Reliability of diagnosis from Mohs slides: interpersonal and intrapersonal agreement on basal cell carcinoma presence and histological subtype

C. van Lee
E. Ip Vai Ching
K. Nasserinejad
H. Neumann
M. Bol
P. Dikrama
N. Kelleners-Smeets
S. Koljenović
K. Munte
V. Noordhoek Hegt
H. de Vijlder
T. Nijsten
R. van den Bos

Br J Dermatol 2016;175:549-54

ABSTRACT

Background: The success of Mohs micrographic surgery (MMS) depends partly on the correct diagnosis of slides.

Objectives: To determine reliability of diagnosis from MMS slides.

Methods: This was a prospective study evaluating the reliability of diagnosis from MMS slides of basal cell carcinoma (BCC) presence, BCC location on the slide and BCC subtype among six raters who independently assessed 50 MMS slides twice with a two-month interval. Slides were randomly selected whereby difficult to diagnose slides were oversampled. For each slide, a reference diagnosis was established by an expert panel. Cohen's kappa (K) was calculated to determine levels of agreement interpersonally (rater vs. reference diagnosis) and intrapersonally (rater at T1 vs. T2). Multivariable logistic regression was used to determine independent risk factors for slides with interpersonal discordant diagnosis. The variables studied were BCC presence, whether a slide was scored as easy or difficult to diagnose, review duration of the 50 slides, profession and years of experience in diagnosis from MMS slides.

Results: Interpersonal and intrapersonal agreement were substantial on BCC presence (K = 0.66 and 0.68) and moderate on BCC subtype (K = 0.45 and 0.55). Slides that were scored as difficult to diagnose were an independent risk factor for interpersonal discordant diagnosis on BCC presence (OR 3.5, 95% CI 1.8-6.8).

Conclusions: Reliability of diagnosis from MMS slides was substantial on BCC presence and moderate on BCC subtype. For slides that are scored difficult to diagnose, a second opinion is recommended to prevent misinterpretation and thereby recurrence of skin cancer.

INTRODUCTION

Mohs micrographic surgery (MMS) is a preferred treatment for aggressive or recurrent facial basal cell carcinoma (BCC), due to the low recurrence rate and preservation of healthy tissue.[1,2] With MMS all surgical margins are microscopically viewed intraoperatively, whereas with conventional excision only a small percentage of the margins are viewed at all, and only postoperatively.

To warrant the additional costs of MMS compared to conventional excision, optimization of the procedure is pivotal. Because clinical recurrences after MMS often become apparent after five to ten years, a critical appraisal of the procedure is needed before MMS is more widely introduced in Europe. Although the success of MMS depends mainly on the correct diagnosis of slides, the reliability of diagnosis from MMS slides is poorly documented.[3,4] A better insight into the reliability of diagnosis from MMS slides might decrease slide misinterpretation. A decrease in slide misinterpretation prevents excision of healthy tissue and results in fewer skin cancer recurrences and, which are both the most important characteristics of MMS.

This is a reliability study on diagnosis from MMS slides. We determined interpersonal and intrapersonal levels of agreement on BCC presence, BCC location on the slide and BCC subtype. Furthermore, to explore possible approaches to improve the reliability of diagnosis from MMS slides, we identified risk factors for slides with interpersonal discordant diagnosis on BCC presence.

METHODS

This prospective reliability study determined interpersonal and intrapersonal levels of agreement on the diagnosis of 50 MMS slides among six raters and an expert panel.

Selection of slides

Slides were obtained from the MMS registry of the Erasmus University Medical Center, Rotterdam, the Netherlands. Frozen MMS slides were stained with haematoxylin and eosin; the distance between each section was 100 µm and the sections were 8 µm thick. Slides of bad quality were excluded.

In total, we selected 300 MMS slides derived from 50 different MMS procedures, all indicated for BCC: six slides per MMS procedure, subsequently cut from one tissue block. In some cases, BCC was present in the deepest cut, while BCC was absent in the outermost

slide. Therefore, one of the six slides was non-randomly selected, marked and used for the final diagnosis (n = 50). Raters were allowed to review all six slides, although they were instructed to diagnose the indicated slide only. All 300 slides were anonymized and coded with a specific study number. Comparable to our daily practice, half of the slides were tumour-free and half of the slides contained BCC. Slides with BCC contained variable BCC subtypes.

Slides were non-randomly selected by the researchers with oversampling of difficult to diagnose slides. We wanted to focus on the histological pitfalls in MMS such as small BCC strands that are easily missed, benign structures mimicking BCC and dense inflammation.[5] We expected that approximately half of all selected slides were difficult to diagnose due to histological pitfalls, which differs from our daily MMS practice where we score approximately five percent of the slides as difficult to diagnose. We chose to oversample difficult slides to determine the minimum level of agreement on diagnosis from MMS slides.

Selection of raters

Six raters (three MMS surgeons and three pathologists) were selected from three different centres: Erasmus University Medical Center Rotterdam (raters A, D), Isala Hospital Zwolle (raters B, E) and Maastricht University Medical Center (raters C, F). Raters had four to fifteen years of experience in diagnosis from MMS slides.

Rating process

Raters were instructed to review all 300 slides independently, twice, with two month intervals. Raters reviewed the slides without any clinical information (e.g. BCC location, BCC subtype, preoperative treatment). They were blinded to their own previous interpretation and to each other's interpretations. Before the second review, the slides were shuffled and recoded to prevent identification. Diagnoses were recorded on standardized study forms including: BCC presence (yes or no); BCC subtype with or without perineural invasion (i.e. superficial, nodular, micronodular, morpheaform or any combination); and exact BCC location. Raters marked the exact BCC location with a waterproof pencil on the slide. One of the researchers copied the marks to translucent paper and then removed the marks from the slides. Furthermore, raters scored the slides easy or difficult to diagnose, the time it took to review all slides, their years of experience in diagnosis from MMS slides, their profession and whether they worked in an academic or non-academic setting. At last, to verify if we indeed oversampled difficult to diagnose slides, raters were asked what percentage of difficult slides they find in their daily MMC practice.

Expert panel consensus-derived reference diagnosis

In the absence of a gold standard, an expert panel consensus-derived reference diagnosis was established for each slide. The expert panel consisted of one MMS surgeon and one pathologist, both with over 15 years of experience in diagnosis from MMS slides. Similar to the rating process, both experts reviewed all 300 slides independently, twice, with 2-month intervals. They were blinded to their own previous interpretation and to each other's interpretation. To establish the reference diagnosis, slides with experts' interpersonal or intrapersonal discordant diagnosis were resolved with consensus discussion. At this consensus discussion, experts were blinded to their own and each other's previous interpretations. If the experts did not reach consensus on BCC presence, deeper paraffin slides of the corresponding tissue blocks were made to reach a consensus diagnosis for all slides. The reference diagnosis included BCC presence (yes or no); BCC subtype with or without perineural invasion (i.e. superficial, nodular, micronodular, morpheaform or any combination); and BCC location on the slide.

Outcome 1: Interpersonal and intrapersonal agreement

Interpersonal agreement was determined between each rater vs. the reference diagnosis. Intrapersonal agreement was determined within each rater at T1 vs. T2. To evaluate the reliability of the reference diagnosis, we also determined the interpersonal and intrapersonal agreement of the two experts. Agreement concerned BCC presence, BCC subtype and BCC location on the slide. To indicate if BCC location was equally marked on a slide, the translucent papers of corresponding slides were stacked.

Outcome 2: Risk factors for slides with interpersonal discordant diagnosis

We determined independent risk factors for slides with discordant diagnosis between raters and the reference diagnosis at T1. Discordancy was based on BCC presence (yes or no), plus equal BCC location on the slide. The variables studied were: BCC presence [1. absent; 2. nonaggressive (i.e. superficial or nodular); 3. aggressive (i.e. micronodular, morpheaform, perineural invasion)]; whether raters scored slides as either easy or difficult to diagnose; the time it took each rater to review all slides; raters' years of experience in diagnosis from MMS slides; and profession (MMS surgeon or pathologist).

Statistics

For BCC presence and subtype, Cohen's kappa coefficients (K) were calculated to determine interpersonal and intrapersonal levels of agreement. Additionally, the t-test with Bonferroni correction was used to calculate whether the levels of agreement differed significantly between the raters. Overall mean K values were computed inside the programme. As described by Malpica et al., levels of agreement were interpreted as slight (K = 0-0.2), fair (K = 0.2-0.4), moderate (K = 0.4-0.6), substantial (K = 0.6-0.8) or almost

perfect (K = 0.8-1.0).[6] Regarding the sample size, to detect a K value of at least 0.6, we needed 50 cases and six raters for a power of 0.9 with alpha = 0.05. K could not be used for the agreement on BCC location. For BCC location we calculated percentages of equally marked BCC out of the total number of slides with concordant BCC presence. Univariable and multivariable logistic regression models were used to determine risk factors for slides with interpersonal discordant diagnosis on BCC presence. All statistical analyses were conducted in the Bayesian framework. Statistical analyses were performed using software package R version3.1.1 and JARG version 3.4.0. Guidelines for reporting reliability and agreement studies were used to report this study.[7]

RESULTS

In total, 800 slide assessments were analysed: six raters and two panel experts diagnosed 50 slides twice.

Reference diagnosis

The experts' diagnosis from MMS slides on BCC presence differed interpersonally at T1 or T2 in 23% (23/100). For expert 1, diagnosis on BCC presence differed intrapersonally in 8% (4/50). For expert 2, diagnosis on BCC presence differed intrapersonally 10% (5/50). At the consensus meeting, the experts did not reach consensus on BCC presence 10% (5/50). Of these five cases, deeper paraffin slides of the corresponding tissue blocks were made. With these deeper slides, the experts reached consensus in all cases. The reference diagnoses of the 50 MMS slides were 'tumour-free' in 26 slides and 'BCC present' in 24 slides. The reference diagnoses on BCC subtypes are summarized in Table 1.

Table 1. Reference diagnosis of the 50 Mohs micrographic surgery slides.

Variable	Cases n = 50 (%)
Tumour free	26 (52)
BCC present[a]	24 (48)
Superficial	1 (2)
Nodular	8 (16)
Morpheaform	10 (20)
Micronodular	3 (6)
Perineural invasion	2 (4)

Percentage were rounded.
BCC, basal cell carcinoma; n, number.
[a] In the case of mixed BCC subtypes, the most aggressive subtype was recorded, i.e. superficial < nodular < micronodular < morpheaform < perineural invasion.

Description of diagnosis from MMS slides at T1

At the first rating session (T1), raters needed 80-140 minutes to diagnose the 50 MMS slides (Table 2). Altogether, the six raters scored 21% (64/300) diagnoses as difficult. In their daily MMS practice, the experts and raters score approximately 6% of cases difficult to diagnose (range 2-10%), which confirms that difficult to diagnose slides were oversampled in this study. Raters diagnosed 'BCC presence' equally to the reference diagnosis in 17-24 slides out of 24 (mean 85%, range 67-100%). Raters diagnosed 'tumour-free' equally to the reference diagnosis in 14-26 slides out of 26 (mean 82%, range 58-100%).

Table 2. Description of the six raters and their diagnosis from Mohs micrographic surgery slides at T1.

Rater	Profession	Academic or Nonacademic	Experience in diagnosis from MMS slides (years)	Duration to diagnose 50 MMS slides (minutes)	Difficult to diagnose MMS slides out of 50 (%)	BCC present n = 24[a] (%)	Tumour free n = 26[b] (%)
A	Dermatologist	Academic	5	120	7 (14)	22 (92)	16 (62)
B	Dermatologist	Nonacademic	7	140	10 (20)	23 (96)	15 (58)
C	Dermatologist	Academic	15	105	17 (34)	21 (88)	22 (85)
D	Pathologist	Academic	8	80	8 (16)	16 (67)	26 100)
E	Pathologist	Nonacademic	4	90	10 (20)	24 (100)	23 (89)
F	Pathologist	Academic	8	100	12 (24)	16 (67)	26 (100)

Percentage were rounded.

BCC, basal cell carcinoma; MMS, Mohs micrographic surgery; n, number.

[a] Numbers of slides that were equally diagnosed as the reference diagnosis of 'BCC present' out of 24.

[b] Numbers of slides that were equally diagnosed as the reference diagnosis of 'tumour free' out of 26.

Experts' interpersonal and intrapersonal agreement

Interpersonal agreement between the two experts was substantial on BCC presence (K = 0.61, 95% CI 0.41-0.79) and moderate on BCC subtype (K = 0.45, 95% CI 0.31-0.61) (Table 3). Equal to the interpersonal agreement, the overall intrapersonal agreement within the two experts was substantial on BCC presence (K = 0.75, 95% CI 0.60-0.86) and moderate on BCC subtype (K = 0.58, 95% CI 0.47-0.69)

Interpersonal levels of agreement between each rater and the reference diagnosis

Overall, interpersonal agreement and the reference diagnosis was substantial on BCC presence (K = 0.66, 95% CI 0.58-0.73) and moderate on BCC subtype (K = 0.45, 95% CI 0.39-0.52) (Table 4). Overall, BCC location on the slide was equal to the reference diagnosis in 95% (range 88-100%). Levels of agreement on BCC presence differed significantly only between raters B and E. Rater B had a significantly lower interpersonal level of agreement (K = 0.49, 95% CI 0.27-0.69), while rater E had a higher interpersonal level of agreement (K = 0.89, 95% CI 0.72-0.97), p = 0.01. Levels of agreement on BCC subtype did not differ significantly between the raters.

Table 3. Interpersonal agreement between the experts and intrapersonal agreement within the experts in 50 Mohs micrographic surgery slides.

Agreement	BCC presence K (95% CI)	BCC location on MMS slide equal (%)	BCC subtype[a] K (95% CI)
Interpersonal	0.61 (0.41-0.79)	100	0.45 (0.31-0.61)
Intrapersonal			
Expert 1	0.68 (0.44-0.85)	96	0.64 (0.47-0.79)
Expert 2	0.84 (0.64-0.95)	88	0.52 (0.37-0.68)
Overall	0.75 (0.60-0.86)	92	0.58 (0.47-0.69)

Percentage were rounded.
BCC, basal cell carcinoma; CI, confidence interval; K, Cohen's kappa; MMS, Mohs micrographic surgery.
[a] BCC subtypes included: superficial, nodular, micronodular, morpheaform, perineural invasion or any combination.

Table 4. Interpersonal agreement between each of the six raters and the reference diagnosis and intrapersonal agreement within each rater in 50 Mohs micrographic surgery slides.

Agreement	BCC presence K (95% CI)	BCC location on slide equal %	BCC subtype[a] K (95% CI)
Interpersonal			
Overall	0.66 (0.58-0.73)	95	0.45 (0.39-0.52)
A	0.53 (0.29-0.72)	95	0.41 (0.25-0.57)
B	0.49 (0.27-0.69)	96	0.35 (0.21-0.51)
C	0.73 (0.50-0.88)	100	0.51 (0.35-0.66)
D	0.67 (0.46-0.84)	88	0.47 (0.32-0.64)
E	0.89 (0.72-0.97)	92	0.55 (0.41-0.70)
F	0.67 (0.46-0.84)	100	0.42 (0.28-0.58)
Intrapersonal			
Overall	0.68 (0.59-0.76)	86	0.55 (0.49-0.62)
A	0.66 (0.42-0.84)	89	0.67 (0.50-0.81)
B	0.57 (0.33-0.78)	85	0.51 (0.35-0.66)
C	0.64 (0.42-0.82)	89	0.51 (0.35-0.66)
D	0.69 (0.44-0.86)	77	0.58 (0.39-0.75)
E	0.75 (0.55-0.90)	89	0.59 (0.44-0.73)
F	0.79 (0.59-0.92)	88	0.46 (0.32-0.61)

Percentage were rounded.
BCC, basal cell carcinoma; CI, confidence interval. K, Cohen's kappa.
[a] BCC subtypes included: superficial, nodular, micronodular, morpheaform, perineural invasion or any combination.

Intrapersonal levels of agreement

Equal to the interpersonal agreement, overall intrapersonal agreement within each rater was substantial on BCC presence (K = 0.68, 95% CI 0.59-0.76) and moderate on BCC subtype (K = 0.55, 95% CI 0.49-0.62) (Table 4). Overall, BCC location similarly scored more

often interpersonally (95%) than intrapersonally (86%). Intrapersonal levels of agreement did not differ significantly between the raters on BCC presence and BCC subtype.

Risk factors for slides with discordant diagnosis on basal cell carcinoma presence

Raters' diagnosis on BCC presence and BCC location differed from the reference diagnosis in 17% (50/300). An independent risk factor for discordant diagnosis on BCC presence were slides that were scored as difficult to diagnose by the raters (OR 3.5, 95% CI 1.8-6.8) (Table 5). Other studied variables did not affect the risk in the univariable or multivariable analysis.

Table 5. Univariable and multivariable analysis of risk factors for slides with discordant diagnosis on BCC presence between raters and the reference diagnosis.

Variable	Univariable OR (95% CI)	Multivariable OR (95% CI)
MMS slides (n = 50)		
Tumour free (n = 26)	1.00	1.00
Nonaggresive BCC subtype (n = 9)[a]	1.17 (0.54-2.46)	1.38 (0.61-3.02)
Aggressive BCC subtype (n = 15)[b]	0.96 (0.50-1.84)	1.11 (0.54-2.24)
Raters profession (n = 6)		
Pathologist (n = 3)	1.00	1.00
Dermatologist (n = 3)	1.52 (0.51-4.27)	0.88 (0.06-6.60)
Difficulty of the interpretation (n = 300)		
Easy (n = 236)	1.00	1.00
Difficult (n = 64)	3.30 (1.71-6.26)	3.54 (1.81-6.84)
MMS experience (years)	1.00 (0.84-1.21)	0.99 (0.76-1.34)
Review duration (minutes)	1.01 (0.99-1.04)	1.02 (0.96-1.09)

BCC, basal cell carcinoma; CI, confidence interval; MMS, Mohs micrographic surgery; OR, odds ratio.

[a] Nonaggressive BCC subtypes included: superficial BCC and nodular BCC.

[b] Aggressive BCC subtypes included: morpheaform BCC, micronodular BCC and BCC with perineural invasion.

DISCUSSION

Reliability of interpretation of MMS slides is pivotal because the success of MMS depends mainly on the correct diagnosis of slides. This study showed substantial interpersonal and intrapersonal agreement on whether BCC was present and moderate agreement on BCC subtype, which is comparable with other fields of diagnostic reliability such as breast pathology and radiology.[8-16] Discordant diagnosis on BCC presence was more frequent when slides were self-scored as difficult to diagnose. This suggests that raters are aware of their uncertainty and should know when to consult others to reduce misin-

terpretation. Although interpersonal and intrapersonal agreement on BCC presence was found to be imperfect, the recurrence rate of skin cancer after MMS is extremely low.[1,2]

Previous studies show high interpersonal agreement between MMS surgeons and pathologists (95-99%) on BCC presence.[17-22] These studies overestimate the reliability of diagnosis from MMS slides because they report concordance rates instead of K values. In contrast to these studies, we oversampled difficult MMS slides, which resulted in an underestimated level of agreement. Besides our focus on challenging aspects of diagnosis from MMS slides, this study had several other unique aspects. Firstly, we established a reference diagnosis and we included six raters to determine interpersonal and intrapersonal agreement. Secondly, agreement on BCC presence, BCC subtype and BCC location on the slide was determined. Thirdly, we determined risk factors for interpersonal discordance.

Remarkably, interpersonal and intrapersonal agreement on BCC presence was comparable, while we expected to find a higher intrapersonal agreement. Intrapersonal diagnosis from MMS slides was found to be less consistent than we anticipated even among experienced raters. This might be because we oversampled slides that were difficult to diagnose.

In line with Nedved et al., we observed a somewhat lower level of agreement on BCC subtyping than on BCC presence.[23] This is logically explained because BCC presence (yes or no) is a binary question while the differentiation between BCC subtypes is less strict and mixed subtypes are common. To improve the diagnostic concordance in BCC subtyping, it might be necessary to further specify and simplify the current World Health Organization classification of BCC subtypes.

Our predictor analysis of discordant BCC diagnosis from MMS slides showed that slides that were self-scored as difficult to diagnose increased the likelihood of discordance three-and-half-fold (OR 3.5, 95% CI 1.8-6.8). As suggested in a previous study, a second opinion might prevent slide misinterpretation and thereby prevent skin cancer recurrence and unnecessary excision of healthy tissue.[24] Other possible measures to improve accurate slide diagnosis include cutting an additional deeper slide or obtaining a (paraffin) slide with additional histochemical stains.[5,25] In exceptionally difficult cases, these actions do not clear the diagnosis. In those cases, to minimize the risk of recurrence an additional small safety margin should be excised and examined microscopically.

A promising development that might further optimize diagnosis from MMS slides is the use of optical devices such as spectroscopy, which aim to eliminate the subjectivity of human diagnosis from MMS slides.[26]

Other studies show that MMS surgeons are as good as pathologists in evaluating MMS slides.[18,21,22] Although this study was not designed to assess differences between MMS surgeons and pathologists, the three MMS surgeons identified almost all 24 BCC-positive slides correctly (high sensitivity), while they were more likely to interpret benign structures as BCC (lower specificity). In contrast, the three pathologists had a lower sensitivity, but a higher specificity.

This study has several limitations. Crucially, there is no gold standard.[27] We assumed that the expert reference diagnosis was more accurate than the raters' diagnoses. Moreover, our study showed that interpersonal levels of agreement were equal between the experts and raters and intrapersonal levels of agreement were only a little higher among the experts than among the raters. Furthermore, the determined reliability is likely to be lower than the reality for the following reasons.[25,27] Firstly, difficult to diagnose slides were oversampled. Secondly, raters could not consult a colleague or pathologist to establish a consensus-derived diagnosis. Thirdly, raters could not ask for a deeper (paraffin) slide or additional staining. Fourthly, raters were not informed about the clinical context. In addition, the generalizability of our findings is limited because the number of participating clinicians was small (n = 6), although we included a heterogeneous group of raters (i.e. MMS surgeons and pathologists of academic and non-academic MMS settings). To further determine the reliability of interpretation of MMS slides, a large international web-based study assessing randomly selected MMS slides is warranted.

This study shows that interpersonal and intrapersonal levels of agreement on diagnosis from MMS slides were substantial for BCC presence and moderate for BCC subtype. Slides that were scored as difficult to diagnose were an independent risk factor for discordant diagnosis. A better understanding of the reliability of diagnosis from MMS slides might decrease slide misinterpretation and thereby prevent recurrences of skin cancer.

Acknowledgement

We would like to thank Loes Zandwijk-Hollestein for her statistical advice.

REFERENCES

1. van Loo E, Mosterd K, Krekels GA et al. Surgical excision versus Mohs' micrographic surgery for basal cell carcinoma of the face: a randomised clinical trial with 10 year follow-up. Eur J Cancer 2014;50:3011-20.
2. Muller FM, Dawe RS, Moseley H et al. Randomized comparison of Mohs micrographic surgery and surgical excision for small nodular basal cell carcinoma: tissue-sparing outcome. Dermatol Surg 2009;35:1349-54.
3. Hruza GJ. Mohs micrographic surgery local recurrences. J Dermatol Surg Oncol 1994;20:573-7.
4. Campbell T, Armstrong AW, Schupp CW et al. Surgeon error and slide quality during Mohs micrographic surgery: is there a relationship with tumour recurrence? J Am Acad Dermatol 2013;69:105-11.
5. Bouzari N, Olbricht S. Histologic pitfalls in the Mohs technique. Dermatol Clin 2011;29:261-72.
6. Malpica A, Matisic JP, Niekirk DV et al. Kappa statistics to measure interrater and intrarater agreement for 1790 cervical biopsy specimens among twelve pathologists: qualitative histopathologic analysis and methodologic issues. Gynecol Oncol 2005;99:S38-52.
7. Kottner J, Audig,e L, Brorson S et al. Guidelines for reporting reliability and agreement studies (GRRAS) were proposed. J Clin Epidemiol 2011;64:96-106.
8. Elmore JG, Feinstein AR. A bibliography of publications on observer variability (final installment). J Clin Epidemiol 1992;45:567-80.
9. Elmore JG, Longton GM, Carney PA et al. Diagnostic concordance among pathologists interpreting breast biopsy specimens. JAMA 2015;313:1122-32.
10. Pournik O, Alavian SM, Ghalichi L et al. Inter-observer and intra-observer agreement in pathological evaluation of non-alcoholic fatty liver disease suspected liver biopsies. Hepat Mon 2014;14:e15167.
11. Okamoto Y, Fujimori T, Ohkura C et al. Histological assessment of intra- and inter-institutional reliabilities in detection of desmoplastic reaction in biopsy specimens of early colorectal carcinomas. Pathol Int 2013;63:539-45.
12. Eaden J, Abrams K, McKay H et al. Gastrointestinal, inter-observer variation between general and specialist gastrointestinal pathologists when grading dysplasia in ulcerative colitis. J Pathol 2001;194:152-7.
13. Zannoni L, Savelli L, Jokubkiene L et al. Intra- and inter-observer agreement with regard to describing adnexal masses using International Ovarian Tumour Analysis terminology: a reproducibility study involving seven observers. Ultrasound Obstet Gynecol 2014;44:100-8.
14. Elmore JG, Wells CK, Lee CH et al. Variability in radiologists' interpretations of mammograms. N Engl J Med 1994; 331:1493-9.
15. Redondo A, Comas M, Macia F et al. Inter- and intraradiologist variability in the BI-RADS assessment and breast density categories for screening mammograms. Br J Radiol 2012;85:1465-70.
16. Bol MG, Baak JP, Buhr-Wildhagen S et al. Reproducibility and prognostic variability of grade and lamina propria invasion in stages Ta, T1 urothelial carcinoma of the bladder. J Urol 2003;169:1291-4.
17. Murphy ME, Brodland DG, Zitelli JA. Errors in the interpretation of Mohs histopathology sections over a 1-year fellowship. Dermatol Surg 2008;34:1637-41.
18. Tan E, Elliott T, Yu L et al. Mohs surgery histopathology concordance in Australia. Australas J Dermatol 2011;52:245-7.
19. Semkova K, Mallipeddi R, Robson A et al. Mohs micrographic surgery concordance between Mohs surgeons and dermatopathologists. Dermatol Surg 2013; 39:1648-53.

20. Grabski WJ, Salasche SJ, McCollough ML et al. Interpretation of Mohs micrographic frozen sections: a peer review comparison study. J Am Acad Dermatol 1989;20:670-4.
21. Mcfarlane L, Waters A, Evans A et al. Seven years' experience of Mohs micrographic surgery in a UK centre, and development of a UK minimum dataset and audit standards. Clin Exp Dermatol 2013;38:262-9.
22. Mariwalla K, Aasi SZ, Glusac EJ et al. Mohs micrographic surgery histopathology concordance. J Am Acad Dermatol 2009; 60:94-8.
23. Nedved D, Tonkovic-Capin V, Hunt E et al. Diagnostic concordance rates in the subtyping of basal cell carcinoma by different dermatopathologists. J Cutan Pathol 2014;41:9-13.
24. van Lee CB, Graafland B, Koljenovic S et al. Additional review of Mohs slides to optimize Mohs micrographic surgery. Br J Dermatol 2015;173:123-7.
25. Allison KH, Reisch LM, Carney PA et al. Understanding diagnostic variability in breast pathology: lessons learned from an expert consensus review panel. Histopathology 2014;65:240-51.
26. Takamori S, Kong K, Varma S et al. Optimization of multimodal spectral imaging for assessment of resection margins during Mohs micrographic surgery for basal cell carcinoma. Biomed Opt Express 2015;6:98-111.
27. Davidson NE, Rimm DL. Expertise vs evidence in assessment of breast biopsies, an atypical science. JAMA 2015; 313:1109-10.