

Original Article

Validation of whole slide imaging for frozen section diagnosis in surgical pathology

Thomas W. Bauer, Renee J. Slaw, Jesse K. McKenney, Deepa T. Patil

Department of Pathology, Robert J. Tomsich Pathology and Laboratory Medicine Institute, Cleveland Clinic, Cleveland, OH 44195, USA

E-mail: *Thomas W. Bauer - bauert@ccf.org

*Corresponding author

Received: 01 July 2015

Accepted: 03 August 2015

Published: 31 August 2015

Abstract

Background: Whole slide imaging (WSI) using high-resolution scanners is gaining acceptance as a platform for consultation as well as for frozen section (FS) evaluation in surgical pathology. We report results of an intra-observer concordance study comparing evaluation of WSI of scanned FS microscope slides with the original interpretation of the same microscope slides after an average lag time of approximately 1-year. **Methods:** A total of 70 FS cases (148 microscope slides) originally interpreted by 2 pathologists were scanned at $\times 20$ using Aperio CS2 scanner (Leica Biosystems, San Diego, CA, USA). Reports were redacted such that the study pathologists reviewed images using eSlide Manager Healthcare Network application (Leica Biosystems) accompanied by the same clinical information available at the time of original FS evaluation. Discrepancies between the original FS diagnosis and WSI diagnosis were categorized as major (impacted patient care) or minor (no impact on patient care). **Results:** Lymph nodes, margins for head and neck cancer resections, and arthroplasty specimens to exclude infection, were the most common FS specimens. The average wash-out interval was 380 days (range: 303–466 days). There was one major discrepancy (1.4% of 70 cases) where the original FS was interpreted as severe squamous dysplasia, and the WSI FS diagnosis was mild dysplasia. There were two minor discrepancies; one where the original FS was called focal moderate squamous dysplasia and WSI FS diagnosis was negative for dysplasia. The second case was an endometrial adenocarcinoma that was originally interpreted as Federation of Gynecology and Obstetrics (FIGO) Grade I, while the WSI FS diagnosis was FIGO Grade II. **Conclusions:** These findings validate and support the use of WSI to provide interpretation of FS in our network of affiliated hospitals and ambulatory surgery centers.

Key words: Digital pathology, frozen sections, validation, whole slide images

Access this article online

Website:

www.jpathinformatics.org

DOI: 10.4103/2153-3539.163988

Quick Response Code:



INTRODUCTION

There have been recent significant advances in the field of telepathology such that the resolution of whole slide imaging (WSI) is now comparable to light microscopic evaluation performed at different magnifications.^[1] A widely recognized potential benefit of using high-resolution scanners to capture WSIs of microscope slides is the ability to interpret those images

This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 3.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as the author is credited and the new creations are licensed under the identical terms.

For reprints contact: reprints@medknow.com

This article may be cited as:

Bauer TW, Slaw RJ, McKenney JK, Patil DT. Validation of whole slide imaging for frozen section diagnosis in surgical pathology. J Pathol Inform 2015;6:49.

Available FREE in open access from: <http://www.jpathinformatics.org/text.asp?2015/6/1/49/163988>

at remote locations. This may be especially useful for consultation cases or frozen sections (FSs) in surgical pathology. We previously validated the use of WSI for primary^[2] and consultation^[3] diagnoses, but now hope to apply the technology in a way that facilitates the interpretation of FSs from a network of regional affiliated hospitals and ambulatory surgery centers by subspecialty pathologists located in our main campus hospital and regional affiliates. Although a few centers have been using telepathology and digital imaging for interpreting remote FSs for years, relatively few reports have documented validation testing across a range of specimens encountered in a busy academic center.^[4-11]

In 2013, a workgroup of the college of american pathologists (CAP) published guidelines to help laboratories validate the use of digitized, WSIs for diagnostic use.^[12] Among other suggestions, the committee recommended that each pathology laboratory implementing WSI carry out its own validation studies that would test intra-observer variability between digital and glass slide interpretation. The committee further suggested that the validation study should closely emulate the real-world clinical environment in which the technology would be used and should include a sample set of at least 60 cases with a “wash-out” interval of at least 2 weeks. Revalidation should be performed whenever a significant change is made to any component of the WSI system, and if a new intended use for WSI is contemplated, a separate validation for the new use should be performed. The purpose of this report is to describe the results of a study intended to test the hypothesis that the interpretation of WSI of FSs is not inferior to interpreting the original FS microscope slides. The design of the study meets or exceeds all of the recommendations of the CAP working group.

METHODS

This is an intra-observer concordance study comparing the interpretation of WSI of scanned FS microscope slides with the original interpretation of the same microscope slides after an average lag time of approximately 1-year.

The Cleveland Clinic Department of Anatomic Pathology is a large, academic department composed of more than 50 staff pathologists as well as residents and fellows. In 2013, 4589 FS cases, comprising 9165 FS microscope slides were performed. Similar to most practices, the most frequent types of cases that we receive for intra-operative consultation include margins for oncologic resections, lymph nodes for cancer staging, lesions for diagnosis, and tissue for lymphoma work-up. In 2013, the subspecialties in which FSs were most frequently performed included head and neck, orthopedic, pulmonary, breast, gynecologic, genitourinary,

and neuropathology. Although all of the pathologists at the main hospital practice in a subspecialized sign-out model, the 59 who rotate on the FS interpretation service also have experience as general surgical pathologists and interpret most of the FS cases on any given day. When an unusually difficult FS is encountered, the FS pathologist may request a consultation from a different pathologist based on the subspecialty. FSs are performed in a dedicated area located close to the operating rooms, and the pathologist who is interpreting those slides has the ability to review the gross specimen and speak directly with the surgical team. Each specimen is accompanied by an accession sheet, which contains patient identification and limited clinical information.

Besides the main hospital, the Cleveland Clinic system includes approximately 12 additional regional affiliated hospitals and ambulatory surgery centers. Surgeons at some of those sites currently need to schedule FSs ahead of time to allow a pathologist time to travel to the site. Surgeons throughout the health care system have also expressed the desire to have subspecialty pathologists available for FS interpretation. As described by Evans *et al.*^[4] the use of telepathology or digital imaging has the potential to facilitate the remote interpretation of FSs by subspecialty pathologists at a main campus. However, even though we have already validated WSI for interpreting primary and consultation cases, it is still appropriate for us to test intra-observer concordance with respect to interpreting FSs.

Similar to a previous study,^[2] we hypothesized that interpreting WSI of FSs is not inferior to interpreting the original FS microscope slides, and it would be considered noninferior if there were no more than 4% major discrepancies between WSI diagnosis and the original FS diagnosis. No one knows the ideal yet realistically acceptable rate of intra-observer concordance, but as described in more detail in our previous study, we selected 4% based on previously reported discrepancy rates in a range of studies (both intra- and intra-observer), including comparing intra-operative FSs with digital images, FSs with final diagnoses, primary diagnoses with those rendered after consultation, and in CAP quality reviews. Although the CAP workgroup recommends a minimum wash-out interval of 2 weeks, in our experience pathologists can easily remember cases for longer than 2 weeks, so we selected a conservative wash-out interval of approximately 1-year. A study coordinator performed a search of our Laboratory Information System to identify consecutive FSs interpreted by each of two study pathologists (TWB and JM) as part of their regular FS rotation. Cases for which either of the two study pathologists had sought consultation from other subspecialty pathologists were excluded, as were cases for which a different pathologist on the FS rotation

had sought a consultative interpretation from either of the two study pathologists. The CAP working group recommended using a minimum of 60 cases,^[12] so we included the first 35 FS cases interpreted by each study pathologist to exceed that total number. As explained in one of our previous studies,^[2] many surgical pathology cases actually have multiple different slides corresponding to different “parts,” each of which may require a separate diagnosis. Among our 70 total FS cases, there were 148 microscope slide interpretations.

Each FS microscope slide was scanned at $\times 20$ using the same brand and model of scanner that we intend to use in the affiliated hospitals and surgical centers (Aperio CS2, Leica Biosystems, San Diego, CA, USA). Reports were redacted under the supervision of the study coordinator and uploaded such that the study pathologists would be able to see the clinical information that was available at the time of original FS interpretation, but would not be able to see the original interpretation or any subsequent diagnoses. Each study pathologist then reviewed the WSI using the same viewing application that we intend to use clinically (eeSlideManager Healthcare Network, Leica Biosystems), and entered the diagnosis in the “results” text box. After all cases had been reviewed, a spreadsheet was prepared by the study coordinator and the original FS diagnosis was compared with the diagnosis based on the WSI. Potentially discrepant cases were identified, and the director of the corresponding subspecialty was asked to classify the potential discrepancy as either concordant (the same or essentially the same diagnosis), minor discrepancy (a different diagnosis that would not influence patient care), or a major discrepancy (a different diagnosis that would potentially influence patient care). These discrepancy categories follow the Association of Directors in Anatomic and Surgical Pathology guidelines that currently exist in our routine FS quality assurance practice.^[13]

RESULTS

The types of cases and number of slides are shown in Table 1. Lymph nodes submitted to exclude metastatic carcinoma were the most frequent FS specimens (20 cases; 39 slides), and margins of resection associated with head-and-neck cases were the next most frequent type of frozen sections. The orthopedic surgeons at Cleveland Clinic also commonly request FSs at the time of revision or re-revision arthroplasty with a request to “rule out infection.” The other FSs represent a variety of surgical procedures commonly performed at tertiary hospitals. The average wash-out interval was 380 days (range: 303–466 days).

There was one major discrepancy: A margin on a head and neck resection for squamous cell carcinoma. The original

FS was interpreted as severe dysplasia while the WSI of that FS was interpreted as mild dysplasia [Figure 1]. This is known to be a difficult interpretation at FS, and evaluation of that microscope slide by a group of pathologists after completion of this study did not yield a consensus concerning the best diagnosis, in part because thermal artifact from the biopsy procedure has distorted a portion of the mucosa. Therefore, the major discrepancy rate was 1.4% of 70 cases, or 0.7% of 148 slides.

There were two minor discrepancies for a minor discrepancy rate of 2.8% of 70 cases or 1.4% of 148 slides. The first minor discrepancy was another margin from the same case described above in which the original FS was interpreted as focal moderate dysplasia while the WSI of that slide was interpreted as negative for dysplasia. The second minor discrepancy was an endometrial adenocarcinoma that was originally interpreted as Federation of Gynecology and Obstetrics (FIGO) Grade I while the WSI of that slide was interpreted as FIGO Grade II [Figure 2]. Interestingly, the final diagnosis based on interpretation of permanent sections was FIGO Grade II. Therefore, the diagnosis based on the WSI more closely matched the final diagnosis than did the diagnosis of the original FS microscope slide,

Table 1: Number of frozen section cases and slides of each case type. There were actually 70 cases, but several cases included frozen sections of different tissue types (e.g., bronchial margin and lymph node)

Case type	Number of cases	Number of slides
Lymph node	20	39
Head/neck	9	30
Ortho	9	21
Gyn	9	11
Lung	8	11
Parathyroid	3	10
Pancreas	5	9
GU	4	7
Soft tissue	5	6
Kidney	2	3
Stomach/esoph	1	1
Total	75	148

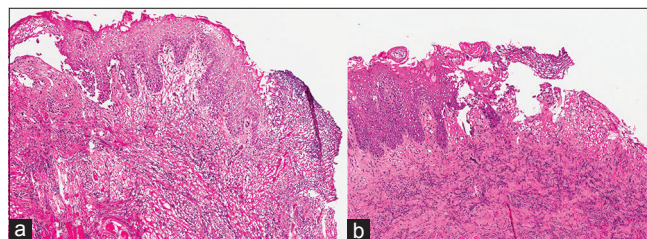


Figure 1: The original frozen section was originally interpreted as severe dysplasia, and the whole slide imaging of that frozen section was interpreted as mild dysplasia. (a). The corresponding permanent section (b) was interpreted as severe dysplasia

resulting in an “adjusted minor discrepancy rate”^[2] of 1.4% of cases, and 0.7% of slides.

CONCLUSIONS

The concept of visualizing and interpreting surgical pathology cases and FSs from remote locations has, over the past several decades encompassed a range of technological tools, including among other things: video telemicroscopy,^[14] transmission of selected static images,^[15] use of a robotic microscope,^[16] and high-resolution scanners with images viewed over the internet.^[4] Telepathology technology was first used for primary FS diagnosis in 1989 by investigators in Norway.^[14] These investigators documented a 100% accuracy rate, while providing intra-operative consultation coverage for 5 remote hospitals.

A recent consolidation of pathology services between the Cleveland Clinic system and approximately 12 additional regional affiliated hospitals and ambulatory surgery centers prompted us to expand our WSI-based system of providing surgical pathology consultations to incorporate WSI-based FS service. CAP panel recommendations for WSI require revalidation of the system when a new application is introduced. To this effect, we performed an intra-observer concordance study with respect to interpreting FSs. In a series of 70 FS cases, with 148 microscope slide interpretations, we found a diagnostic concordance rate of 98.6% when only major discrepancies were included (one major), and a concordance rate of 97.1% (one major and one minor) when both major and minor discrepancies were included.

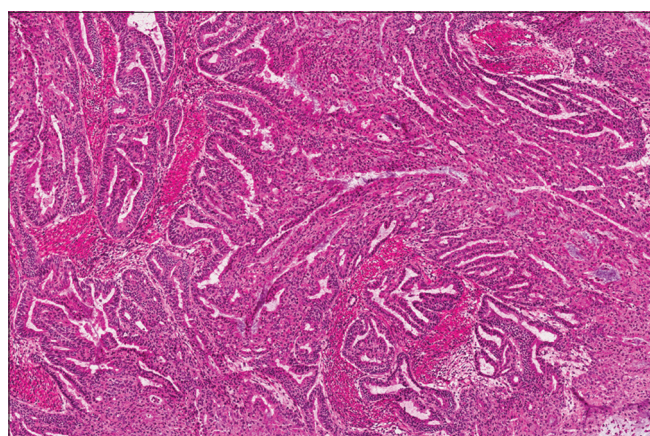


Figure 2: This endometrial adenocarcinoma was graded as Federation of Gynecology and Obstetrics Grade I on the original frozen section, but Federation of Gynecology and Obstetrics Grade II on the whole slide imaging, therefore representing a minor discrepancy. The permanent sections for the case were ultimately considered Federation of Gynecology and Obstetrics Grade II. Use of annotation tools in the viewing application for the whole slide imaging makes it easier to measure the depth of invasion than estimation via light microscopy

In one of the first published studies intended to test the use of telepathology to interpret FSs, Kaplan *et al.*^[16] measured the intra-observer variability of 120 cases for which both frozen and permanent sections had already been interpreted. A robotic telepathology system was installed at the Walter Reed Army Medical Center and several other affiliated institutions. Slides that had been used for the original FS diagnosis were viewed via telepathology, and later by the same pathologists using conventional light microscopy. The authors reported 100% intra-observer concordance and suggested that the use of telepathology for FS diagnosis “meets community standards of diagnostic accuracy” and would help optimize the number and distribution of pathologists to support surgical services.

Slodkowska *et al.*^[11] described their experience with telepathology systems in which two different digital imaging systems (“Aperio” and “Coolscope”) were used to interpret pulmonary pathology cases among cities in Poland. Part of the manuscript relates to cytology and small biopsies, but the authors also reported 114 cases in which FSs were obtained of lung tumors, bronchial margins, lymph nodes, and other tissues. Those sections were scanned and interpreted remotely. The authors reported that: “For the FSs, the primary telediagnoses were concordant with the light microscopy paraffin section diagnosed in 100% for Aperio and in 97.5% for Coolscope.”

Fallon *et al.*^[7] compared the original FS diagnosis with subsequent review of digital images of those FS slides performed by two different pathologists for 52 cases (72 slides) of ovarian tumors. Each of the two pathologists was concordant with respect to malignant, borderline, or uncertain in 50 of the 52 cases (96%). Each under-diagnosed 2 borderline malignant cases as benign cysts for a discrepancy rate of 4%. They concluded that correlation between the original FS and WSI diagnoses was very high, and the discordant cases represent a common differential diagnostic problem in ovarian pathology.

Evans *et al.* have described perhaps the largest published experience to date with the use of digital imaging to interpret FSs.^[4,5] They reported evolving from a robotic microscopic telepathology system to digital scanning system to interpret FSs remotely within the University Health Network in Toronto, Canada. As of 2009, they had used these systems to interpret FSs of primarily neuropathology cases, including 633 cases in which a digital scanner had been used. They reported a diagnostic accuracy of 97.5% with the robotic microscope system, and 98% with the digital scanning system. Most of the discrepancies were considered minor, most commonly misclassification of benign tumors. Repeat scanning was requested in 12 cases (2%), most commonly related to

poor quality initial FS microscope slides.

Ramey *et al.* recently tested the ability of 6 pathologists to interpret scanned images of FSs with the use of the iPad.^[10] Approximately 210 FS slides from 67 nearly consecutive cases were scanned at $\times 20$ and linked to the clinical information available at the time the FS slides were originally interpreted. Six pathologists used a wireless network and other applications to interpret between 4 and 9 cases each. The authors reported a diagnostic accuracy of 89% for interpreting FS using the iPad. Most of the discrepancies were minor; there were 3% clinically significant discrepancies (two from the same difficult head-and-neck case).

In several studies, Horbinski *et al.*^[9,17,18] described the use of a telepathology system to interpret FSs from an affiliated hospital in Pittsburgh. Diagnostic discrepancy rates ranged from 2.4 to 5.8% over several years, and most of the discrepancies were minor and related to tumor classifications. As summarized by Evans *et al.*,^[4] the experiences in Toronto with neuropathology FSs as well as the discrepancy rates reported by Hutarew *et al.*,^[19] and Horbinski *et al.*^[9] are not significantly different than the discrepancy rate of 2.7% reported by Plesec and Prayson^[20] for more than 2100 FSs of neoplastic neuropathology lesions.

FS discrepancies using WSI system can be attributed to several factors, including selection of appropriate tissue for sectioning (grossing related sampling error), sampling error due to inadequate sectioning of the tissue (not facing the block completely), poor image quality related to FS artifacts (folded sections, thick sections, excessive mounting medium), multiple tissue fragments on one slide, and inherent diagnostic error related to pathologist's experience with the system. Our diagnostic concordance rate was found to be similar to that described in the aforementioned studies. While the major discrepancy in our study was interpretive in nature, the tissue submitted for frozen had significant thermal artifact, and thus, the error could be attributed to the nature of sample rather than the technology used for evaluating the tissue. Furthermore, our major and minor discrepancies pertaining to the extent of dysplasia of squamous mucosa reflect a relatively common problem in diagnostic pathology.

Although our study had very few major discrepancies, each of the two reviewers found that it took longer to review a WSI of a FS compared to the microscope slide itself, although our efficiency may improve with additional experience and ultimately the use of a different interface than a mouse to navigate the images. We found screening lymph nodes in search of metastatic carcinoma and viewing the peri-prosthetic fibrous tissue to rule out acute inflammation to be especially time-consuming. On the other hand, the use of annotation tools to measure distances made it easier and faster to judge the depth of

invasion (for example, of endometrial adenocarcinoma) than the use of a microscope alone. Visualizing inked margins of resection on WSI was not difficult. Saving time by directing dissection and viewing the FSs from the pathologist's office rather than walking to and from the FS area near the operating room (often in another building) offers a way that digital pathology might balance the extra time needed for interpretation.

The strengths of our study include fulfilling or exceeding the recommendations of the CAP working group, including the use of a consecutive series of FSs with an average wash-out interval of just over 1-year. Limitations include, by chance, the lack of breast cancer cases although FSs of the breast are now rarely performed in our hospital. This study also lacked lymphoproliferative lesions, although at our institution the main role of FSs of lymphoproliferative lesions is to document the adequacy of tissue rather than to provide a diagnosis. No neuropathology cases were included in this study because the vast majority of neuropathology FSs at Cleveland Clinic are evaluated by neuropathology subspecialists. However, previous studies have reported very low discrepancy rates for the interpretation of neuropathology cases by telepathology or WSI.^[4,9,17,18] Our study was not designed to evaluate the preanalytic components of WSI-based intra-operative consultation, such as grossing errors, tissue-sectioning errors, and hardware and software related issues. We recognize that participation of the pathologist in the dissection and sampling of a specimen at the time of intra-operative consultation is important, and as we implement the technology we have installed web-based cameras that allow the remote pathologist to visualize and guide specimen orientation and sampling performed at the remote site by a physician's assistant. We hope to share this experience when we implement this technology in the coming months. In spite of those limitations, the results of our study are encouraging and support the use of WSI to provide interpretation of FSs in our network of affiliated hospitals and ambulatory surgery centers.

Financial Support and Sponsorship

Nil.

Conflicts of Interest

Thomas W. Bauer: Medical/Scientific Advisory Boards of Leica Biosystems and Xifin, Inc. Renee J. Slaw: Leica BioSystems Imaging Advisory Board.

REFERENCES

1. Wilbur DC, Madi K, Colvin RB, Duncan LM, Faquin WC, Ferry JA, *et al.* Whole-slide imaging digital pathology as a platform for teleconsultation: A pilot study using paired subspecialist correlations. *Arch Pathol Lab Med* 2009;133:1949-53.
2. Bauer TW, Schoenfeld L, Slaw RJ, Yerian L, Sun Z, Henricks WH. Validation of whole slide imaging for primary diagnosis in surgical pathology. *Arch Pathol Lab Med* 2013;137:518-24.
3. Bauer TW, Slaw RJ. Validating whole-slide imaging for consultation diagnoses

- in surgical pathology. *Arch Pathol Lab Med* 2014;138:1459-65.
4. Evans AJ, Chetty R, Clarke BA, Croul S, Ghazarian DM, Kiehl TR, et al. Primary frozen section diagnosis by robotic microscopy and virtual slide telepathology: The university health network experience. *Hum Pathol* 2009;40:1070-81.
 5. Evans AJ, Kiehl TR, Croul S. Frequently asked questions concerning the use of whole-slide imaging telepathology for neuropathology frozen sections. *Semin Diagn Pathol* 2010;27:160-6.
 6. Fallon MA, Wilbur DC, Prasad M. Ovarian frozen section diagnosis: Use of whole-slide imaging shows excellent correlation between virtual slide and original interpretations in a large series of cases. *Arch Pathol Lab Med* 2010;134:1020-3.
 7. Ferreiro JA, Gisvold JJ, Bostwick DG. Accuracy of frozen-section diagnosis of mammographically directed breast biopsies. Results of 1,490 consecutive cases. *Am J Surg Pathol* 1995;19:1267-71.
 8. Gould PV, Saikali S. A comparison of digitized frozen section and smear preparations for intraoperative neurotelepathology. *Anal Cell Pathol (Amst)* 2012;35:85-91.
 9. Horbinski C, Fine JL, Medina-Flores R, Yagi Y, Wiley CA. Telepathology for intraoperative neuropathologic consultations at an academic medical center: A 5-year report. *J Neuropathol Exp Neurol* 2007;66:750-9.
 10. Ramey J, Fung KM, Hassell LA. Use of mobile high-resolution device for remote frozen section evaluation of whole slide images. *J Pathol Inform* 2011;2:41.
 11. Slodkowska J, Pankowski J, Siemiatkowska K, Chyczewski L. Use of the virtual slide and the dynamic real-time telepathology systems for a consultation and the frozen section intra-operative diagnosis in thoracic/pulmonary pathology. *Folia Histochem Cytobiol* 2009;47:679-84.
 12. Pantanowitz L, Sinard JH, Henricks WH, Fatheree LA, Carter AB, Contis L, et al. Validating whole slide imaging for diagnostic purposes in pathology: Guideline from the College of American Pathologists Pathology and Laboratory Quality Center. *Arch Pathol Lab Med* 2013;137:1710-22.
 13. Recommendations on quality control and quality assurance in anatomic pathology. Association of Directors of Anatomic and Surgical Pathology. *Am J Surg Pathol* 1991;15:1007-9.
 14. Nordrum I, Eide TJ. Remote frozen section service in Norway. *Arch Anat Cytol Pathol* 1995;43:253-6.
 15. Frierson HF Jr, Galgano MT. Frozen-section diagnosis by wireless telepathology and ultra portable computer: Use in pathology resident/faculty consultation. *Hum Pathol* 2007;38:1330-4.
 16. Kaplan KJ, Burgess JR, Sandberg GD, Myers CP, Bigott TR, Greenspan RB. Use of robotic telepathology for frozen-section diagnosis: A retrospective trial of a telepathology system for intraoperative consultation. *Mod Pathol* 2002;15:1197-204.
 17. Horbinski C, Hamilton RL. Application of telepathology for neuropathologic intraoperative consultations. *Brain Pathol* 2009;19:317-22.
 18. Horbinski C, Wiley CA. Comparison of telepathology systems in neuropathological intraoperative consultations. *Neuropathology* 2009;29:655-63.
 19. Hutarew G, Schlicker HU, Idriceanu C, Strasser F, Dietze O. Four years experience with teleneuropathology. *J Telemed Telecare* 2006;12:387-91.
 20. Plesec TP, Prayson RA. Frozen section discrepancy in the evaluation of central nervous system tumors. *Arch Pathol Lab Med* 2007;131:1532-40.