistique. Cependant, une amélioration importante a été constatée entre P1 et P3 (p = 0,0184). *Conclusion :* Cette étude constitue une vérification pratique et instructive pour le pathologiste et les résidents. Elle confirme l'exactitude des diagnostics neuropathologiques extemporanés et éclaire notre compréhension des sources d'erreurs. La plupart des erreurs étaient dues à une combinaison de facteurs reliés à l'échantillonnage, à la technique et à l'interprétation. Une amélioration significative de la compétence diagnostique a été observée à mesure que l'expérience croissait.

Specialists institute a variety of diagnostic audits to identify areas of weakness and to optimize clinical performance[1-5]. In pathology, one common diagnostic event well-suited to this type of analysis is the intraoperative consultation where touch, smear and frozen section preparations have long been used to assist surgical decision making[6-9].

Neurosurgeons incorporate a neuropathologist's interpretation into the surgical plan through intraoperative consultations on small biopsies[5,6,8,10]. These consultations represent a diagnostic challenge for the neuropathologist that imposes sampling and technical limitations under a relative time constraint. Without the luxury of optimal tissue processing and ancillary tests (special stains, immunohistochemistry, electron microscopy, or molecular techniques) intraoperative

consultations rely on the gross and microscopic interpretive skills of the pathologist and, at times, their ability to weigh clinical data. While intraoperative pathology consultations have proven to be an accurate and reliable diagnostic tool[8,11-14],

From the Department of Pathology (RRH) and Clinical Neurological Sciences (RRH, JM), Department of Epidemiology and Biostatistics (MM), Schulich School of Medicine and Dentistry, University of Western Ontario, London Health Sciences Centre, London; Department of Pathology (JKR), Sunnybrook Health Sciences Centre, University of Toronto, Toronto, Ontario, Canada; Department of Pathology (HR), Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, Massachusetts, USA.

accuracy is variable and dependant upon the pathology[15-17] and the pathologist[2,3,14].

Intraoperative consultations are followed shortly (typically within 24 to 48 hours) by "permanent sections" and a "final diagnosis" using fixed, paraffin-embedded samples. Unfortunately, intraoperative diagnostic errors have the potential to misguide patient care until a revised final diagnosis is passed. These errors are usually quickly addressed, which in addition to optimizing patient care provides case-specific feedback to the pathologist. However, there is often no system for cumulative longitudinal feedback.

Paired diagnostic events over time provide an opportunity to observe procedural limitations, contributing factors, diagnostic trends as well as the relevance of experience and diagnostic phraseology. This study aimed to perform a longitudinal analysis of diagnostic accuracy during intraoperative consultation for a single neuropathologist. We hypothesized that errors in intra-operative diagnosis could be secondary to sampling, technical, and interpretive factors. We also hypothesized that the error rate would decline with increasing experience and that the accuracy of "qualified" diagnoses would improve.

METHODS

Cases were retrieved by an electronic search of the Pathology archives of the London Health Sciences Centre for all intraoperative consultations made by one neuropathologist. The study pathologist and reviewers received their training in neuropathology in Canada at the University of Western Ontario. All intraoperative diagnoses rendered in the present study were made by the study pathologist. Final diagnoses were rendered by the study pathologist in 90% of cases and by colleagues or collective opinion in the remainder. The first (P1), second (P2) and most recent (P3) 100 suitable cases were analyzed. Cases previously known to the study pathologist were excluded so as to not inflate measures of accuracy. For similar reasons, in surgeries requiring multiple intraoperative consultations, the diagnosis of record was the final intraoperative diagnosis rendered.

It is the study pathologist's custom to examine the tissue in a blinded fashion to arrive at a specific diagnosis without clinical history or radiographic details. The latter are available on all cases at the time of consultation, but are used after arriving at a tissue diagnosis as ancillary information with which to provide a clinicopathological correlation as needed. All cases are examined by frozen section and smear preparation with a hematoxylin and eosin stain.

Specific diagnoses were placed into nine generalized sub-groups to facilitate data analysis: high grade glioma, WHO grades III and IV), low grade glioma (WHO grades I and II), metastatic, lymphoma, high grade other (for example, sarcoma), low grade other (for example, meningioma), infectious or reactive, neuronal and glioneuronal tumours and non-neoplastic other (for example, gliosis). Comparison of the distribution of subgroups was conducted to ensure consistency in case mix between time periods.

Each intraoperative diagnosis was placed into one of three categories; correct, minor error or major error as compared to the final diagnosis. Incorrect diagnoses were judged to be of potentially minor or major significance on clinical-pathological grounds. Major errors differed from minor ones primarily by their potential to alter the operative plan and/or perioperative care (until the final diagnosis was communicated). Minor errors were reclassifications of little or no intra- or perioperative significance. Incorrect diagnoses were analyzed for sources of error and any misinterpretive trends. In brief, two senior trainees (JK and HR) independently reviewed all pathological materials from all incorrect diagnoses and recorded whether in their opinion sampling, technical or interpretive factors played a role. If either reviewer felt that a given diagnostic factor was

Table 1: The table lists the diagnostic subgroups and their frequencies in each of the three periods studied (P1, P2 and P3)

Diagnostic Group	# of Cases P1	% Correct P1	# of Cases P2	% Correct P2	# of Cases P3	% Correct P3	Totals
HGG	30	93%	26	93%	33	97%	**89 (30%)**
LGG	13	77%	11	100%	8	100%	**32 (11%)**
MET	9	78%	15	93%	15	100%	**39 (13%)**
LYM	4	75%	6	83%	4	100%	**14 (5%)**
HGO	5	80%	4	50%	2	100%	**11 (4%)**
LGO	30	93%	28	89%	28	96%	**86 (29%)**
IR	5	100%	3	100%	2	100%	**10 (3%)**
MISC	4	100%	7	86%	8	100%	**19 (6%)**
Totals	**100**	**89%**	**100**	**91%**	**100**	**98%**	**300**

HGG = high grade glioma; LGG = low grade glioma; MET = metastatic tumour; LYM = lymphoma; HGO = high grade other; LGO = low grade other; IR = infectious or reactive; MISC = miscellaneous

Table 2: The table summarizes each revised diagnosis by error class and source of error

Time Period	Intraoperative Diagnosis	Final Diagnosis	Error class		Source of error		
			Minor	Major	Sampling	Interpretive	Technical
P1	Anaplastic Astrocytoma	Glioblastoma	x			x	
	Anaplastic Oligodendroglioma	Glioblastoma	x		x		
	Astrocytoma	Oligodendroglioma	x		x	x	
	Meningioma	Hemangioma		x	x	x	x
	Metastatic tumour	Plasmacytoma		x		x	
	Low grade glioma	Dysembryoplastic neuroepithelial tumour	x		x	x	x
	Lymphoma	Rhabdoid		x		x	
	Schwannoma	Meningioma	x			x	x
	Abscess	Metastatic carcinoma		x	x	x	
	High grade glioma	Metastatic carcinoma		x	x	x	
	Inflammation	Ganglioglioma		x	x	x	
P2	Neurofibroma	Schwannoma	x		x		x
	High grade glioma	Primitive neuroectodermal tumour	x		x	x	x
	Pilocytic astrocytoma	Anaplastic mixed glioma		x	x		x
	No tumour present	Craniopharyngioma		x	x		
	Astrocytoma	Lymphoid infiltrate		x	x	x	
	Inflammation	Metastatic carcinoma		x	x	x	
	Reactive changes	Vasculitis		x		x	x
	Low grade glioma	Hemangioblastoma	x		x		
	Pleomorphic xanthoastrocytoma	Malignant fibrous histiocytoma		x	x	x	
P3	Oligodendroglioma	Anaplastic oligodendroglioma		x	x		
	Schwannoma	Meningioma	x		x	x	
		Totals	**9**	**13**	**17**	**16**	**7**

significant in a particular case, it was scored as such. Sampling factors were deemed to be significant if essential elements for the final diagnosis were not present in the sample examined intraoperatively. Sampling is primarily influenced by the surgeon, in consideration of anatomy, radiographic features and biopsy modality. At times and on a smaller scale, sampling is influenced by the pathologist if only part of a biopsy is selected for examination at the time of the intraoperative consult. Technical factors were deemed significant if essential diagnostic elements were obscured by handling or staining artifact. Interpretive factors were deemed significant if, given the final histopathology, the diagnosis could possibly have been reached, even if only in retrospect.

Although we consider intraoperative diagnoses to be preliminary, intraoperative diagnoses frequently included qualifiers ("favour", "probable", "query", "defer to permanent sections") within the diagnosis to communicate a degree of greater uncertainty. Intraoperative diagnoses not using these terms were regarded as relatively "confident". For each of the three time periods, the accuracy of qualified versus confident diagnoses was compared.

Case-mix distribution was tested using two tailed chi-squared (χ^2) analysis at α=0.05. Due to the low frequencies of misdiagnoses observed, diagnostic accuracy (both over time and amongst "qualified" diagnoses) was performed using a two-tailed, mid-p Fisher's exact test at α=0.05. All analyses were performed with SAS version 9.2 for Windows (SAS Institute Inc. Cary, NC, USA).

RESULTS

According to the permanent diagnoses rendered, the three sample sets (P1, P2 and P3) were comparable in diagnostic range and difficulty (p=0.817, Table 1). The percentage of biopsies performed by needle or stereotaxy increased from 7% to 23% between periods P1 and P2 and has thereafter remained stable, accounting for approximately 25% of all biopsies.

Of the 22 minor and major diagnostic errors (Table 2), 17 were influenced by sampling factors (7 in P1, 8 in P2 and 2 in P3), 16 by interpretive factors (10 in P1, 5 in P2 and 1 in P3) and 7 by technical factors (3 in P1, 4 in P2). Fifteen were judged to be due to a combination of two or three factors (sampling, technical and interpretive). The sole source of diagnostic error was sampling in four cases (1 in P1, 2 in P2 and 1 in P3) and interpretive in three (all in P1). Technical aspects were never judged to be the sole source of diagnostic error.

Eleven errors were made in P1 (6 major, 5 minor), nine in P2 (6 major, 3 minor), and two in P3 (1 major, 1 minor). Diagnostic accuracy (defined as no minor or major diagnostic error) showed

no significant improvement between P1 and P2 (p=0.8143), failed to reach statistical significance between P2 and P3 (p=0.0582), but was found to be significantly improved between P1 and P3 (p=0.0184). The level of accuracy between diagnostic subgroups varied although modest improvement was noted in most subgroups over time (Table 1).

Diagnostic qualifiers ("favour", "probable", "query", "defer to permanent sections") were used at a consistent frequency over time: 40% in P1 compared with 41% in P2 and 44% in P3. The accuracy of qualified diagnoses was not significantly different from that for unqualified diagnoses in any of the time periods. Diagnostic accuracy amongst qualified versus "confident" diagnoses was 88% vs. 90% (p=0.7508), 88% vs. 93% (p= 0.4776), and 95% vs. 100% (p=0.1911) in P1, P2, and P3 respectively.

DISCUSSION

Frozen section diagnoses are an important element in surgical management and an invaluable skill set for pathology departments. They incorporate technical and medical expertise and when compared with final diagnoses, represent a practical opportunity to audit a pathology practice. Thanks to electronic databases, recent data of this nature is readily obtained.

In re-examining 22 minor and major diagnostic errors for the present study (Table 2), the majority (68%) were judged to be due to a combination of two or all three factors (sampling, technical and interpretive) confirming our first hypothesis. Sampling was the sole source of diagnostic error in four cases while interpretive factors were the sole source in three (all occurring in P1). Technical aspects were never judged to be the sole source of misinterpretation. The latter is most likely due to the fact that most technical artifacts can be remedied immediately at the time of consult.

The four errors principally due to sampling (1 in P1, 2 in P2 and 1 in P3) were evenly distributed over time. Sampling was also judged to be a significant factor in an additional 13 cases. These instances call attention to the need for continued diligence in technical precision of sample preparation and effective dialogue between the radiologist, surgeon and pathologist to ensure optimal lesional sampling. With the trend towards smaller open or stereotactic biopsies in neuropathology, diagnostic errors due to sampling bias will continue to occur but must be kept to a minimum.

The results of this study confirm our hypothesis that experience would result in reduced diagnostic error over time. Interpretive factors contributed to diagnostic error in 16 cases in total (10 in P1, 5 in P2 and 1 in P3). No errors due solely to interpretive factors occurred after P1. Assuming that the influence of sampling and technical factors was relatively constant, the statistical improvement in accuracy over time provides further evidence of an association between interpretive skill and procedural accuracy (of the intraoperative consultation) over time.

Intraoperative diagnostic qualifiers (favour, probable, etc.) were intended to communicate a greater degree of uncertainty in the interpretation. However, the present study suggests that no difference in accuracy exists between qualified and "confident" diagnoses, thereby questioning the relevance of such qualifiers in diagnostic phraseology. Instead, it could be argued that all parties

would benefit from the mindset that intraoperative diagnoses are, by definition, preliminary. Nonetheless, it is recognized that the relevance and application of diagnostic qualifiers may vary by pathologist and by specialty.

Potential limitations of the study include possible bias on the part of the case reviewers. The vast majority (90%) of cases with intraoperative consultations were finalized by the study Pathologist as is customary in our group practice. Although routine consultation was performed on challenging cases, the potential for bias in arriving at the final diagnosis remains. Similarly, professional relationships between the study pathologist and the reviewers introduces another potential source of bias. Future similar studies may be strengthened by recruiting external reviewers, blinded to the identity of the study pathologist, to improve validity. The generalizability of the present study may be limited by variable preferences in diagnostic phraseology, in that individual pathologists may routinely elect to use less specific terminology (e.g. "neoplasm") in intraoperative consultations, reducing the likelihood of discrepant diagnoses. Finally, the effect of optimized case selection and lesional sampling by advances in imaging and image-guidance were not measured. This may have influenced the accuracy of intraoperative consultations over time.

Prospectively or by retrospective electronic searches, pathologists and other physicians can readily study analogous diagnostic patterns and interpretive skills in their own practice and in so doing, optimize this element of patient care. For trainees, the experience consolidates their own intraoperative consultative skills and provides further insight into diagnostic pitfalls and limitations using actual case material.

CONCLUSIONS

The present study confirms that intraoperative neuro-pathology consultation remains an accurate tool. The longitudinal comparison of intraoperative and final diagnoses proved to be a simple and practical audit of diagnostic accuracy and interpretive and technical skills. Re-examining cumulative intraoperative diagnostic errors was highly instructive in providing a glimpse into personal diagnostic trends and greater insight into the limitations of intraoperative consultations.

REFERENCES

1. Evans CA, Suvarna SK. Intraoperative diagnosis using the frozen section technique. J Clin Pathol. 2006; 59(3):334.
2. Khoo JJ. An audit of intraoperative frozen section in Johor. Med J Malaysia. 2004; 59(1):50-5.
3. Wen MC, Chen JT, Ho WL. Frozen-section diagnosis in surgical pathology: a quality assurance study. Kaohsiung J Med Sci. 1997; 13(9):534-9.
4. Raab SS, Tworek JA, Souers R, Zarbo RJ. The value of monitoring frozen section-permanent section correlation data over time. Arch Pathol Lab Med. 2006; 130(3):337-42.

5. Tilgner J, Herr M, Ostertag C, Volk B. Validation of intraoperative diagnoses using smear preparations from stereotactic brain biopsies: intraoperative versus final diagnosis--influence of clinical factors. Neurosurgery. 2005; 56(2):257-65.

6. Brender E, Burke A, Glass RM. Frozen section biopsy. JAMA. 2005; 294(24):3200.

7. Powell SZ. Intraoperative consultation, cytologic preparations, and frozen section in the central nervous system. Arch Pathol Lab Med. 2005; 129(12):1635-52.

8. Ferreiro JA, Myers JL, Bostwick DG. Accuracy of frozen section diagnosis in surgical pathology: review of a 1-year experience with 24,880 cases at Mayo Clinic Rochester. Mayo Clin Proc. 1995; 70(12):1137-41.

9. Kim JE, Kim DG, Paek SH, Jung HW. Stereotactic biopsy for intracranial lesions: reliability and its impact on the planning of treatment. Acta Neurochir (Wien). 2003; 145(7):547-54.

10. Heper AO, Erden E, Savas A, Ceyhan K, Erden I, Akyar S, et al. An analysis of stereotactic biopsy of brain tumors and nonneoplastic lesions: a prospective clinicopathologic study. Surg Neurol. 2005; 64 Suppl 2:S82-8.

11. Savargaonkar P, Farmer PM. Utility of intra-operative consultations for the diagnosis of central nervous system lesions. Ann Clin Lab Sci. 2001; 31(2):133-9.

12. Folkerth RD. Smears and frozen sections in the intraoperative diagnosis of central nervous system lesions. Neurosurg Clin N Am. 1994; 5(1):1-18.

13. Reyes MG, Homsi MF, McDonald LW, Glick RP. Imprints, smears, and frozen sections of brain tumors. Neurosurgery. 1991; 29(4): 575-9.

14. Lechago J. The frozen section: pathology in the trenches. Arch Pathol Lab Med. 2005; 129(12):1529-31.

15. Gaudin PB, Sherman ME, Brat DJ, Zahurak M, Erozan YS. Accuracy of grading gliomas on CT-guided stereotactic biopsies: a survival analysis. Diagn Cytopathol. 1997; 17(6):461-6.

16. Takekawa Y, Kinukawa N, Nemoto N, Sakurai I, Komatsu K, Seki T, et al. [Usefulness of cytology applied simultaneously to frozen section at rapid intraoperative diagnosis of intracranial tumors]. Rinsho Byori. 1998; 46(9):954-8.

17. Chandrasoma PT, Smith MM, Apuzzo ML. Stereotactic biopsy in the diagnosis of brain masses: comparison of results of biopsy and resected surgical specimen. Neurosurgery. 1989; 24(2): 160-5.