Validation of An Image Analysis Algorithm for Quantifying CD138 Immunohistochemical Marker in Plasma Cell Neoplasms

Capstone Project

By

Ngoc Thy Bao Tran, MD

Department of Medical Informatics & Clinical Epidemiology

School of Medicine

Oregon Health & Science University

Spring 2020

School of Medicine

Oregon Health & Science University

Certificate of Approval

This is to certify that the Master's Capstone Project of

Ngoc Thy Bao Tran, MD

*"*Department of Medical Informatics & Clinical Epidemiology*:*

Validation of an image analysis algorithm for quantifying CD138 immunohistochemical marker in plasma cell neoplasms"

Has been approved

_____

William R. Hersh, MD – Capstone Advisor

Table of Contents

CHAPTER 1. INTRODUCTION 4

CHAPTER 2. BACKGROUND ..... **Error! Bookmark not defined.**

CHAPTER 3. MATERIAL AND METHODS 6

3.1 CASE SELECTION 6

3.2 IMMUNOHISTOCHEMISTRY 6

3.3 7

INFORMATION ABOUT INSTRUMENT AND ITS RELATED SOFTWARE 7

Slide scanner 7

Slide viewing software and image analysis 7

3.4 DESIGN 8

Scanning and preparing digital slides 8

Generating testing images and obtaining ground truth 8

Testing phase with Aperio image analysis 8

Testing phase with hematopathologists 9

3.5 9

STATISTICS 9

CHAPTER 4. RESULTS ..... **Error! Bookmark not defined.**

CHAPTER 5. DISCUSSION 18

CHAPTER 6. CONCLUSIONS 19

APPENDICES 21

Appendix A. Raw data of plasma cell percentages obtained from four pathologists, the algorithm, and ground truth 21

Appendix B. Two-tiered classification based on diagnostic impact of estimated plasma cell percentages 22

Appendix C. Three-tiered classification based on estimated plasma cell percentages. 24

GLOSSARY 25

REFERENCES 26

CHAPTER 1. INTRODUCTION

Plasma cell neoplasms are a spectrum of disorders caused by the proliferation of monoclonal plasma cells in the bone marrow leading to the production of monoclonal immunoglobulin. Included in this spectrum, from least severe to most severe, are monoclonal gammopathy of undetermined significance, smoldering multiple myeloma, and plasma cell myeloma. Each diagnosis carries a different prognosis and requires different clinical management. For example, the median overall survival of patients with Stage I, II, and III PCM has been shown to be 62, 44, and 29 months, respectively, while the median overall survival of patients with monoclonal gammopathy of undetermined significance is expected to be only slightly shorter than that of age-matched controls.[1-4] Thus, making the correct diagnosis amongst this spectrum of disorders is paramount for the prognostication and proper clinical management of plasma cell neoplasms.

Monoclonal gammopathy of undetermined significance (MGUS) is a precancerous proliferation of plasma cells, while smoldering multiple myeloma (SMM) is a malignant form of plasma cell neoplasms. Both MGUS and SMM lack all signs of end-organ or tissue impairment such as hypercalcemia, renal insufficiency, anemia, or bony lesions (referred to as CRAB features).[5] The difference between the two lies on the percentage of monoclonal plasma cells in a bone marrow biopsy (BMBx). For MGUS, this percentage is required to be <10%, while that is between 10% and 60% for SMM.[6] Finally, the diagnosis of plasma cell myeloma (PCM) requires either ≥10% plasma cells on a BMBx and at least one end-organ damage or ≥60% plasma cells on a BMBx with or without end-organ damages (Table 1). [6]

Chronically, plasma cell percentage (PC%) was first achieved by performing differential counts on a Wright-Giemsa-stained bone marrow aspirate smear. This practice was followed by using a CD138 stain, a plasma cell immunohistochemical biomarker in hematology, on a bone marrow trephine biopsy or related blood clot as an additional measurement to improve accuracy. This is because many factors reportedly can affect the quality of the aspirate smears[7], leading to the fact that PC% in aspirate smears tend to be lower than that in trephine bone marrow biopsies.[7-10] Adding CD138 staining to the routine practice has provided a more accurate PC% when pathologists are expected to report the higher PC% estimate amongst the two approaches according to the most current practice standard from the International Myeloma Working Group[6].

However, PC% obtained from visual estimation of a BMBx at a lower power field under a microscope can create potential inter- and intraobserver variability and inaccuracy. The contributing factors for these issues are the inherently uneven distribution of fat cells, normal nucleated bone marrow cells, and infiltrating, neoplastic plasma cells in a trephine BMBx. In addition, the contamination of peripheral non-nucleated red blood cells in the BMBx sections or especially blood clot from the aspirate is another challenge for pathologists to accurately estimate PC% (Figure 1). Of note, PC% is calculated by dividing the number of plasma cells present in the BMBx or clot by the number of total nucleated cells in the bone marrow that do not include contaminated peripheral non-nucleated red blood cells. The ability to quantify CD138 marker expression using computer-aided image analysis could provide a promising

solution to improve these issues.[11] As seen in Table 1 and the above mentioned information, we can appreciate how crucial the accurate and consistent PC% in a BMBx is to the diagnosis, treatment, and outcome of a patient with a plasma cell neoplasm.

Table 1. Plasma cell neoplasms in relation to end-organ damage and monoclonal plasma cell percentage

Plasma cell neoplasm	Classification	End-organ damage	Monoclonal plasma cell %
MGUS	Precancerous	Absent	<10%
SMM	Cancer	Absent	10% ≤ PC% ≤ 60%
PCM	Cancer	Present	10% ≤ PC% ≤ 60%
		Absent or Present	≥ 60%

MGUS: Monoclonal gammopathy of undetermined significance; **SMM**: Smoldering multiple myeloma; **PCM**: Plasma cell myeloma

CHAPTER 2. BACKGROUND

Literature shows that there are many studies that applied image analysis techniques toward quantifying biomarkers such as HER2 (associated with breast, esophageal, and gastric cancers)[12] and Ki-67 (associated with cell proliferation in many cancers)[13], but only a few were conducted on CD138 marker[9-11, 14]. All of these studies recommended image analysis as a good tool to obtain objective PC% from BMBx.

However, each of these CD138 studies solely used an area-based rather than a cell count-based approach for calculating PC%, and therefore none of those compared the image analysis-calculated PC% to a manually labeled ground truth[8, 9, 11], except a study by Lee et al. that had manual counts of selected bone marrow areas but no reported detail about the image analysis tool and how it measured PC%[10].

Our aims in this study are 1/ to evaluate the accuracy, interobserver variability of pathologist-estimated PC% and 2/ to validate an image analysis membrane algorithm of PC% both by comparing these PC% to those obtained from ground truth. This study also introduces the use of a cell count-based approach on our current image analysis tool for PC% calculation, which to the best of our knowledge has never been shown in the literature.

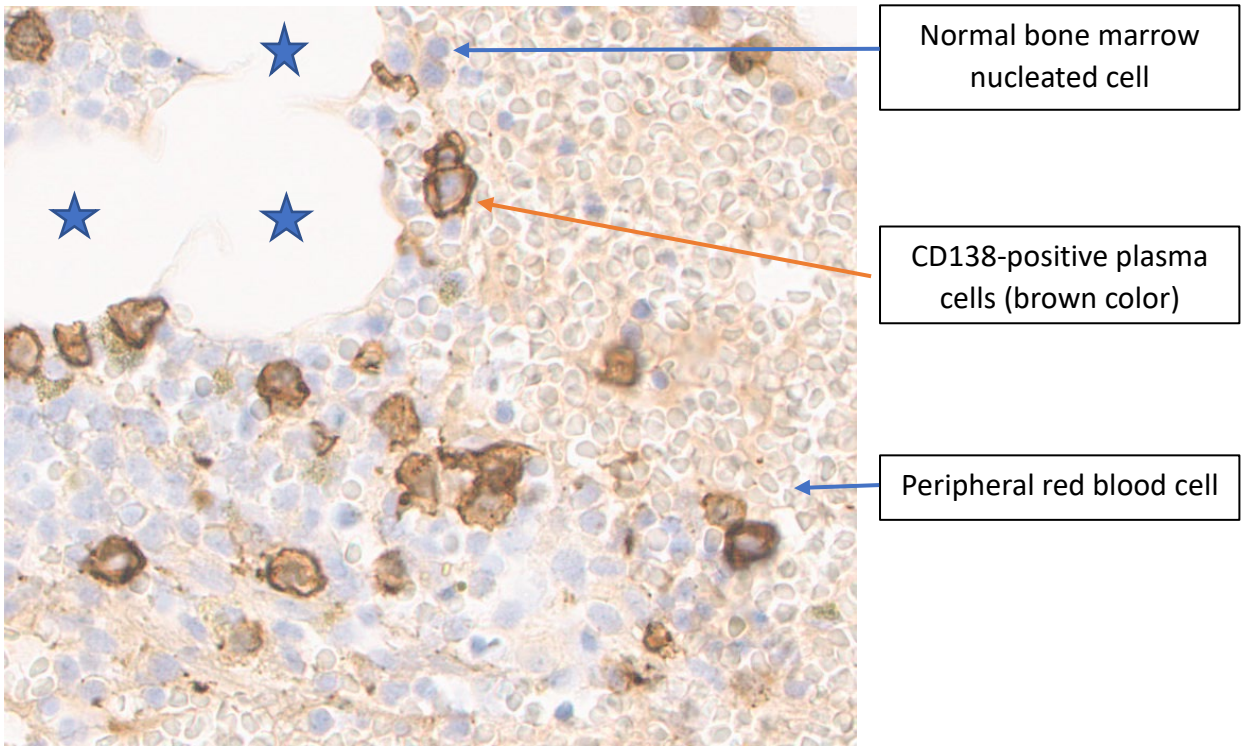


Figure 1. Section of bone marrow clot (40X) with CD138-negative nucleated cells, CD138-positive plasma cells (brown), and contaminated non-nucleated peripheral blood cells (smaller size, pale cytoplasm, some with crescent shape). The blue stars are fat vacuoles that are empty after tissue processing in which alcohol was used.

CHAPTER 3. MATERIAL AND METHODS

3.1 CASE SELECTION

The study was reviewed and approved by the Institutional Review Board. A third-party business intelligence software (SAP America, Inc., Newtown Square, PA, USA) that was integrated into the OHSU EHR (EPIC, Verona, WI, USA) and laboratory information system (BEAKER, EPIC, Verona, WI, USA) was queried to retrospectively identify patients with plasma cell neoplasms. Only adult patients (> 18 years of age) were included from a period from 2009 to 2019. These cases were cross-referenced with bone marrow biopsy reports to include only those samples for which there is also a bone marrow specimen from the same patient. A total of 53 cases were selected from the data set, not based on plasma cell percentages. Case numbers from these slides were deidentified.

3.2 IMMUNOHISTOCHEMISTRY

Three to three and a half micron sections of neutral buffered formalin fixed paraffin embedded bone marrow core biopsy or clot from bone marrow biopsy procedure when a core biopsy cannot be obtained were stained with immunohistochemical technique using CD138, a plasma cell biomarker in bone marrow specimens. This CD138 marker or syndecan-1 (B-A38, predilute; Cell Marque^{TM}, Rocklin, CA, USA) is a membranous, mouse monoclonal antibody. Benchmark

® Ultra automated stainer (Ventana Medical System, Tucson, AZ, USA) was used. The CD138 staining protocol has a pretreatment with CC1, pH 8.0, protease III (Ventana Medical System), the antigen retrieval with high pH for 36 minutes, followed by incubation with the CD138 antibody in 16 minutes at 42^0C degree. The antigen-antibody complexes are detected by ULTRAVIEW (UVIEW) DAB detection kit (Ventana Medical System). This CD138 protocol has been the same during the period that the biopsy samples were collected for this study.

3.3 INFORMATION ABOUT INSTRUMENT AND ITS RELATED SOFTWARE

Slide scanner

OHSU laboratory uses a commercial slide scanner, model Aperio ScanScope AT2 (Leica biosystems, Vista, CA, USA), to scan tissue on glass slides and convert the image of the tissue to a digital format (.svs) for multiple purposes including education, tumor board presentation, frozen section diagnosis, and research. These digital files are stored on the laboratory database with regulated access.

This Aperio ScanScope AT2 slide scanner (Figure 2) provides high volume, digital whole slide scanning function with up to 400 slide capacity. The objective lenses are capable of 40x scanning with 2x optical magnifying changer. The scanner has dimensions of 23.5 inches (H) x 16 inches (W) x 25.5 inches (D) and weight of 129 pounds.

Slide viewing software and image analysis

The slide viewing software, ImageScope – version 12.4.3.7001 (Leica biosystems, Vista, CA, USA), provided by the vendor, with no additional cost, helps users open, view, and navigate the digital tissue images. This software has a set of optionally purchased algorithms (membranous, nuclear, and cytoplasmic staining patterns) integrated into the image analysis tool to calculate the percentage of a certain cell type of interest. Since the algorithms are originally developed for research use only, coupled with their generic use purpose, we need to modify and validate when testing on a specific immunohistochemical biomarker before using for routine practice similar to the College of American Pathologists guideline[15] for HER2 immunohistochemistry on breast cancer.

Since CD138 antibody-antigen complexes express on plasma cell membrane, the membranous algorithm was chosen for this project. The algorithm has multiple parameters for users to modify. It had been tested on a training image set. Every version of the modified algorithm was saved, and these versions were compared across the training set in order to select the best algorithm that was used on the testing image set.

Figure 2. Slide scanner, model Aperio AT2 from Leica, used to scan slides for this project (from https://www.leicabiosystems.com/digital-pathology/scan/aperio-at2/)

3.4 DESIGN

Scanning and preparing digital slides

The 53 deidentified, selected slides described in the case selection were scanned by the Aperio ScanScope AT2 after being cleaned to remove dust by alcohol pads. These digital files were stored in the laboratory IT server managed by a laboratory IT specialist.

Generating testing images and obtaining ground truth

One of this study aims is to validate the accuracy of the membranous algorithm in quantifying CD138-positive cell (plasma cells) in a bone marrow biopsy. This can be achieved by comparing results from gold standard or manual counting method, which is ground truth, to the results generated by the algorithm or hematopathologists. The barrier to obtaining ground truth is that one entire core bone marrow biopsy (0.2 cm in width by 1.5 – 2.0 cm in length) or clot contains thousands of cells, making it time-consuming for manual counting to obtain ground truth on all 53 biopsies. Therefore, the study approach was extracting a 1000 x 4000-pixel image from an area that represents highest PC% on that slide. Ground truth was obtained by manually counting CD138 positive and negative cells on each image by using another open-source slide viewing software named QuPath.[16]

Testing phase with Aperio image analysis

These images were uploaded to the vendor-provided EslideManager website in order for these files to be saved and analyzed by Aperio's server. The website requires user authorization to login when users need to run its algorithms. The workflow of image analysis is seamless with all

cases being selected by the “select all” button and analyzed by clicking “analyze” button. The server processed and returned the number and percentage of CD138 positive cells with tiered positive intensity (i.e. 0+, 1+, 2+, 3+) (Figure 3-A &B). This study defined that the reported PC% was the total percentages of 3+ and 2+. These results were downloaded under a .csv format.

Testing phase with hematopathologists

A PowerPoint slide was created with each slide having one image along with its newly assigned number for the de-identification purpose. Each slide or image was set to automatically advance every 30 seconds. Four OHSU hematopathologists with different number of years of practice independently viewed the PowerPoint slide and documented their estimate for each case in an Excel file with the provided image names in the same order seen in the PowerPoint slide to facilitate the documentation and potentially prevent skipping case.

3.5 STATISTICS

All the raw data obtained from the study was checked for normality by using a Shapiro-Wilk test. Results from the algorithm and estimates from each hematopathologist were compared to ground truth using concordance coefficient correlation (CCC), Spearman, and Pearson tests. Since the cutoff 10% and 60% are crucial criteria in diagnosing MGUS, SMM, and PCM, the classification of the impact of each image’s PC% among four pathologists based on these cutoffs was expressed as a nominal variable (with and without impact on classification) and compared by a Kappa test. A focus sample of only misclassified images present at any pathologist’s estimation was divided into 3 ordinal categories: underestimated, expected, and overestimated (Table 4) and compared by a Kappa test. The total number of misclassifications based on these cutoffs from the algorithm or pathologists were compared using Chi-square test. Statistical significance was defined as two-sided P-value less than 0.05. Statistical analysis was performed using Stata 14.2 (StataCorp, College Station, TX, USA).

CHAPTER 4. RESULTS

Results obtained by Aperio algorithm on 53 images extracted from 53 cases show highest concordance coefficient correlation (0.991; 95% CI [0.987 - 0.996]) and highest Pearson correlation (0.992; 95% CI [0.986 – 0.995]) compared with these numbers from four pathologists. Spearman correlation test, which measures the monotone association, shows the highest correlation coefficient (0.948; 95% CI [0.91-0.97]) from pathologist 3 – whose PC%

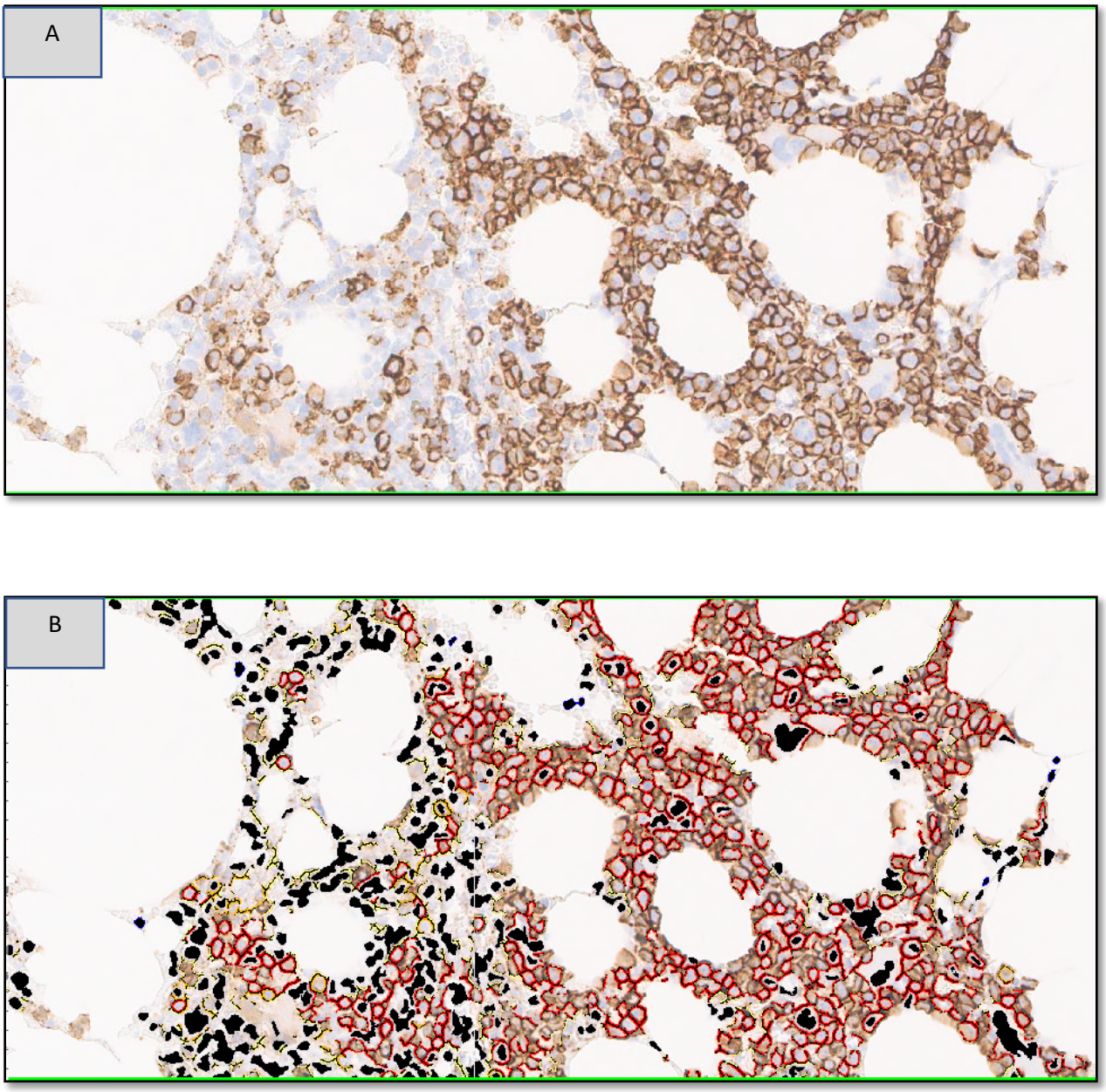


Figure 3. 3A – An illustrated bone marrow image (17X) of CD138 stain for plasma cells (brown) before being analyzed by the image analysis algorithm; 3B – The same image with markup after being analyzed by the image analysis algorithm (red color =3+; orange =2+; yellow =1+; black =nucleus)

estimates have almost one pattern that has most of the values higher than ground truth (Table 4 and Figure 6). Estimated PC% from pathologist 1 has two trends in terms of misclassification including under- and over-estimation. Thus, the Spearman coefficient (0.731; 95% CI [0.574 – 0.836]) from pathologist 1 has the lowest value among all the pathologists (Table 4 and Figures 4

and 9). Overall, Aperio algorithm and the pathologists provided good to excellent estimation of PC% with the algorithm having the best performance (Table 2).

Since PC% playing as one of the main criteria in plasma cell neoplasm classification and diagnosis, values from the PC% estimation in this study were examined as groups based on the 10% and 60% cutoffs (Table 4) to evaluate the impact of misquantification in changing this criterion (Table 1). Also based on Table 1, we can see that under- or over-estimating PC% surrounding the 10% and 60% cutoffs cause shifting in diagnostic criteria. Misquantified PC% between 10% and less than 60% do not have the same significant impact with regard to criteria but do have clinical impact if bone marrow biopsy is repeated after treatment for surveillance due to the need to assess treatment response by comparing pre- and post-treatment PC% in bone marrow. The pathologists were always aware of these cutoffs when making their estimation, similar to recognizing boundaries. In other words, although being able to pick a continuous number of PC% for an image with a wide range from 0 to 100%, they might simultaneously check their estimation with the cutoffs and then adjust the estimation to make it fit with the overall diagnostic impression for the image. This could explain why there is no estimated value of 8, 9 or 58, 59. Therefore, the values are not totally random.

To assess inter-observer agreement on the estimation of PC%, the statistical analysis had two steps. The first one is classifying each pathologist' estimate based on its relation to the cutoffs and ground truth as having a diagnostic impact or not. It was found that Kappa is 0.43 ($p<0.001$), which shows a *moderate* agreement among the pathologists across all 53 images. The second step is labeling or classifying discrepant estimated PC% with ground truth in terms of the 10% and 60% cutoffs as under-estimated, or within an accepted range, or over-estimated (Table 4). These labels were assigned as 1, 2, and 3, respectively. Kappa test was conducted, and a combined Kappa for these three outcomes is 0.162 ($p=0.018$), which means there is a *slight* agreement among the pathologists.

The number of total significant misquantifications of each pathologist and the algorithm was compared using a one-way Chi-square test and shows no difference in these numbers among the pathologists and the algorithm ($p=0.84$) (Table 4). The total percent of absolute difference (a positive number) for each pathologist and the algorithm shows that two of four pathologists have these numbers close to the lowest number that is from the algorithm.

Figure 9 with Bland – Altman plots demonstrates the bias or difference between estimated values and ground truth in relation to the mean of these two values. The plot from the algorithm shows the smallest differences with a consistent trend throughout the 53 images, while it seems like the bias in pathologists' plots are bigger when the PC% is higher.

Table 2. Summary statistics of Leica image analysis algorithm' and pathologists' estimation compared with ground truth

	Algorithm	Pathologist 1	Pathologist 2	Pathologist 3	Pathologist 4
Mean ± SD	13.2 ± 20.3	13.00 ± 23.88	11.79 ± 21.48	18.67 ± 26.04	14.24 ± 24.50
Bias ± SD	-0.27 ± 2.64	-0.06 ± 7.55	1.15 ± 5.67	-5.72 ± 8.42	-1.31 ± 6.72
CCC (95% CI)	0.991 (0.987 - 0.996)	0.940 (0.914 – 0.966)	0.960 (0.940 – 0.980)	0.904 (0.869 – 0.940)	0.952 (0.935 – 0.970)
Spearman Correlation (95% CI)	0.944 (0.905 – 0.968)	0.731 (0.574 – 0.836)	0.937 (0.893 – 0.963)	0.948 (0.910 – 0.970)	0.923 (0.870 – 0.955)
Pearson Correlation (95% CI)	0.992 (0.986 – 0.995)	0.959 (0.930 – 0.976)	0.967 (0.942 – 0.980)	0.973 (0.953 – 0.984)	0.979 (0.964 – 0.988)

Table 3. Misquantified PC% present at least at one column based on cutoffs of 10% and 60%

Image Name	Pathologist 1	Pathologist 2	Pathologist 3	Pathologist 4	Algorithm	Ground truth
4	10	0	1	1	0	**1**
28	1	2	10	5	4	**4**
30	1	4	15	5	3	**4**
31	5	5	10	5	7	**5**
33	5	5	15	7	13	**8**
34	5	5	10	5	5	**8**
35	10	10	25	10	11	**9**
36	1	20	15	2	5	**10**
37	15	5	20	10	11	**10**
38	1	5	15	7	8	**11**
39	0	10	20	7	7	**11**
40	1	5	15	7	8	**12**
45	5	5	20	15	21	**24**
46	20	15	60	30	30	**29**
48	60	60	70	60	57	**56**
49	80	50	70	65	56	**57**
50	90	80	90	90	69	**64**
51	60	70	80	80	69	**65**
52	80	80	90	80	68	**65**
53	70	70	80	90	65	**66**

Table 4. Details of cases with misclassified PC% based on cutoff 10% and 60%

Cutoff	Image Name	Pathologist 1	Pathologist 2	Pathologist 3	Pathologist 4	Algorithm
<10%	4	O (+9)				
	28			O (+6)		
	30			O (+11)		
	31			O (+5)		
	33			O (+7)		O (+5)
	34			O (+2)		
≥10%	36	U (-9)			U (-8)	U (-5)
	37		U (-5)			
	38	U (-10)	U (-6)		U (-4)	U (-3)
	39	U (-10)			U (-4)	U (-4)
	40	U (-11)	U (-7)		U (-5)	U (-4)
	45	U (-19)	U (-19)			
<60%	46			O (+30)		
	48	O (+4)	O (+4)	O (+14)	O (+4)	
	49	O (+23)		O (+13)	O (+8)	
	Total mistakes based on cutoffs ($p = 0.84$; χ^2)	8	5	8	6	5
	Total absolute difference (%) from ground truth	95	41	87	33	21

"**O**": overestimated compared to ground truth; "**U**": underestimated compared to ground truth; empty cells have the estimates falling within the ranges defined by the cutoffs.

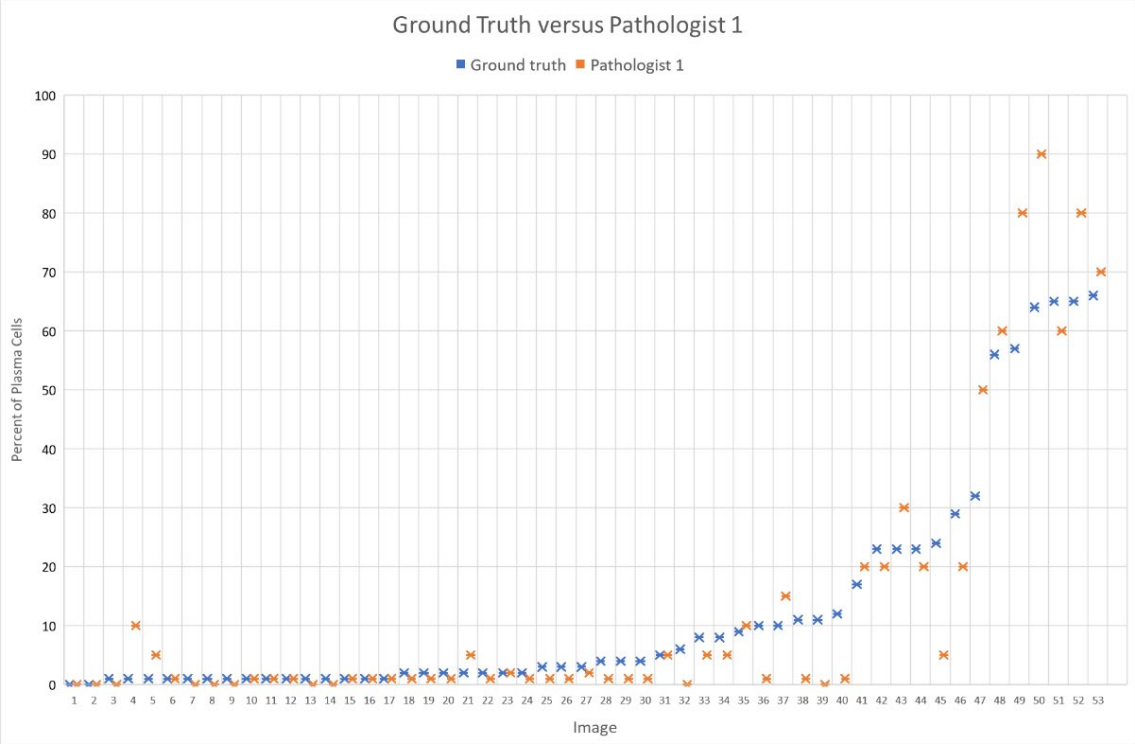


Figure 4. Estimates from pathologist 1 versus ground truth

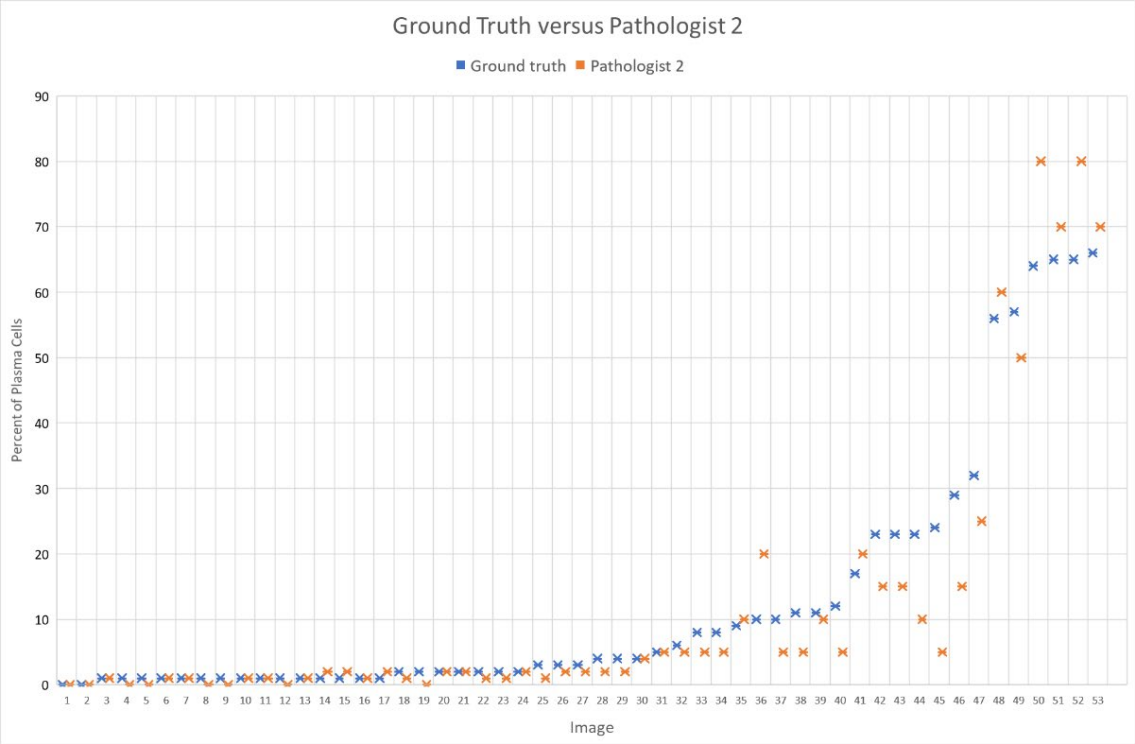


Figure 5. Estimates from pathologist 2 versus ground truth

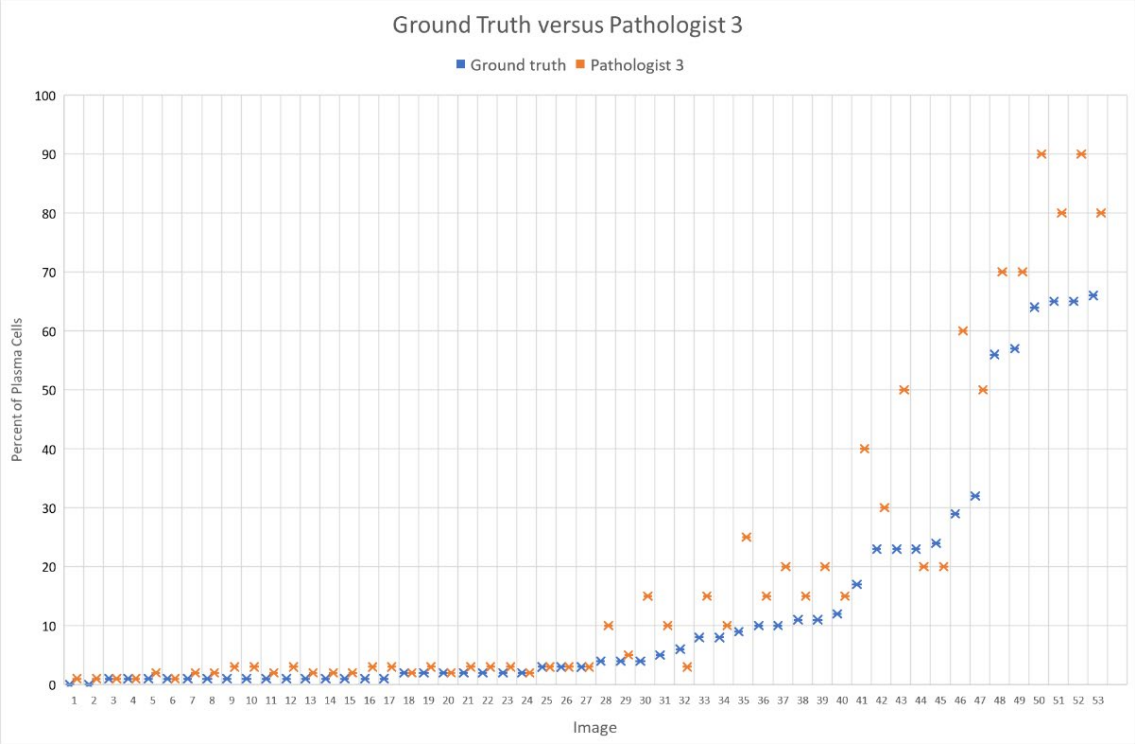


Figure 6. Estimates from pathologist 3 versus ground truth

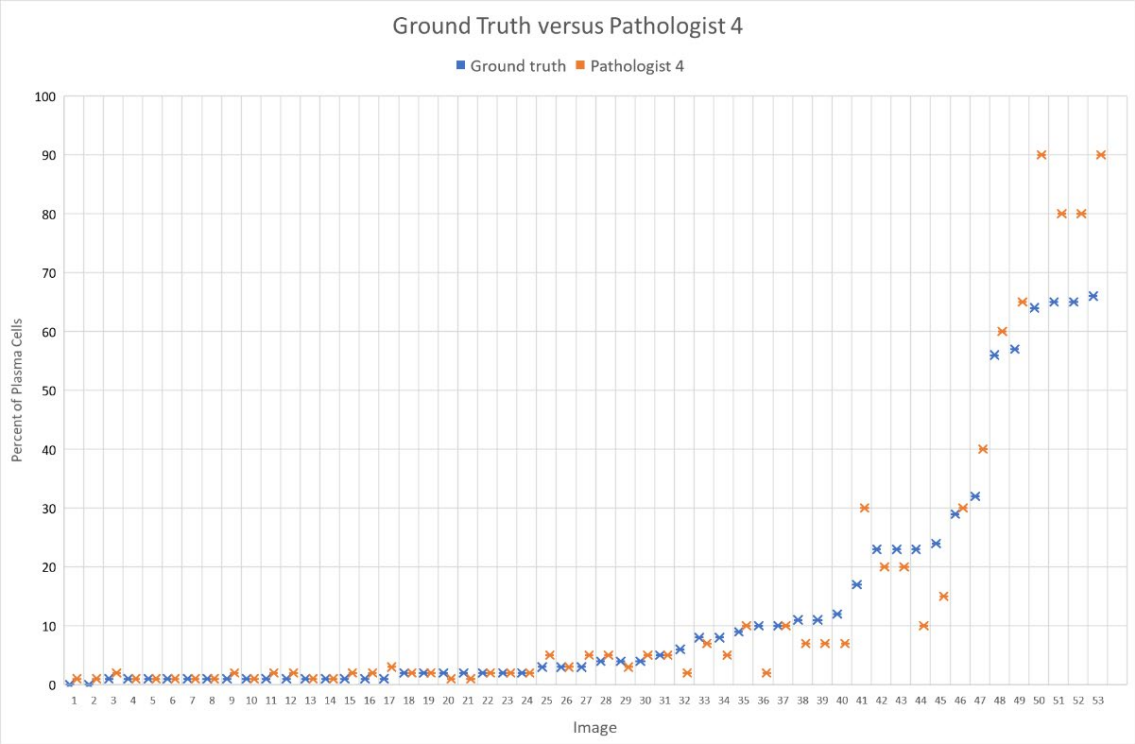


Figure 7. Estimates from pathologist 4 versus ground truth

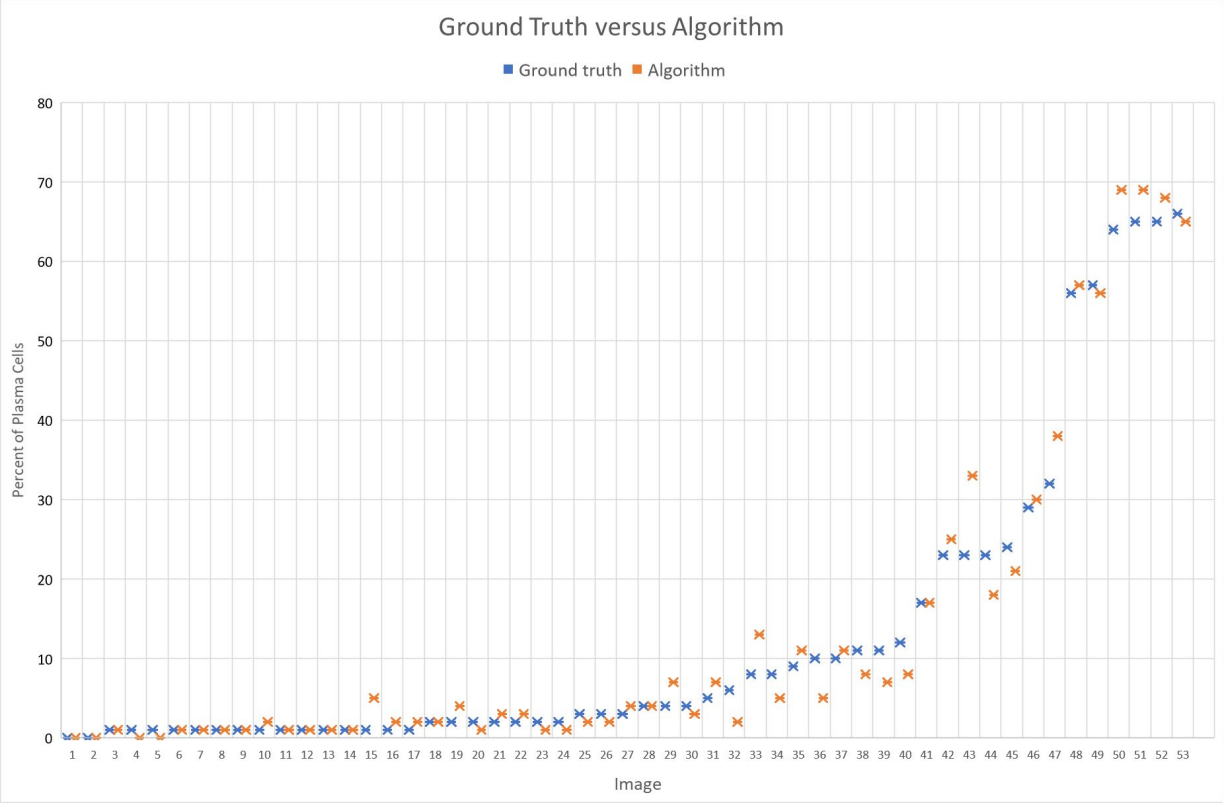
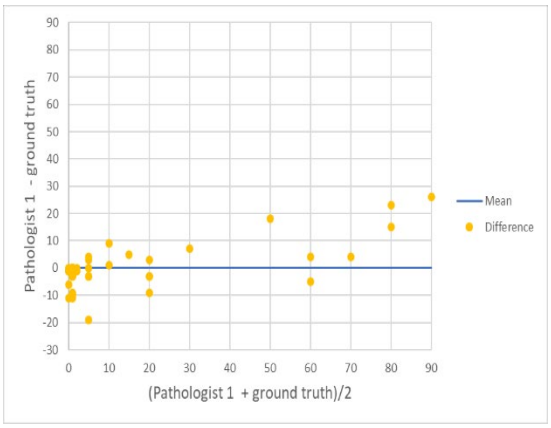
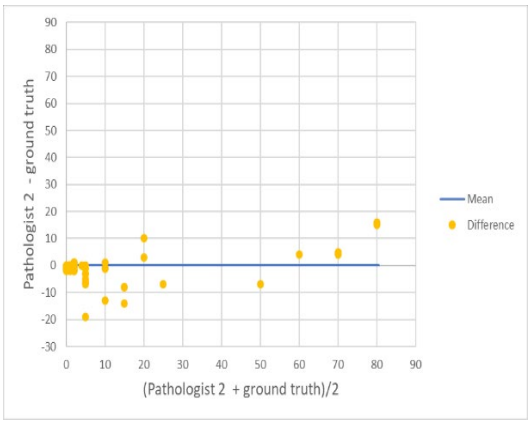


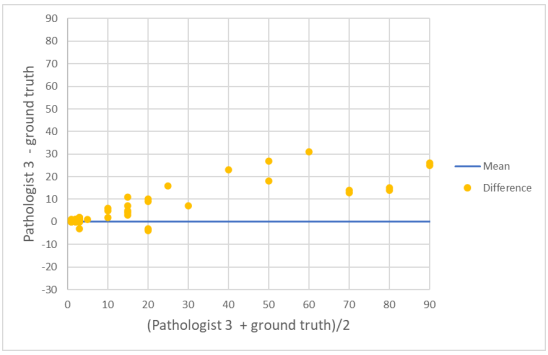
Figure 8. Algorithm's estimates versus ground truth



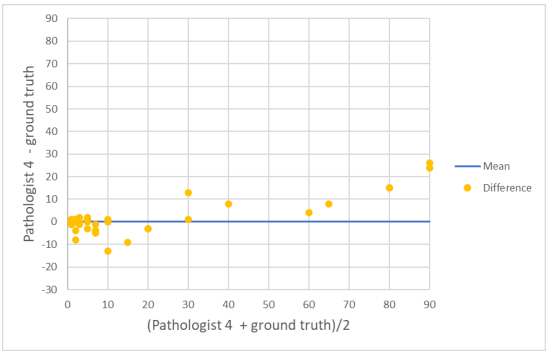
Pathologist 1



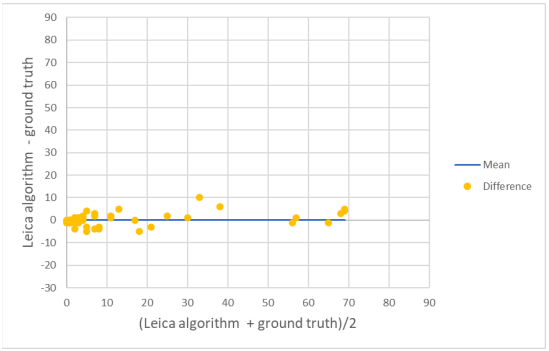
Pathologist 2



Pathologist 3



Pathologist 4



Aperio algorithm

Figure 9. Bland – Altman plots demonstrating bias and variability of the pathologists' estimates of PC%. The algorithm has the lowest bias and variability, while these seem larger with higher PC%.

CHAPTER 5. DISCUSSION

It would be ideal for this study to examine a whole BMBx core to simulate the real practice setting where pathologists give PC% based on the highest PC% between a core biopsy and a bone marrow aspirate[6], with the latter obtained from manual counting and the former obtained from visual estimation. However, without ground truth obtained for the 53 biopsies due to limitation of time and personnel, it would be suboptimal to compare PC% by visual estimation given by pathologists to that by image analysis algorithm in order to decide which method provides more accurate and consistent results. Ultimately, this study aims to validate the membrane algorithm developed by Aperio in quantifying PC% for routine use. Therefore, establishing ground truth is paramount for the comparison and subsequent decision.

Table 5. Summary of plasma cell neoplasm studies with their image analysis characteristics, scale of examination, testing method, and gold standard

Plasma cell neoplasm studies	Image analysis	Scale of examination	Testing method used by pathologists	Gold standard or assessment method
Stifter et al.[9]	Relative area of positive stain	Whole BMBx core	Visual estimation	Survival
Lee N et al.[10]	Not reported	Whole BMBx core	Visual estimation	Survival
Went et al.[11]	Positive pixel count	Whole core of BMBx	Visual estimation	None
Smith et al.[14]	Quantitative nuclear antigen algorithm	Selected image area	Manual counting under microscope using a cell counter	Ground truth (from manual counting)
This study	Membrane cell count algorithm	Selected image area	Visual estimation	Ground truth (from manual counting)

As seen in Table 5, only a few studies[9-11, 14] evaluated the use of image analysis on plasma cell quantification. When whole bone marrow core was examined, testing method had only one option that is visual estimation with using survival outcome as gold standard for the assessment of the image analysis algorithms. However, with a different approach by using selected image area from the core biopsy, ground truth could be achieved. Smith et al.[14] designed their study similar to this present study in terms of using ground truth as a gold standard and choosing selected image area for testing. The difference was that the pathologists in their study used a cell counter under a microscope to quantify plasma cells and then compute PC%. This study aims to compare the accuracy of PC% given by pathologists' estimates and by the algorithm when the ground truth is available as a gold standard. With this approach, this study shows that overall, both pathologists and the algorithm provided high concordance correlation coefficients ($r>0.9$) with the algorithm having the highest coefficient of 0.99. This result is promising for the application of this algorithm to routine practice in terms of accuracy; other factors must be considered such as laboratory personnel to help with slide scanning, ensured slide quality (stain consistency, cleanliness, tissue thickness), stabled image analysis software, software maintenance, pathologist buy-in, pathologist training to use the software, and creating a new workflow.

The other advantage of having ground truth is that identification of the diagnostic impact on each image based on its PC% compared to ground truth gave us a detailed picture of the estimation characteristics of the pathologists and how much agreement they had in all the images and especially on misclassified ones based on the 10% and 60% cutoffs. Three of four pathologists had underestimated PC% that shifted the diagnostic impact on these images when true PC% is less than 10%, and overestimated PC% without diagnostic impact when PC% is close or higher than 60%. One pathologist had a consistent trend of overestimating PC% with diagnostic impact with true PC% of less than 10%, leading to upgrading the diagnostic criterion (Table 4). There were a moderate variability in the diagnostic impact of 53 images among four pathologists (Kappa = 0.42) and a wide variability in the subsample of only impacted images from at least one pathologist (Kappa = 0.16). In general, the total *numbers* of *misclassifications* among the pathologists and the algorithm are not statistically different (Table 4).

With other immunohistochemical biomarkers such as HER2 and Ki-67[13], the value of quantifying these marker percentage aids to cancer prognostication and treatment, while CD138 marker quantification plays not only in diagnosis but also in treatment. Lee et al.[17] studied the prognostic impact of bone marrow PC% assessment before autologous stem cell transplant and found that patients with bone marrow PC% less than 5% coupled with negative serologic complete response had a significant progression-free survival. Therefore, achieving more accurate and consistent PC% in bone marrow biopsies from a patient is extremely important for oncologists to have the right treatment therapy for a multiple myeloma patient.

Although this study shows the advantage of using an image analysis algorithm in quantifying bone marrow plasma cells, it does not imply that manual counting on bone marrow aspirate should be eliminated. The guideline from the International Myeloma Working Group in 2014[6] is continuously followed where the final PC% is the higher of the two values from a biopsy (core or clot) and an aspirate.

The question for this study is that whether we should establish a range for the cutoffs 10% and 60% rather than a discrete number because the reality is that when the results obtained by the algorithm are close to 10% or 60%, for instance, 8%, 9%, 58% or 59%, then whether we should consider they are 10% and 60% or not. How wide these ranges can be accepted is an open question for pathologists who are considering applying this algorithm in practice. When classifying which case was significantly misquantified in Table 4, this study used the raw data. If a range of 8-12% were considered to meet the cutoff 10%, then the algorithm would have only one significantly misquantified/misclassified case rather than five cases out of the 53 cases.

CHAPTER 6. CONCLUSIONS

The Aperio membrane algorithm shows a promising approach to accurate and consistent quantification of CD138-positive plasma cell percentage in bone marrow biopsy specimen from patients with plasma cell neoplasms. Overall, pathologists' visual estimation is as good as that by the algorithm. With the moderate variability in pathologists' assessment of bone marrow PC%, consistency is not well achieved. Thus, applying a computer-assisted image analysis tool into

routine practice for plasma cell neoplasms should be considered to increase diagnosis accuracy and consistency, which may improve patients' treatments and outcomes.

APPENDICES

Appendix A. Raw data of plasma cell percentages obtained from four pathologists, the algorithm, and ground truth

Image Name	Pathologist 1 (%)	Pathologist 2 (%)	Pathologist 3 (%)	Pathologist 4 (%)	Algorithm (%)	Ground truth (%)
1	0	0	1	1	0	0
2	0	0	1	1	0	0
3	0	1	1	2	1	1
4	10	0	1	1	0	1
5	5	0	2	1	0	1
6	1	1	1	1	1	1
7	0	1	2	1	1	1
8	0	0	2	1	1	1
9	0	0	3	2	1	1
10	1	1	3	1	2	1
11	1	1	2	2	1	1
12	1	0	3	2	1	1
13	0	1	2	1	1	1
14	0	2	2	1	1	1
15	1	2	2	2	5	1
16	1	1	3	2	2	1
17	1	2	3	3	2	1
18	1	1	2	2	2	2
19	1	0	3	2	4	2
20	1	2	2	1	1	2
21	5	2	3	1	3	2
22	1	1	3	2	3	2
23	2	1	3	2	1	2
24	1	2	2	2	1	2
25	1	1	3	5	2	3
26	1	2	3	3	2	3
27	2	2	3	5	4	3
28	1	2	10	5	4	4
29	1	2	5	3	7	4
30	1	4	15	5	3	4
31	5	5	10	5	7	5
32	0	5	3	2	2	6

Image Name	Pathologist 1 (%)	Pathologist 2 (%)	Pathologist 3 (%)	Pathologist 4 (%)	Algorithm (%)	Ground truth (%)
33	5	5	15	7	13	8
34	5	5	10	5	5	8
35	10	10	25	10	11	9
36	1	20	15	2	5	10
37	15	5	20	10	11	10
38	1	5	15	7	8	11
39	0	10	20	7	7	11
40	1	5	15	7	8	12
41	20	20	40	30	17	17
42	20	15	30	20	25	23
43	30	15	50	20	33	23
44	20	10	20	10	18	23
45	5	5	20	15	21	24
46	20	15	60	30	30	29
47	50	25	50	40	38	32
48	60	60	70	60	57	56
49	80	50	70	65	56	57
50	90	80	90	90	69	64
51	60	70	80	80	69	65
52	80	80	90	80	68	65
53	70	70	80	90	65	66

Appendix B. Two-tiered classification based on diagnostic impact of estimated plasma cell percentages (1=no impact; 2= with impact)

For example, PC% from pathologist 1 on image #5 from Appendix A is 5%, while ground truth is 1%. This discrepancy does not change the diagnostic impact on this image. Therefore, this case was classified as no impact or 1. In contrast, PC% from pathologist 1 on image #4 is 10%, while ground truth is 1%. This misquantified 10% on this image has a diagnostic impact because it meets the cutoff 10%, which is the main criterion for diagnosis of plasma cell neoplasms. This image was classified as 2.

Image Name	Pathologist 1	Pathologist 2	Pathologist 3	Pathologist 4	Algorithm
1	1	1	1	1	1
2	1	1	1	1	1
3	1	1	1	1	1
4	2	1	1	1	1
5	1	1	1	1	1

Image Name	Pathologist 1	Pathologist 2	Pathologist 3	Pathologist 4	Algorithm
6	1	1	1	1	1
7	1	1	1	1	1
8	1	1	1	1	1
9	1	1	1	1	1
10	1	1	1	1	1
11	1	1	1	1	1
12	1	1	1	1	1
13	1	1	1	1	1
14	1	1	1	1	1
15	1	1	1	1	1
16	1	1	1	1	1
17	1	1	1	1	1
18	1	1	1	1	1
19	1	1	1	1	1
20	1	1	1	1	1
21	1	1	1	1	1
22	1	1	1	1	1
23	1	1	1	1	1
24	1	1	1	1	1
25	1	1	1	1	1
26	1	1	1	1	1
27	1	1	1	1	1
28	1	1	2	1	1
29	1	1	1	1	1
30	1	1	2	1	1
31	1	1	2	1	1
32	1	1	1	1	1
33	2	2	1	2	1
34	2	2	1	2	2
35	1	1	1	1	1
36	2	1	1	2	2
37	1	2	1	1	1
38	2	2	1	2	1
39	2	1	1	2	2
40	2	2	1	2	1
41	1	1	1	1	1
42	1	1	1	1	1
43	1	1	1	1	1
44	1	1	1	1	1

Image Name	Pathologist 1	Pathologist 2	Pathologist 3	Pathologist 4	Algorithm
45	2	2	1	1	1
46	1	1	2	1	1
47	1	1	1	1	1
48	2	2	2	2	1
49	2	1	2	2	1
50	1	1	1	1	1
51	1	1	1	1	1
52	1	1	1	1	1
53	1	1	1	1	1

Appendix C. Three-tiered classification based on estimated plasma cell percentages.

(1=underestimated; 2=within expected range; 3=overestimated)

Cases with misquantified plasma cell percentages above 60% were not included due to not being clinical significance. The values in the below table are related to the 10% cutoff.

Image Name	Pathologist 1	Pathologist 2	Pathologist 3	Pathologist 4
4	3	2	2	2
28	2	2	3	2
30	2	2	3	2
31	2	2	3	2
33	2	2	3	2
36	1	2	1	1
37	2	1	2	2
38	1	1	2	1
39	1	2	2	1
40	1	1	2	1
45	1	1	2	2

GLOSSARY

BMBx	bone marrow biopsy
CCC	concordance coefficient correlation
MGUS	monoclonal gammopathy of undetermined significance
SMM	smoldering multiple myeloma
PCM	plasma cell myeloma
PC%	plasma cell percentage
O	overestimated compared to ground truth
U	underestimated compared to ground truth

REFERENCES

1. Kristinsson SY, Bjorkholm M, Andersson TM, Eloranta S, Dickman PW, Goldin LR, et al. Patterns of survival and causes of death following a diagnosis of monoclonal gammopathy of undetermined significance: a population-based study. Haematologica. 2009;94(12):1714-20.
2. Kyle RA, Larson DR, Therneau TM, Dispenzieri A, Kumar S, Cerhan JR, et al. Long-Term Follow-up of Monoclonal Gammopathy of Undetermined Significance. N Engl J Med. 2018;378(3):241-9.
3. Palumbo A, Avet-Loiseau H, Oliva S, Lokhorst HM, Goldschmidt H, Rosinol L, et al. Revised International Staging System for Multiple Myeloma: A Report From International Myeloma Working Group. J Clin Oncol. 2015;33(26):2863-9.
4. Varettoni M, Corso A, Cocito F, Mangiacavalli S, Pascutto C, Zappasodi P, et al. Changing pattern of presentation in monoclonal gammopathy of undetermined significance: a single-center experience with 1400 patients. Medicine (Baltimore). 2010;89(4):211-6.
5. Berenson JR, Anderson KC, Audell RA, Boccia RV, Coleman M, Dimopoulos MA, et al. Monoclonal gammopathy of undetermined significance: a consensus statement. Br J Haematol. 2010;150(1):28-38.
6. Rajkumar SV, Dimopoulos MA, Palumbo A, Blade J, Merlini G, Mateos MV, et al. International Myeloma Working Group updated criteria for the diagnosis of multiple myeloma. Lancet Oncol. 2014;15(12):e538-48.
7. Matsue K, Matsue Y, Kumata K, Usui Y, Suehara Y, Fukumoto K, et al. Quantification of bone marrow plasma cell infiltration in multiple myeloma: usefulness of bone marrow aspirate clot with CD138 immunohistochemistry. Hematol Oncol. 2017;35(3):323-8.
8. Al-Quran SZ, Yang L, Magill JM, Braylan RC, Douglas-Nikitin VK. Assessment of bone marrow plasma cell infiltrates in multiple myeloma: the added value of CD138 immunohistochemistry. Hum Pathol. 2007;38(12):1779-87.
9. Štifter S, Babarović E, Valković T, Seili-Bekafigo I, Štemberger C, Načinović A, et al. Combined evaluation of bone marrow aspirate and biopsy is superior in the prognosis of multiple myeloma. Diagn Pathol. 2010;5(1):30.
10. Lee N, Moon S, Lee J, Park H, Kong S, Bang SM, et al. Discrepancies between the percentage of plasma cells in bone marrow aspiration and BM biopsy: Impact on the revised IMWG diagnostic criteria of multiple myeloma. Blood Cancer J. 2017;7(2):e530-e.
11. Went P, Mayer S, Oberholzer M, Dirnhofer S. Plasma cell quantification in bone marrow by computer-assisted image analysis. Histol Histopathol. 2006.
12. Koopman T, de Bock GH, Buikema HJ, Smits MM, Louwen M, Hage M, et al. Digital image analysis of HER2 immunohistochemistry in gastric-and oesophageal adenocarcinoma: a validation study on biopsies and surgical specimens. Histopathology. 2018;72(2):191-200.
13. Volynskaya Z, Mete O, Pakbaz S, Al-Ghamdi D, Asa SL. Ki67 quantitative interpretation: Insights using image analysis. J Pathol Inform. 2019;10.
14. Smith FB, Elnawawi A. A counting strategy for estimating plasma cell number in CD138-stained bone marrow core biopsy sections. Ann Clin Lab Sci. 2008;38(2):138-42.
15. Bui MM, Riben MW, Allison KH, Chlipala E, Colasacco C, Kahn AG, et al. Quantitative Image Analysis of Human Epidermal Growth Factor Receptor 2 Immunohistochemistry for Breast Cancer: Guideline From the College of American Pathologists. Arch Pathol Lab Med. 2019;143(10):1180-95.
16. Bankhead P, Loughrey MB, Fernández JA, Dombrowski Y, McArt DG, Dunne PD, et al. QuPath: Open source software for digital pathology image analysis. Sci Rep. 2017;7(1):1-7.
17. Lee S-E, Yoon J-H, Shin S-H, Eom K-S, Kim Y-J, Kim H-J, et al. Bone marrow plasma cell assessment before peripheral blood stem cell mobilization in patients with multiple myeloma undergoing autologous stem cell transplantation. BioMed research international. 2014;2014.