

Interobserver variation is a significant limitation in the diagnosis of Burkitt lymphoma

Swapnil Ulhas Rane,
Tanuja Shet, Epari Sridhar,
Sanica Bhele¹, Vaishali
Gaikwad², Shubhangi Agale³,
Sweety Shinde⁴, Daksha Prabhat⁵,
Gwendolyn Fernandes⁵,
Meenal Hastak⁶, Chandralekha
Tampi⁶, Swati Narurkar⁷,
Keyuri Patel⁸, Chitra Madiwale⁹,
Ketki Shah¹⁰, Laxmi Shah¹⁰,
Satyakam Sawaimoon¹¹,
Purnima Lad¹²

Department of Pathology, Tata Memorial Centre, ¹Department of Pathology, Prince Aly Khan Hospital, Mumbai, ²Department of Pathology, L.T. Medical College and Sion Hospital, Mumbai, ³Department of Pathology, Grant Medical College and Sir J. J. Hospital, Mumbai, ⁴Department of Pathology, T.N. Medical College and Nair Hospital, Mumbai, ⁵Department of Pathology, Seth G.S. Medical College and K.E.M. Hospital, Mumbai, ⁶Department of Pathology, Lilavati Hospital, Mumbai, ⁷Dr. Narurkar's Laboratory, Mumbai, ⁸Department of Pathology, Bombay Hospital, Mumbai, ⁹Department of Pathology, Hinduja Hospital, Mumbai, ¹⁰Department of Pathology, Nanavati Hospital, Mumbai, ¹¹Tata Medical Centre, Kolkatta, ¹²Specialty Ranbaxy Ltd, Mumbai, India

Address for correspondence:

Tanuja Shet
Professor & Pathologist,
Department of Pathology,
8th Floor, Annexe Building, Tata
Memorial Hospital, Parel,
Mumbai - 400 092, India.
Email:tanujashet5@gmail.com

ABSTRACT

Context: The pathology of classic Burkitt lymphoma (BL) remains a challenge despite being a well-defined entity, in view of the significant overlap with atypical BL and B-cell lymphoma intermediate between DLBL (diffuse large B cell lymphoma) and BL. They are difficult to be segregated in resource-limited setups which lack molecular testing facilities. This is further affected by interobserver variability and experience of the reporting pathologist. **Aims:** The aim of our study was to quantitate variability among a group of pathologists with an interest in lymphomas (albeit with variable levels of experience) and quantitate the benefit of joint discussions as a tool to increase accuracy and reduce interobserver variability of pathologists, in the diagnosis of BL in a resource-limited setup. **Materials and Methods:** A set of 25 non-Hodgkin lymphoma cases in which a diagnosis of BL was entertained were circulated to 14 participating pathologist within the Mumbai lymphoma study group. A proforma recorded the morphologic and immunohistochemical features perceived during the initial independent diagnosis followed by a consensus meeting for discussion on morphology and additional information pertinent to the case. **Statistical analysis and Results:** The concordance was poor for independent diagnosis among all the pathologists with kappa statistics (\pm SE) of 0.168 (\pm 0.018). Expert lymphoma pathologists had the highest (albeit only fair) concordance (kappa = 0.373 \pm 0.071) and general pathologists the lowest concordance (kappa = 0.138 \pm 0.035). Concordance for morphological diagnosis was highest among expert lymphoma pathologists (kappa = 0.356 \pm 0.127). Revision of diagnoses after consensus meeting was highest for B-cell lymphoma intermediate between DLB and BL. To conclude, interobserver variation is a significant problem in BL in the post WHO 2008 classification era. Experience with a larger number of cases and joint discussion exercises such as the one we conducted are needed as they represent a simple and effective way of improving diagnostic accuracy of pathologists working in a resource-limited setup.

Key words: Burkitt lymphoma, concordance, inter-observer variation, non-Hodgkin lymphoma

Access this article online

Quick Response Code:



Website:
www.ijmpo.org

DOI:
10.4103/0971-5851.133721

INTRODUCTION

The diagnosis of Burkitt lymphoma is a challenge partly due to the dramatic impact on treatment and chance of cure and partly due to the significant morphologic overlap

it shares with diffuse large B-cell lymphoma (DLBL). The current WHO classification of hematolymphoid neoplasms 2008 acknowledges this difficulty with the inclusion of a diagnostic category of “B-cell lymphoma unclassified with features intermediate between Burkitt lymphoma and diffuse large B-cell lymphoma”. This term should however be used after the three diagnostic points, namely morphology, immunohistochemistry and cytogenetic studies fail to classify a tumor as Burkitt lymphoma. But in countries with limited resources, urgency for diagnosis results in therapy of patients based on morphology and immunohistochemistry alone. Classically, BL is a high-grade B-cell non-Hodgkin lymphoma composed of sheets of small non-cleaved cells with high proliferation index ($\geq 95\%$) and typically expresses BCL6, CD10 and is negative for BCL2. The recurrent t(8;14)(q24;q32) translocation involving the *IgH* and *C-MYC* genes is classical for BL though not absolutely specific. All of the above three can be seen in DLBL. The difficulty in diagnosis of BL is worsened by paucity of good immunohistochemistry and molecular genetics laboratories in resource-limited countries and lack of adequate training in these modalities for pathologists. The presence of focused study groups such as the Mumbai Lymphoma study group facilitates an interrogation of this variation and provides a platform for development of consensus among pathologists for diagnosis of BL in a resource-limited set up. The aim of our study was to evaluate the ability and interobserver variability of pathologists, with an interest in lymphomas but with varying levels of experience, to diagnose BL in a resource-limited set up.

MATERIALS AND METHODS

The study was conducted among the members of the Mumbai Lymphoma Study group starting from March 2011. The study involved initial independent assessment by the pathologists on a set of 25 cases and subsequently, slides were reviewed at the concluding meeting where a consensus diagnosis was reached after joint discussion.

Participating pathologists

Fourteen pathologists from 11 centers (tertiary hospitals, diagnostic centers and private hospitals) with varying levels of expertise with diagnosing lymphomas participated in the study. For the purpose of the study, the pathologists were divided into three groups, namely expert lymphoma pathologists (A1-A3) who worked in a diagnostic center with >500 lymphoma cases/year, pathologists with experience in lymphomas were specialist hematopathologists working in general hospitals with some training in lymphoma (B1-B4) and other pathologist involved in diagnostic surgical pathology (C1-C7).

Cases

A set of 25 cases were selected where in a diagnosis of BL was considered either based on clinical features, morphological features and immunophenotype. These cases were drawn from four centers (TMH, KEMH, BNH and TNMC), with the largest series being from TMH (18/25). The clinicopathological details along with the initial diagnoses and final consensus diagnoses of these cases are shown in Table 1.

Initial independent assessment

During the initial independent assessment, all 14 pathologists recorded their findings (morphological and immunohistochemistry) and final impression in one of the three diagnostic categories, namely BL/BL with atypical features, B-cell lymphoma unclassifiable-intermediate between BL/DLBL and DLBL. They were provided with basic clinical information like age, sex, clinical presentation and site of biopsy along with hematoxylin–eosin-stained slides and a limited immunohistochemistry panel comprising of CD3, CD20, CD10, BCL2, BCL6 and MIB1. The basic panel of immunohistochemistry used was CD20/CD3/CD10/MIB1/bcl2 in all cases. BCL6 was additionally provided in 8 cases and at consensus meeting, information on CMYC status was provided in 11 and EBER-ISH in 3 of the 25 cases. The pathologists were provided with a proforma whereby they were given clinical details of patients like site, age, sex, performance status, LDH and stage of disease and a table where they recorded their assessment of cell size, nuclear contour, degree of anisonucleosis, type of nucleoli, individual immunohistochemistry results and a final impression.

Since the cases were acquired from different institutes/laboratories with varying levels of technical expertise, the histopathology and immunohistochemistry sections were graded for their technical adequacy in terms of fixation, staining quality and overall acceptability on a semi-quantitative scale of 1-3 with 1 being least acceptable for reporting while 3 being ideal technical processing.

Consensus building

A consensus meeting was held after the initial independent assessment of all the 14 pathologists was recorded. At the consensus meeting the slides were discussed again with a discussion on the WHO 2008 criteria for diagnosis of BL, diagnostic dilemmas and pitfalls, in addition contributory clinical information like type of treatment received, response to therapy and information about results of EBER-ISH and *c-MYC* FISH were also given. Consensus was defined when $\geq 75\%$ of pathologists agreed on a single diagnosis.

Table 1: Clinico-pathological details including the original diagnosis at parent institute and the consensus diagnosis.

Case no	Age/Sex	Clinical Details	Original Diagnosis	Consensus Diagnosis
1	50/F	left inguinal node, LDH-742 IU/L, Stage IA	DLBL	DLBL
2	50/F	Tobacco chewer, progressively increasing right parotid region swelling since 2 months, Stage IIAX, LDH-172 IU/L	B-Int	No consensus
3	4/M	Left maxillary mass eroding alveolus, orbit and nose with involved bone marrow. LDH-1194 IU/L, stage IV BXE, ECOG 2	BL	BL
4	6/F	Abdominal pain and vomiting. CT scan reveals an abdominal mass extending from the subhepatic region to right iliac fossa. LDH-972 IU/L, ECOG-3	BL	BL
5	33/M	Anterior chest wall mass extending into mediastinum with SVC syndrome. LDH-1020 IU/L, ECOG-3	AtyBL	AtyBL
6	8/M	Ileocaecal mass with intussusception	AtyBL	AtyBL
7	11/F	Bowel mass with ovarian and omental deposits	AtyBL	AtyBL
8	13/M	B/l pleural effusion with ascites, inoperable retroperitoneal nodal mass	AtyBL	AtyBL
9	5/M	Multiple enlarged retroperitoneal lymph nodes	AtyBL	AtyBL
10	48/F	Right axillary node	DLBL	DLBL
11	21/M	Lower abdominal pain and malena. On USG, ileocaecal mass with intussusception	AtyBL	B-Int
12	30/F	12.9x4.4 cm left psoas mass, LDH-485 IU/L, ECOG-2.	DLBL	B-Int
13	24/M	Small intestinal mass	BL	AtyBL
14	6/M	Jejunal mass	BL	AtyBL
15	47/M	HIV positive, with enlarged right cervical node and bone marrow involved by same tumor	BL	BL
16	45/F	HIV positive, right axillary nodal mass along with, nasopharyngeal mass, bilateral level IIb and V nodes. LDH-413 IU/L.	DLBL	No consensus
17	56/F	Abdominal pain, vomiting, with a mass in second part of duodenum along with liver lesions. LDH-1028 IU/L.	B-Int	DLBL
18	54/M	HIV positive, with bilateral tonsillar fossa mass, extending into soft palate and inferiorly bilateral valleculae and lingual surface of epiglottis. LDH-249 IU/L.	DLBL	DLBL
19	-/M	HIV positive with right axillary node. Bone marrow involved by the same tumor. LDH-259 IU/L, ECOG-1	BL	AtyBL
20	43/M	Enlarged inguinal node	AtyBL	AtyBL
21	5/M	Ileoileal intussusception with mass. LDH-267 IU/L, ECOG-3.	BL	BL
22	22/F	Married lady with breast mass, multiple intraabdominal, renal, colonic and ovarian masses	BL	BL
23	16/-	HIV positive with enlarged supraclavicular node. LDH-477 IU/L, ECOG-1	DLBL	AtyBL
24	15/F	Nasopharyngeal mass with change in voice and dysphagia	B-Int	B-Int
25	4.5/M	Multiple masses in epigastrium, anterior to bladder and involving the kidneys and pancreas	BL	BL

M – Male, F – Female, BL – Classic Burkitt Lymphoma, AtyBL – Burkitt Lymphoma with Atypical Features, B-Int – B-cell Lymphoma Intermediate between BL and DLBL, DLBL – Diffuse Large B-Cell Lymphoma

Statistical analysis

Analysis was performed by an independent pathologist who had not seen the slides and the names of participants were blinded. All analyses were performed on SPSS for Windows v19 on a standard computer running Windows 7.

Variabilities among the pathologists were assessed using multiple parameters. Interobserver variabilities in the entire set of pathologists, between each pair of pathologists as well as between the three groups of pathologists, for morphological features and final independent diagnosis were studied using kappa statistics. Accuracy (agreement with the consensus diagnosis) was assessed by comparing the morphological diagnoses, independent final diagnoses and revised diagnoses after discussion with the consensus

diagnosis. The benefits of immunohistochemistry and of discussion were assessed by measuring the change in accuracy of independent final diagnosis over morphological diagnosis and revised diagnosis over independent final diagnosis, respectively. Finally, sensitivity and specificity of each group of pathologists to diagnose BL (typical and atypical) were calculated. Since, the pathologists participating in this study represent only a sample of the larger population of pathologists, it was concurred that a point estimate would be an over simplification. Hence, Monte-Carlo simulation was applied for sensitivity and specificity for each group of pathologists (using the Beta distribution). The Monte-Carlo simulation was run using the Oracle Crystal Ball Fusion Edition, Release 11.1.2.2.000 (www.oracle.com/crystalball) and Microsoft Excel 2010.

RESULTS

Initial independent assessment

Three pathologists committed to a diagnosis in 24/25 cases, one pathologist committed to a diagnosis in 23/25 cases, while the rest ten pathologists committed to a diagnosis in all the 25 cases. A summary of the different diagnoses offered is illustrated in Figure 1.

The concordance was poor for independent diagnosis among all the pathologists with kappa statistics (\pm SE) of 0.168 (\pm 0.018). Level of experience with

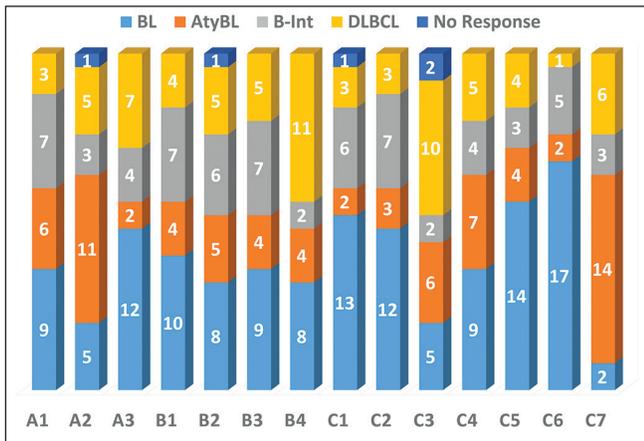


Figure 1: Bar diagram showing the various diagnoses offered by 14 pathologists. While 10 pathologists submitted responses in all 25 cases, 3 pathologists submitted responses in only 24 cases while 1 pathologist submitted responses in 23 cases. A1-A3: Expert lymphoma pathologists; B1-B4: Pathologists with experience in lymphomas; C1-C7: General lymphoma pathologists

lymphoma diagnosis showed direct correlation with the kappa statistics, with expert lymphoma pathologists having the highest (albeit only fair) concordance (kappa = 0.373 \pm 0.071) and general pathologists the lowest concordance (kappa = 0.138 \pm 0.035).

Interobserver variation in morphological features

Concordance for the morphological features tested was very low. A summary of the kappa statistics is presented in Table 2. The concordance for morphological diagnosis was highest among expert lymphoma pathologists (kappa = 0.356 \pm 0.127). Among all the morphological features assessed, highest concordance was noted for nuclear contour (kappa = 0.896 \pm 0.110) and lowest for nucleolar prominence (kappa = -0.062 \pm 0.124).

What parameters did pathologists use to differentiate between classic BL, atypical BL and B-cell lymphoma intermediate between Burkitt's and DLBL?

Morphological and key immunohistochemical features used by the 14 pathologists were cross-tabulated against the independent final diagnosis offered, a graphical summary of which is presented in Figure 2. It is evident that pathologist were least likely to accept deviation from certain features perceived to be very characteristics of BL, namely intermediate cell size, CD10 positivity and MIB-1 labeling of greater than 90%. They were however willing to accept BCL2 positivity, irregular nuclear contours, and multiple conspicuous nucleoli as well as greater anisonucleosis than what is expected in a classic BL. Greater the deviation in these features, pathologists were more likely to classify the case as either atypical BL or B-cell lymphoma intermediate

Table 2: Interobserver variability for morphological features, independent diagnosis and revised diagnosis along with its concordance with consensus diagnosis of 14 pathologists in this set of 25 cases of suspected Burkitt's lymphoma

Criteria		All Pathologists	Expert Lymphoma Pathologists	Pathologists with Experience in Lymphomas	General Pathologists
Pathologists			Actively examining >500 cases/yr	Experienced in lymphoma, but less than that of experts	No specialized training/experience in lymphomas
Pathologists			A1, A2 and A3	B1, B2, B3 and B4	C1, C2, C3, C4, C5, C6 and C7
Interobserver variation (kappa ^[1] \pm ASE)	Morphological Diagnosis	0.192 \pm 0.050	0.355 \pm 0.127	-0.333 \pm 0.356	-0.166 \pm 0.176
	Cell size	0.217 \pm 0.072	-0.333 \pm 0.117	0.085 \pm 0.131	0.286 \pm 0.050
	Nucleolar prominence	-0.062 \pm 0.124	0.134 \pm 0.095	-0.263 \pm 0.266	0.228 \pm 0.059
	Anisonucleosis	0.161 \pm 0.088	-0.095 \pm 0.089	0.197 \pm 0.126	0.123 \pm 0.095
	Nuclear contour*	0.896 \pm 0.110	0.052 \pm 0.118	0.528 \pm 0.188	0.284 \pm 0.091
	Independent Diagnosis	0.168 \pm 0.018	0.373 \pm 0.071	0.222 \pm 0.052	0.138 \pm 0.035
	Revised Diagnosis	0.835 \pm 0.021	0.852 \pm 0.074	0.796 \pm 0.055	0.891 \pm 0.044
Concordance with Consensus Diagnosis (kappa ^[2] \pm ASE)	Independent diagnosis	0.259 \pm 0.039	0.513 \pm 0.140	0.308 \pm 0.066	0.122 \pm 0.038
	Revised Diagnosis	0.633 \pm 0.011	0.651 \pm 0.032	0.592 \pm 0.030	0.655 \pm 0.021
	P#	0.00017	0.332	0.054	0.00045

[1]Fleiss kappa, [2]Cohen's kappa; Note: Kappa statistic values range from -1 to +1 indicating perfect discordance and perfect concordance respectively; *Maximum concordance for assessment of morphological features was noted for "Nuclear Contours"; #p value obtained from a paired t-test for statistical significance differences between the independent diagnosis and revised diagnosis for concordance with the consensus diagnosis

between Burkitt's and DLBL. Greater deviations from a typical morphology were accepted more easily for classic BL in pediatric patients during the independent diagnosis assessment though this trend was not statistically significant. These trends were seen in all the three groups of pathologists. However, no group of pathologists was homogenous in their interpretations, though the maximum concordance was evident in the group of expert lymphoma pathologists.

Consensus diagnosis

Twelve out of 14 pathologists participating in the study attended the consensus meeting. Consensus was reached on the final diagnosis in 23 out of 25 cases. Consensus was reached unanimously among the 12 pathologists in 19 cases, while consensus was assumed in 4 cases when 8 or more of 12 ($\geq 75\%$) pathologists agreed upon a single diagnosis. Consensus was not reached in two cases. The morphology of one of those cases is depicted in Figure 3; this was a 50-year-old lady in good general condition with a Stage I AX disease involving parotid node where nine pathologists labeled it as BL and four pathologists including the three

expert pathologists labeled it as B-cell unclassifiable. The other case was a HIV-positive male with a nasopharyngeal mass where three pathologists still labeled as BL in spite of having a large cell size in view of high proliferation, while four pathologists diagnosed it as DLBL and six as B-cell lymphoma unclassified. C-MYC/bcl2/bcl6 FISH was attempted in both cases but failed due to poor quality of paraffin processing.

The revised diagnoses offered by each pathologist were tabulated and their concordance with each other as well as with the consensus diagnoses was estimated. The level of agreement among the pathologists for the revised diagnosis after consensus meeting was assessed and was found to be very high ($\text{kappa} = 0.835 \pm 0.021$) and was similar across each group of pathologists.

Revision of diagnosis at consensus meeting was highest in the group of general pathologists and least in the group of expert lymphoma pathologist, though not statistically significant ($P = 0.121$) [Table 3]. Revision of diagnoses was highest for cases which were diagnosed as either atypical BL

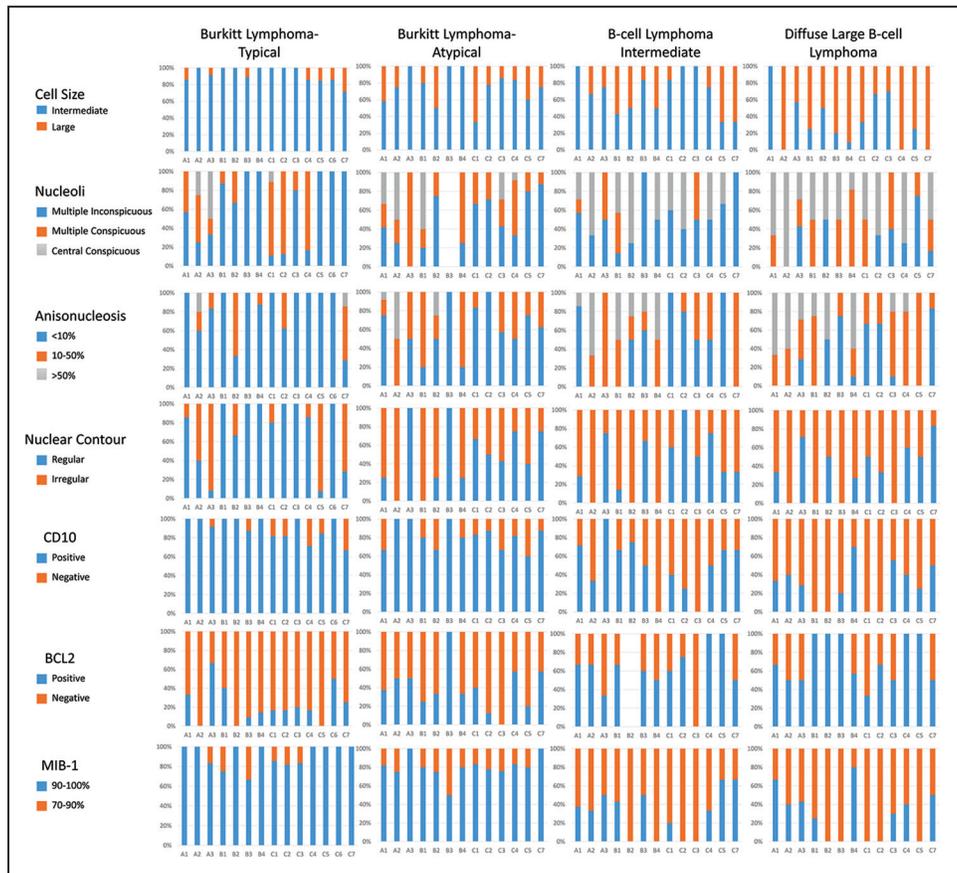


Figure 2: Composite bar diagram summarizing the salient morphological characters and immunohistochemistry parameters reported by each pathologist in different diagnostic categories. A1-A3: Expert lymphoma pathologists; B1-B4: pathologists with experience in lymphomas; C1-C7: general lymphoma pathologists. The chart enables at-a-glance visualization of diagnostic criteria used by different pathologists to offer a particular diagnosis and thus the variability among them

or B-cell lymphoma intermediate between BL and DLBL [Figure 4] and minimum revision occurred in classic BL group ($P = 0.001$).

Concordance with the consensus diagnosis

The concordance of independent diagnosis offered by each pathologist with the consensus diagnosis was low (mean kappa statistics (\pm SE) of 0.259 ± 0.039 and median of 0.207) and at the same time highly variable with a range of kappa statistics from -0.131 to 0.667. Moreover, the concordance with consensus diagnosis increased and variability reduced significantly with increasing experience of diagnosing lymphomas [Figure 5]. The concordance of the revised diagnoses with the consensus diagnoses shifted to very high (mean kappa statistics (\pm SE) of 0.633 ± 0.011 and median of 0.656) across the three groups of pathologists.

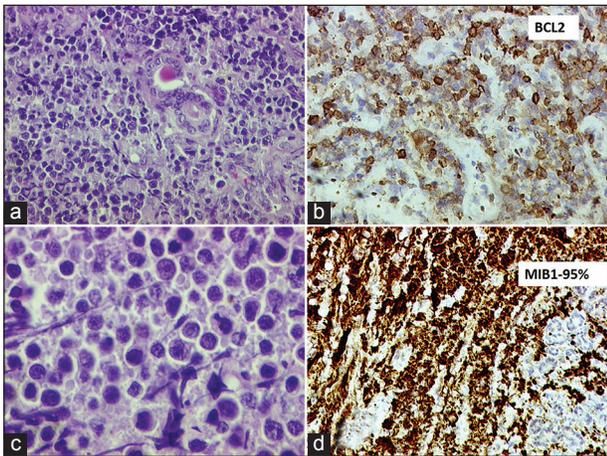


Figure 3: Panel of photomicrographs depicting the findings of one case in which consensus was not reached regarding the diagnosis. (a) Low-power photomicrograph (H and E stain, 100x original magnification) showing monomorphic lymphomatous infiltrate within salivary gland tissue. (b) High-power photomicrograph (H and E stain, 400x original magnification) showing non-cleaved cells with occasional prominent nucleoli. (c) BCL2 immunohistochemistry (200x original magnification, automated IHC using Ventana BenchMark XT) showing strong positivity in the lymphoma cells. (d) MIB-1 labeling index was >95%.

Effect of tissue fixation, age group and provision of additional information on revision of diagnoses

All biopsies were graded for quality of fixation and staining by morphological assessment on a semi quantitative scale of 1-3 with 3 being the best possible fixation and staining. There was no difference in the distribution of fixation and staining scores across the diagnostic categories ($P = 0.654$). There was no effect of fixation on the proportion of cases being reclassified at consensus meeting with equal proportion of cases being re-classified in all the three grades of fixation (means of 54.167 ± 29.167 , 47.222 ± 7.217 and 50 ± 6.989 in grades 1,2 and 3 respectively, ANNOVA, $P = 0.931$).

There was no effect of *C-MYC* status, EBER-ISH results or of BCL6 IHC results on the frequency of revision of diagnoses. Similarly there was no difference in the

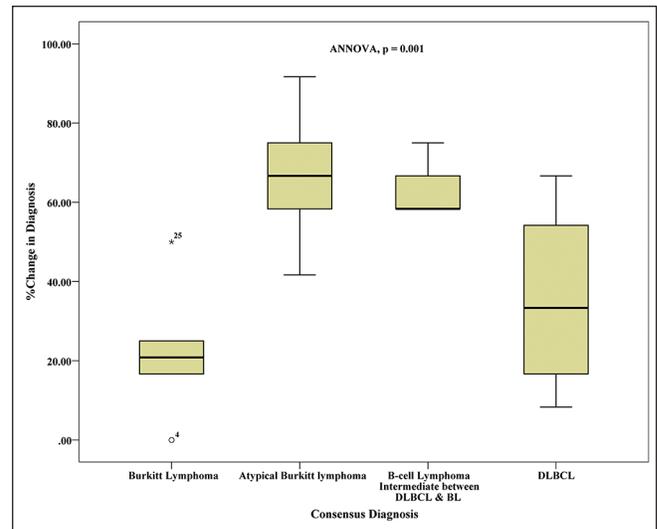


Figure 4: Box plot showing the proportion of revision of diagnosis at consensus meeting in each diagnostic category. The change was maximum in the diagnostic categories of Burkitt lymphoma with atypical features and B-cell lymphoma intermediate between BL and DLBL, while it was least with Burkitt lymphoma

Table 3: Summary of percentage revision of diagnosis and accuracy of diagnoses by the three groups of pathologists

	All Pathologists	Expert Lymphoma Pathologists	Pathologists with experience in lymphomas	General Pathologists	p*
Revision of diagnosis (%)	51.667 \pm 5.313	38.67 \pm 13.131	46.0 \pm 8.718	64.0 \pm 4.381	0.121
Accuracy (%)					
Morphological Diagnosis (mean \pm SE)	36.79 \pm 2.631	41.889 \pm 3.466	33.333 \pm 6.943	36.086 \pm 3.742	0.249
Independent diagnosis	45.963 \pm 4.903	66.667 \pm 13.825	51.087 \pm 4.820	34.161 \pm 3.727	0.014
Revised diagnosis after consensus meeting	95.652 \pm 1.311	97.101 \pm 2.898	92.391 \pm 2.735	97.391 \pm 1.469	0.288
Accuracy (%)					
Burkitt lymphoma (mean \pm SE)	72.619 \pm 7.536	83.333 \pm 9.622	70.833 \pm 12.5	69.047 \pm 13.327	0.784
Burkitt lymphoma with atypical features	24.186 \pm 7.026	48.148 \pm 25.926	21.667 \pm 0.556	15.357 \pm 7.951	0.198
Intermediate	35.714 \pm 10.166	55.556 \pm 29.397	66.667 \pm 13.608	9.524 \pm 6.147	0.017
DLBL	58.928 \pm 8.535	91.667 \pm 8.333	68.75 \pm 11.968	39.286 \pm 10.714	0.029

*p values obtained by Annova for comparing the accuracy between the three groups of pathologists followed by Bonferroni test as the *post-hoc* test wherever necessary

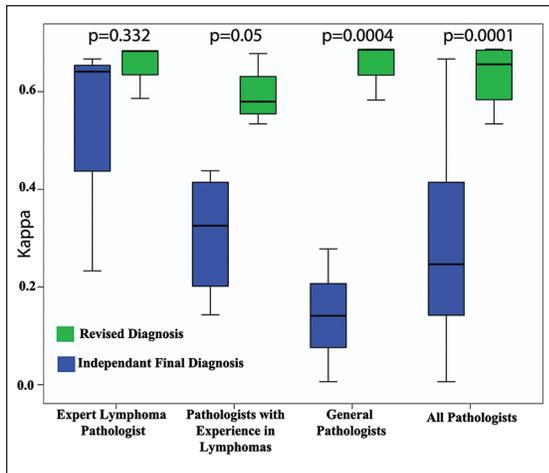


Figure 5: Composite box plot showing the concordance of initial independent diagnosis and revised diagnosis with the consensus diagnosis for each group of pathologists

rates of revision of diagnoses at consensus meetings whether the case belonged to an adult or pediatric age group (mean revision of $45.513 \pm 6.579\%$ and 53.472 ± 7.429 in pediatric (age < 18 years) and adult age groups, respectively, independent t-test, $P = 0.429$), whether patient was HIV positive (mean revisions $46.667 \pm 15.275\%$) or HIV negative (mean revisions $50 \pm 5.058\%$, $P = 0.792$) or whether the clinical presentation was classical of BL (mean revision 46.795 ± 7.580) or not classical of BL (mean revisions $52.083 \pm 6.333\%$, $P = 0.601$). There was also no statistically significant difference in the consensus diagnosis (i.e. Burkitt's vs. others) for age groups ($P = 0.073$), HIV status ($P = 0.740$) or for site of disease ($P = 0.291$).

Accuracy of pathologists (morphology, immunohistochemistry and consensus discussions)

We measured the accuracy (agreement with consensus diagnosis) of our set of 14 pathologists to make a correct diagnosis with the consensus diagnosis being taken as the gold standard [Table 3].

Accuracy of morphological diagnosis with respect to the consensus diagnosis was just $36.79 \pm 2.631\%$ with the expert lymphoma pathologists having the highest accuracy ($\sim 42\%$) while the least accuracy was seen in the group of pathologists with lesser experience in lymphomas ($\sim 33\%$).

After immunohistochemistry (independent final diagnosis), the accuracy of the entire group of pathologists increased to $45.963 \pm 13.825\%$, with expert lymphoma pathologists having a mean accuracy of $66.667 \pm 13.825\%$ (mean \pm SE), while pathologists with experience in lymphoma made a correct diagnosis $51.087 \pm 4.820\%$ of the times and the general pathologists got it right only $34.161 \pm 3.727\%$ of times. The group of “expert lymphoma pathologists” had a significantly higher likelihood of making a correct

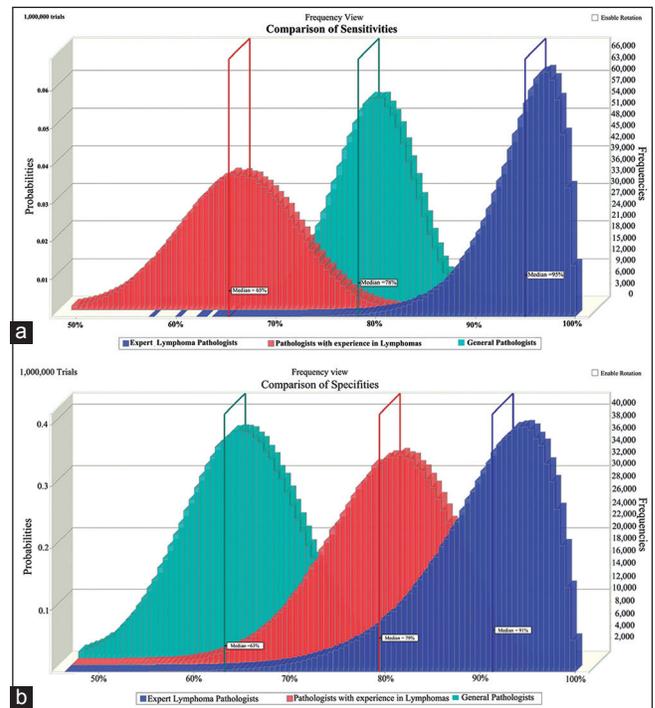


Figure 6: Overlay histograms obtained from Monte-Carlo simulation depicting the (a) sensitivities and (b) specificities of the three groups of pathologists to diagnose Burkitt's lymphoma. Monte-Carlo simulations were performed for 1 million trials. The median values of each group are depicted

diagnosis as compared to both the “pathologists with experience in lymphomas” (OR = 3.14, $P = 0.012$) and the “general pathologists” (OR = 5.3, $P = 0.00032$). Similarly, a higher likelihood of correct diagnosis was seen with the “pathologists with experience in lymphoma” as compared to the “general pathologists”, but was not statistically significant (OR = 1.69, $P = 0.062$). The accuracy of all pathologists increased after the consensus meeting with the accuracy of revised diagnoses being $95.652 \pm 1.311\%$, with very little variation within each group of pathologists ($97.101 \pm 2.898\%$, 92.391 ± 2.735 and $97.391 \pm 1.469\%$ accuracy for the group of expert lymphoma pathologists, pathologists with experience in lymphomas and general pathologists, respectively).

The mean change of accuracy by immunohistochemistry over morphology was 9.698 ± 4.799 , while the mean change of accuracy by discussion/consensus meeting over that by immunohistochemistry was $47.464 \pm 5.039\%$.

Accuracy was also the highest in the classic Burkitt lymphoma group with a mean accuracy of $72.619 \pm 7.536\%$ followed by accuracy in DLBL at $58.928 \pm 8.535\%$. Accuracy was lowest in the cases of atypical BL and B-cell lymphoma intermediate between BL and DLBL ($24.186 \pm 7.026\%$ and $35.714 \pm 10.166\%$, respectively). Again, the accuracy in each category of consensus diagnosis

was highest among the expert lymphoma pathologists and lowest among the group of general pathologists.

Sensitivity and specificity to diagnose Burkitt's lymphoma

The sensitivity and specificity of the group of expert lymphoma pathologists to diagnose BL (typical and atypical cases) was the highest (96.88% and 94.44%, respectively). Interestingly, the group of general pathologists had a higher sensitivity than the group B pathologists (78.57% vs. 65.63%, respectively), though the later had a much higher specificity (80.56% vs. 63.49%). As these were just point estimates, Monte-Carlo simulation was applied which revealed similar findings, though there were significant overlaps in both sensitivity and specificity [Figure 6].

DISCUSSION

The dexterity of lymphoma diagnosis has always been experience driven, however not many studies focus in details on the interobserver variation in lymphoma diagnosis and methods to reduce it. In 2005, El-Zimaity *et al.*^[1] demonstrated that pathologists with greater experience in interpretation of gastrointestinal biopsies and hemolymphoid biopsies had greater odds of diagnosing MALT lymphomas and its look-alikes correctly than the general pathologists. Provision of greater clinical details, endoscopic findings as well as provision of adequate amount of tissue for histopathological assessment facilitated greater confidence to the pathologist in one's own diagnosis. While MALT lymphoma often is a diagnostic challenge, our results in a monomorphic blastic tumor like BL and its look-alikes (where we expect more concordance) are no different.

Burkitt lymphoma is an highly aggressive disease which however is nearly curable with upfront aggressive combination chemotherapy at least in children.^[2] This dictates the criticality of diagnosis in Burkitt lymphoma. We tried to investigate the factors contributing to the inter-observer variability in the diagnosis of BL and its look-alikes in a group of pathologists working within a relatively resource-constrained setting. The morphologic spectrum of Burkitt lymphoma has expanded since its original description.

In 1994, when the REAL classification was published, the Burkitt-like lymphoma (BLL) category was not thought to be a histologically reproducible category or a distinct clinicopathologic entity. The SWOG study in the post REAL classification era addressed the issue of reproducibility of BLL, Burkitt lymphoma and DLBL diagnosis for the first time.^[3] Survival studies indicated a

reduction of more than 50% in median survival time for BLL patients compared with DLBL patients (1.2 years versus 2.5 years), though the 5-year survival rates were remarkably similar (40% of BLL patients versus 42% of DLBL patients) indicating that BLL might indeed be closer to BL in behavior though it shows greater morphologic variation. The level of agreement on the diagnosis of BLL in their study (52%) was similar to that seen among the general pathologist in our study and less than expert lymphoma pathologist. The improvement observed in this study may be due to our better understanding of Burkitt lymphoma morphologic spectrum in the post WHO 2008 and in the gene profiling era.

While the gold standard for Burkitt remains at a molecular level, translation of that standard into practice has evolved in last decade. Dave *et al.*^[4] documented that "Molecular Burkitt" as defined by gene-expression profiling signature identified all 25 cases of classic BL that had been verified by an expert panel of hematopathologists. However, nine aggressive lymphomas that were diagnosed as DLBL or high-grade lymphoma by the panel were classified as BL on the basis of gene-expression profile clearly outlining that a small percentage of Burkitt lymphomas are missed.^[4]

Using a set of 220 cases of mature aggressive B-cell lymphomas, Hummel *et al.* identified "a Burkitt's gene signature" of 58 genes.^[5] A major contribution of this gene profiling study was that it documented that Burkitt lymphoma is "MYC-simple," namely: - it has *IGH-MYC* fusions and a low chromosomal complexity score (<6) with no *IGH-BCL2* and *BCL6* translocations. On other hand "MYC-complex" or lymphomas with non-*IGH-MYC* fusions or with *IGH-MYC* fusions that have a high chromosomal complexity score, an *IGH-BCL2* fusion, or *BCL6* breakpoint, or any combination of these were more likely to be aggressive DLBL. This then translated into the use of the triple translocations (*MYC/Bcl2/Bcl6*) by FISH for a more accurate diagnosis of Burkitt lymphoma and separating them from the triple hit or double hit DLBL harboring two or three of these translocations.

Though the gold standard may be molecular, morphology continues to be the base for evaluation and the initial therapy in Burkitt lymphomas world over.

Our study shows that interobserver variability in the interpretation of BL is high (albeit the least) even among "Expert Lymphoma Pathologists". This variability stems from differences in the interpretation of morphological and immunohistochemical findings which were directly proportional to the level of experience in diagnosing lymphomas. There is no study evaluating interobserver variability in the diagnosis of BL and its look-alikes

after publication of the WHO2008 classification of haematolymphoid malignancies. In a study evaluating inter-rater agreement in BL based on the R.E.A.L. classification by Lones *et al.*,^[6] there was good agreement among the pathologists with respect to the consensus diagnosis for BL (88%) and DLBL (80%) while it was low for “Burkitt-like lymphomas” (42%). The results of our study also show similar findings, with highest accuracy (i.e. agreement with the consensus diagnosis) in the cases of classic BL (~73%), followed by DLBL (~60%) and least among cases of Burkitt lymphoma with atypical features (~24%) and B-cell lymphoma intermediate between Burkitt and DLBL (~36%).

In a study from Uganda, Ogwang *et al.*^[7] documented that morphological assessment of biopsies alone improved diagnostic accuracy of clinically diagnosed BL only marginally from 75% to 82%. They highlighted the importance of optimal fixation and processing on the variability in morphological assessment as well as on immunohistochemistry, but did not evaluate the additional benefit of immunohistochemistry over morphology in improving diagnostic accuracy. In our study, the mean increase in accuracy after immunohistochemistry was ~10% over and above morphological assessment. On the other hand, the accuracy in our set of pathologists more than doubled after the consensus meeting with the maximum increase being seen in the group of general pathologists. A previous German study documented that immunohistochemistry-based algorithm was better than morphology in classifying BL. Two algorithms for classification were followed: algorithm A used a two-step review by four hematopathologists and algorithm B a set of only biologic markers (Ki-67 > or = 90%, CD10+, bcl6+, bcl2-, MYC breakpoint+, BCL2 and BCL6 breakpoint-). BL according to algorithm B was more homogeneous with respect to clinical presentation (gender and localization) than BL defined by algorithm A.^[8]

Clinical information is no doubt crucial in assessing histopathological information, clinical factors like adult vs. pediatric age group, HIV status and site of disease presentation (classical for BL vs. not classical of BL) did not show statistically significant impact on the revision of diagnosis during the consensus meeting, implying that these factors hardly affect the variability in the diagnosis of BL.

Naresh *et al.*^[9] proposed an algorithmic approach in differentiating BL and its look-alikes in a three-phase approach with phase I (morphology with CD10 and BCL2), phase II (CD38, Ki-67 and CD44) and phase III (FISH on paraffin sections for MYC, BCL6, BCL2 and IG gene rearrangements) being able to categorize 82%, 92% and 95% of these cases, respectively. In their

study, basic morphology i.e. lymphomas composed of monomorphic infiltrate of medium-sized lymphoid cells with fine chromatin and lack of conspicuous nucleoli, absence of TdT and/or cyclin D1 expression were essential before the scoring system is employed. While we do not have backup of molecular profiling and the study set is small what we wished to focus was on the wider morphologic spectrum of Burkitt lymphoma and the limited IHC workup that would be available in most resource-constrained laboratories.

We found no effect of provision of EBER-ISH and C-MYC FISH data at the consensus meeting on the proportion of revision of diagnosis but the interpretation is limited by the fact that they were often not satisfactory due to the referral nature of the material. As we did not do FISH for BCL6, BCL2 and IG gene rearrangements, we are not in a position to comment with certainty on the additional benefit of molecular information in categorizing BL and its differentials. There was no gold standard and a triad of clinical features especially site and evidence for high doubling time; morphology and immunohistochemistry was used for diagnosis.

Though we did not follow any algorithmic approach, the proforma provided included all morphologic parameters to understand where pathologist err maximally and it appears that there is a clearly wider range for use of cell size and nuclear anisonucleosis among pathologist with lower concordance rates. However, a joint discussion of morphologic features and patient outcomes improved understanding of this spectrum and the concordance rates. The chief conclusion thus is that experience with a larger number of cases and consensus exercises such as the one we conducted is a simple and more effective way of improving accuracy than provision of high end molecular data or a more extensive immunohistochemical panel of antibodies in countries with limited resources. The benefits of discussions are more evident for a general pathologist than for an expert lymphoma pathologist. A more meaningful assessment of benefit of discussion might be possible by repeating the experiment a few days after discussion/consensus meeting.

CONCLUSIONS

Interobserver variability in the interpretation of BL and its look-alikes is significantly large. One of the biggest reasons is the differences in interpretations in addition to the significant overlap of morphological and immunohistochemical features of BL, BL with atypical features, B-cell lymphoma intermediate between BL and DLBL and DLBL. Interpretation is also affected by the

level of experience with lymphomas as is evident in our study. Furthermore, pathologists with even similar level of expertise were more likely to differ if they offered a diagnosis independently than when they offered a diagnosis after a discussion.

ACKNOWLEDGEMENT

The authors would like to acknowledge the sound guidance and help offered by Mr. Huybert Groenendaal from Epix Analytics (HYPERLINK “<http://www.epixanalytics.com>” www.epixanalytics.com) with the Monte Carlo Simulation.

REFERENCES

1. El-Zimaity HM, Wotherspoon A, de Jong D, Houston MALT lymphoma Workshop. Interobserver variation in the histopathological assessment of malt/malt lymphoma: Towards a consensus. *Blood Cells Mol Dis* 2005;34:6-16.
2. Magrath I. Treatment of Burkitt lymphoma in children and adults: Lessons from Africa. *Curr Hematol Malig Rep* 2006;1:230-40.
3. Brazier RM, Arber DA, Slovak ML, Gulley ML, Spier C, Kjeldsberg C, *et al.* The Burkitt-like lymphomas: A Southwest Oncology Group study delineating phenotypic, genotypic, and clinical features. *Blood* 2001;97:3713-20.
4. Dave SS, Fu K, Wright GW, Lam LT, Kluin P, Boerma EJ, *et al.* Molecular Diagnosis of Burkitt's Lymphoma. *N Engl J Med* 2006;354:2431-42.
5. Hummel M, Bentink S, Berger H, Klapper W, Wessendorf S, Barth TF, *et al.* A Biologic Definition of Burkitt's Lymphoma from Transcriptional and Genomic Profiling. *N Engl J Med* 2006;354:2419-30.
6. Lones MA, Auperin A, Raphael M, McCarthy K, Perkins SL, MacLennan KA, *et al.* Mature B-cell lymphoma/leukemia in children and adolescents: Intergroup pathologist consensus with the revised European-American Lymphoma Classification. *Ann Oncol* 2000;11:47-51.
7. Ogwang MD, Zhao W, Ayers LW, Mbulaiteye SM. Accuracy of Burkitt lymphoma diagnosis in constrained pathology settings: Importance to epidemiology. *Arch Pathol Lab Med* 2011;135:445-50.
8. Haralambieva E, Boerma EJ, van Imhoff GW, Rosati S, Schuurin E, Muller-Hermelink HK, *et al.* Clinical, immunophenotypic, and genetic analysis of adult lymphomas with morphologic features of Burkitt lymphoma. *Am J Surg Pathol* 2005;29:1086-94.
9. Naresh KN, Ibrahim HA, Lazzi S, Rince P, Onorati M, Ambrosio MR, *et al.* Diagnosis of Burkitt lymphoma using an algorithmic approach - applicable in both resource-poor and resource-rich countries. *Br J Haematol* 2011;154:770-6.

How to cite this article: Rane SU, Shet T, Sridhar E, Bhele S, Gaikwad V, Agale S, *et al.* Interobserver variation is a significant limitation in the diagnosis of Burkitt lymphoma. *Indian J Med Paediatr Oncol* 2014;35:44-53.
Source of Support: Nil. **Conflict of Interest:** None declared.

Staying in touch with the journal

1) Table of Contents (TOC) email alert

Receive an email alert containing the TOC when a new complete issue of the journal is made available online. To register for TOC alerts go to www.ijmpo.org/signup.asp.

2) RSS feeds

Really Simple Syndication (RSS) helps you to get alerts on new publication right on your desktop without going to the journal's website. You need a software (e.g. RSSReader, Feed Demon, FeedReader, My Yahoo!, NewsGator and NewzCrawler) to get advantage of this tool. RSS feeds can also be read through FireFox or Microsoft Outlook 2007. Once any of these small (and mostly free) software is installed, add www.ijmpo.org/rssfeed.asp as one of the feeds.