Leukocyte Cell Population Data for Hematology Analyzer-Based Distinction of Clonal-versus-Non-Clonal Lymphocytosis: A Real-World Testing Experience

Pulkit Rastogi¹ · Prashant Sharma¹ · Neelam Varma¹ · Dmitry Sukhachev² · Naveen Kaushal³ · Ishwar Bihana¹ · Man Updesh Singh Sachdeva¹ · Shano Naseem¹ · Pankaj Malhotra⁴

Received: 24 June 2017 / Accepted: 10 January 2018 / Published online: 20 January 2018
© Indian Society of Hematology and Blood Transfusion 2018

Abstract Automated blood counts revealing lymphocytosis necessitate smear reviews. Even expert morphological evaluation may however, fail to differentiate a benign-versus-malignant etiology without further testing. Automated analyser-derived quantitative data on leukocyte cell populations remain undertested for distinguishing such etiologies. Instrument manufacturers claim that if successful, they may be used to generate software flags that help under-resourced laboratories better triage hemogram specimens requiring further testing. We tested the diagnostic accuracy of volume-conductivity-scatter (VCS) indices together with complete blood count (CBC) parameters in such scenarios. We compared LH780-derived (Beckman Coulter, FL, USA) CBC + VCS parameters from patients with clonal lymphoproliferations ($n = 42$, including 30 chronic lymphocytic leukemia cases) versus 83 controls with absolute or relative lymphocytosis (derivation cohort). Diagnostic performances of

11 logistic regression equations derived were subsequently evaluated on two specific validation cohorts ($n = 130$ and $n = 1465$). Clonal lymphocytoses showed significantly lower hemoglobin and higher leukocyte counts but similar lymphocyte percentages (LY %) vis-à-vis controls. The most significant, albeit overlapping predictor of clonality was the absolute lymphocyte count, LY# ($47.8 \pm 48.4 \times 10^9/L$ vs. $2.9 \pm 1.4 \times 10^9/L$ in clonal vs. benign cases). In eleven logistic regression equations constructed using four combinatorial approaches, only the models with LY# (highest sensitivity/specificity of 99.3%/100%) and the lymphocytic VCS parameters alone (highest sensitivity/specificity of 76.2%/90.2%) performed consistently in both validation cohorts. Lymphocytic VCS parameters were moderately successful in distinguishing benign-versus-malignant lymphocytes. Other approaches of CBC-plus-VCS parameters did not sustain their initial excellent performances in the validation cohorts, highlighting a need for careful appraisal and better standardization of automated cellular analysis technologies.

Prior presentation: A part of this research work was presented as a poster at ISLH 2016, the XXIX International Symposium on Technological Innovations in Laboratory Hematology at Milano, Italy from May 12–14, 2016 (Berend Houwen Travel Award).

✉ Prashant Sharma
sharma.prashant@pgIMER.edu.in

¹ Department of Hematology, Postgraduate Institute of Medical Education and Research, Level 5, Research Block A, Sector 12, Chandigarh 160012, India

² LabTech Manpower Ltd, Saint Petersburg, Russia

³ Department of Biophysics, Panjab University, Chandigarh, India

⁴ Adult Clinical Hematology Unit, Department of Internal Medicine, Postgraduate Institute of Medical Education and Research, Chandigarh, India

Keywords Automated hematology analysers · Leukocyte cell population data · Logistic regression equations · Lymphocytosis · Lymphoproliferative disorders · Volume-conductivity-scatter

Introduction

Routinely flagged peripheral blood lymphocytosis on an automated hematology analyzer (HA) poses a significant diagnostic challenge to clinical laboratories in terms of distinguishing reactive or benign from clonal or malignant etiologies. Such flagged specimens require blood smear review followed by correlation with the clinical

background and further sophisticated, time-consuming and expensive investigations like flow cytometry, serological studies and/or various molecular tests.

Marked technological advancements enable modern HAs to better characterize various circulating blood cell populations. Many new research parameters, are now available, often with the routine analyses at no additional cost [1]. These results may be utilized for flagging as well as to derive algorithmic approaches for identifying various disease states [2, 3].

The volume, conductivity and scatter (VCS) indices obtained from Beckman Coulter instruments have, in the past, been shown to be potentially useful in differentiating the various causes of lymphoproliferations [4–6]. Prior studies have proposed certain cut-offs as well as diagnostic algorithms using lymphocytic VCS parameters to characterize such proliferations. However, they had relatively low diagnostic accuracies with significant overlaps between benign and malignant lymphoid expansions [4–6].

Lymphoproliferations are often accompanied by derangements in blood counts due to effects on other blood cell lineages [7]. In a previous study, we found changes in VCS indices of other cells too (apart from lymphocytes) in such cases [8]. To further evaluate this aspect, we planned this study to check the combined ability of various VCS indices in combination with complete blood count (CBC) data to discriminate various clonal versus benign causes of lymphocytosis. The ultimate aim was to generate regression equations which on incorporation into the laboratory information system (LIS) could help screen and classify cases of lymphocytosis.

Materials and Methods

This laboratory-based prospective case–control study was conducted in two phases. A study group enrolled between January and September, 2015 (phase 1) served as a “derivation cohort”. In phase 2, two “validation cohorts” were enrolled: the first one consisted of 130 healthy adult controls whose blood samples were run parallel to the derivation cohort on the same HA instrument. The second validation cohort comprised of 1465 consecutive blood samples of the patients visiting the hospital in the month of August 2017 (approximately 2 years after Phase 1). These were tested on the same HA so as to study the time and machine-dependent stability of the various parameters. All the samples were of the age group ≥ 18 years. The study was conducted after the approval by Institute’s Ethics Committee. The study protocol was in compliance with the Declaration of Helsinki and the Statement for Reporting Studies of Diagnostic Accuracy (STARD) recommendations [9].

Phase 1 (Derivation Cohort)

Group 1: Patients with a Clonal Lymphocytosis (n = 42)

These comprised 30 consecutive untreated cases of chronic lymphocytic leukemia (CLL) diagnosed on the basis of International Workshop on CLL (IW-CLL) recommendations [10] of absolute lymphocyte count and immunophenotypic analyses along with 12 consecutive untreated cases of non-CLL clonal lymphoproliferative disorders with peripheral blood involvement that were diagnosed as per the WHO classification, 2008 [11].

Flow cytometric immunophenotyping was performed on a BD FACSCanto II instrument (Becton–Dickinson Biosciences, San Jose, USA). The antibody panel used included CD19, CD20, CD79b, CD10, CD23, CD43, FMC-7, surface kappa and lambda immunoglobulin light chains, CD200, CD38, CD11c, CD25, CD103, CD123, CD3, CD4, CD5 and CD8. Further antibodies were tested as and when indicated by the results of this primary panel.

Group 2: Control Group of Patients with Benign Non-clonal Lymphocytosis (n = 83)

These included 83 consecutive cases of absolute or relative lymphocytosis of varied but proven non-clonal etiology including marrow aplasia ($n = 30$), and viral infections [HIV-positive patients on highly active anti-retroviral therapy ($n = 50$), EBV infection ($n = 2$) and hepatitis A ($n = 1$)] that had been confirmed by the appropriate serological, immunophenotypic or molecular studies in conjunction with the appropriate clinical background.

All 125 blood samples anticoagulated in dipotassium EDTA (1.5 mg per ml of blood) underwent CBC and differential counts on one of two LH780 HAs (Beckman Coulter Inc., Miami, FL) within 6 h of collection. Rigorous quality control was exercised including VCS optimization and regular 5-CTM and LatronTM controls.

Nineteen CBC parameters [hemoglobin, total leukocyte count (TLC), hematocrit, platelet count, RBC indices (MVC, MCH, MCHC, RDW), mean platelet volume (MPV) and percentages as well as absolute counts of leukocytes (NE %, LY %, MO %, EO %, BA %, NE#, LY#, MO#, EO#, BA#)] were recorded. The eighteen VCS parameters comprised means of the volume, conductivity and scatter of neutrophils (MNV, MNC, MNS), lymphocytes (MLV, MLC, MLS) and monocytes (MMV, MMC, MMS) and their respective standard deviations (SD) i.e. MNV-SD, MNC-SD, MNS-SD, MLV-SD, MLC-SD, MLS-SD, MMV-SD, MMC-SD and MMS-SD. Eosinophil VCS indices were not retrievable in a database format, and were also not given by the HA in a significant number of cases. Hence, they were excluded from analysis.

Phase 2 (Validation Cohorts)

Validation cohort 1 (healthy controls, $n = 130$) These comprised of blood samples from healthy controls whose CBC and related routine indices were within the normal range established for the laboratory. All samples were run on the same HA and within same time period as for the derivation cohort (January to September 2015).

Validation cohort 2 (consecutive routine blood samples, $n = 1465$) These comprised of consecutive blood samples from out-patients visiting this tertiary level multispeciality hospital for various ailments. The timing for this validation cohort 2 was post-2 years of the initial healthy control set. These samples were run on the same HA as that for the derivation cohort. This cohort was tested in order to assess the correct predictive power of the various models (discussed later) obtained from the derivation cohort with respect to the instrument-specific stability of parameters (VCS and others) over time. Clinical and laboratory information for this cohort 2 was exhaustively sought for the cases indicated as clonal lymphocytosis by high-performing models as well as for all cases with $LY\# > 5 \times 10^9/L$.

Statistical Analysis

Initially, the 19 CBC and 18 VCS parameters were compared between the two groups of the derivation cohort using the Mann–Whitney *U*-test. Since most individual CBC or VCS parameters were not powerfully discriminatory between the two groups, combinatorial approaches were used to construct LIS-compatible logistic regression equations/models with binary outcomes i.e. 0 or 1 corresponding to benign or clonal etiologies [12]. Different strategies were adopted to obtain these models—forward, backward and manual stepwise regression procedures. We attempted to combine the largest possible variety of parameters into the best fitting models. As the fitting characteristics of several of the computed models were rather high (i.e. overfitting), we focused on the percentage of correctly classified cases rather than ROC curves in selecting best models. The models' optimal cut-offs were derived using receiver operating characteristics (ROC) curves [12].

The various regression equations (models) obtained were subsequently tested on both of the validation cohorts and percentages of correctly classified cases were obtained. All analyses were done using GraphPad PrismTM 6 (Version 6.07 for Windows, La Jolla CA, USA) and MedCalc Statistical Software (Version 16.1, bvba, Ostend, Belgium).

Results

Derivation Cohort Results

The patients in group 1 comprised of 30 CLL, 4 mantle cell lymphomas, 3 splenic marginal zone lymphomas, 2 hairy cell leukemias, 2 prolymphocytic leukemias and 1 follicular lymphoma. No clonal T cell lymphoproliferations were included. For group 2, flow cytometric immunophenotyping was required in 4 control cases where the distinction between clonal and benign lymphocytosis could not be made despite expert morphological evaluation and assessment of the clinical data.

Comparison of basic hematological and VCS data These are listed in Table 1. Patients with clonal lymphocytosis had significantly lower Hb and higher total leukocyte counts as compared to benign controls. The percentage of lymphocytes in the differential count was similar in the two groups, while the absolute neutrophil count in group 2 was lower, reflecting the influence of cases with relative lymphocytosis due to neutropenia. The most significant predictor of clonal nature was the absolute lymphocyte count ($LY\#$ in $10^9/L$, 47.8 ± 48.4 in clonal versus 2.9 ± 1.4 in benign cases). However, this too at the lower end of its range, overlapped with the benign cases, with $LY\#$ range being 1.0–181.9 in the clonal and 0.3–8.5 in the benign group ($LY\#$ between 1.0 and 8.5, being indeterminate). 21 cases in group 2 had $LY\# > 4.0 \times 10^9/L$ (range 4.0 – $6.43 \times 10^9/L$).

Combinatorial approaches to construct LIS-compatible binary logistic regression equations The various models created are listed in Table 2, and their performance characteristics are compared in Table 3.

An initial set of four models using all available CBC and VCS parameters was generated. The most complex of these models utilized nine CBC parameters whereas the simplest one used only the $LY\#$ (Table 2, approach 1). The percentage of cases correctly classified ranged from 93.6 to 98.4% (Table 3).

In the next step, models excluding WBC count, $LY\%$ and $LY\#$ were created (Table 2, approach 2). This was done to assess the ability of VCS + CBC indices to pick up clonal lymphocytosis cases even if the TLC and $LY\#$ were not as markedly raised as in our cohorts. The percentages of cases correctly classified by the 3 models generated by this approach were comparable with the first set of models (Table 3).

However, all of these second set of models utilized both the $NE\%$ and the $NE\#$ to distinguish clonal from benign cases. Since 30 out of the 83 group 2 cases had bone marrow aplasia, we next also excluded these two neutrophil indices (in addition to the WBC count, $LY\%$ and $LY\#$) to

Table 1 The CBC and VCS parameters incorporated in the regression equations

Parameter	Group 1 (n = 42)	Group 2 (n = 83)	p value*
Age (years), mean $\pm$ SD (range)	59.3 $\pm$ 30.4 (17–86)	–	–
Hematological parameters			
Hemoglobin (gm %)	10.1 $\pm$ 2.7 (3.3–14.3), 10.8	11.6 $\pm$ 3.3 (3.5–16.6), 12.2	0.0067
Hematocrit (%)	31.7 $\pm$ 8.2, (11.4–44.3), 34.9	36 $\pm$ 10.5 (10.4–51.0), 39.1	0.0082
Total leukocyte count ($\times 10^9/L$)	59.1 $\pm$ 60.3 (2.4–248.1), 34.855	6.8 $\pm$ 4.2 (0.5–18.2), 6.8	< 0.0001
Platelet count ($\times 10^9/L$)	154.1 $\pm$ 90.6 (18.5–344), 141.9	150 $\pm$ 115.7 (7.3–434.1), 136.0	0.5018
Mean corpuscular volume, MCV (fL)	88 $\pm$ 10.4 (58–118.2), 87.8	100.7 $\pm$ 13.0 (73.4–127.7), 96.7	< 0.0001
Mean corpuscular hemoglobin, MCH (pg)	28.0 $\pm$ 3.3 (17.8–34.1), 28.1	32.6 $\pm$ 4.3 (23.0–43.9), 31.5	< 0.0001
Mean corpuscular hemoglobin concentration, MCHC (gm %)	31.8 $\pm$ 1.3 (28.3–34.3), 32.0	32.4 $\pm$ 1.0 (29.7–34.5), 32.5	0.0200
Red cell distribution width, RDW-CV (%)	17.9 $\pm$ 3.4 (13.2–26.5), 16.6	16.6 $\pm$ 3.3 (12.3–25.5), 15.4	0.0079
Neutrophil %, NE %	19.2 $\pm$ 16.6 (2.5–73.9), 17.5	39.1 $\pm$ 14.4 (9.5–68.3), 42.1	< 0.0001
Lymphocyte %, LY %	70.4 $\pm$ 19.2 (20.6–96.1), 74.3	49.7 $\pm$ 16.2 (22.1–87.7), 46.6	< 0.0001
Monocyte %, MO %	8.5 $\pm$ 9.9 (0.6–48.1), 5.4	7.0 $\pm$ 3.0 (2.0–24.3), 7.0	0.2263
Eosinophil %, EO %	1.1 $\pm$ 2.2 (0.0–13.8), 0.5	3.6 $\pm$ 6.3 (0.0–49.1), 1.7	< 0.0001
Basophil %, BA %	0.7 $\pm$ 1.2 (0.0–4.6), 0.3	0.6 $\pm$ 0.8 (0.0–7.0), 0.5	0.0116
Absolute neutrophil count (NE#, $\times 10^9/L$)	6.6 $\pm$ 6.5 (0.2–32), 4.9	3.0 $\pm$ 2.4 (0.1–10.4), 2.8	< 0.0001
Absolute lymphocyte count (LY#, $\times 10^9/L$)	47.8 $\pm$ 48.4 (1.0–181.9), 27.4	2.9 $\pm$ 1.4 (0.3–8.5), 3.1	< 0.0001
Absolute monocyte count (MO#, $\times 10^9/L$)	5.7 $\pm$ 13.1 (0.1–83.5), 1.8	0.5 $\pm$ 0.3 (0.0–1.4), 0.5	< 0.0001
Absolute eosinophil count (EO#, $\times 10^9/L$)	0.4 $\pm$ 1.3 (0.0–8.7), 0.1	0.4 $\pm$ 1 (0.0–8.9), 0.1	0.8854
Absolute basophil count (BA#, $\times 10^9/L$)	0.4 $\pm$ 0.8 (0.0–4.4), 0.1	0.05 $\pm$ 0.09 (0.0–0.71), 0.03	< 0.0001
Mean platelet volume, MPV (fl)	9.4 $\pm$ 1.8 (6.6–14.6), 8.9	10.0 $\pm$ 1.7 (6.7–15.2), 9.7	0.0561
VCS Parameters			
MNV (Mean neutrophil volume)	164 $\pm$ 12.1 (143.9–206.8), 162.8	158.4 $\pm$ 14.2 (134.1–195.6), 156.3	0.0169
MNC (Mean neutrophil conductivity)	144.7 $\pm$ 12.5 (128.3–183.1), 142.8	145.6 $\pm$ 9.0 (122.5–166.6), 147.2	0.1448
MNS (Mean neutrophil scatter)	136.6 $\pm$ 9.0 (108.4–149.3), 138.6	135.5 $\pm$ 9.0 (92.2–152.4), 136.7	0.2145
MNV-SD (Mean neutrophil volume-SD)	33.3 $\pm$ 8.0 (20.1–51.6), 33.3	26.2 $\pm$ 8.3 (17.8–67.5), 23.2	< 0.0001
MNC-SD (Mean neutrophil conductivity-SD)	9.6 $\pm$ 3.1 (5.5–20.8), 9.1	9.3 $\pm$ 4.0 (4.7–23.8), 7.4	0.1257
MNS-SD (Mean neutrophil scatter-SD)	13.9 $\pm$ 2.6 (9.9–19.7), 13.2	13.9 $\pm$ 4.0 (8.9–35.9), 12.7	0.4659
MLV (Mean lymphocyte volume)	82.7 $\pm$ 15.0 (56.0–117.9), 84.3	84.4 $\pm$ 9.0 (58.0–103.8), 85.0	0.6213
MLC (Mean lymphocyte conductivity)	115.8 $\pm$ 13.5 (101.5–163.4), 114.5	116.2 $\pm$ 12.1 (98.5–179.5), 115.6	0.3156
MLS (Mean lymphocyte scatter)	62.3 $\pm$ 9.3 (45.0–76.1), 63.3	62.4 $\pm$ 8.6 (43.5–84.6), 61.6	0.7498
MLV-SD (Mean lymphocyte volume-SD)	20.8 $\pm$ 4.2 (11.3–33.1), 21.0	16.4 $\pm$ 3.9 (11.3–28.2), 14.9	< 0.0001
MLC-SD (Mean lymphocyte conductivity-SD)	15.5 $\pm$ 5.3 (6.4–32.0), 15.5	15.1 $\pm$ 7.6 (6.7–50.6), 13.0	0.0751
MLS-SD (Mean lymphocyte scatter-SD)	17.7 $\pm$ 3.0 (11.8–27.4), 17.8	19.1 $\pm$ 3.7 (13.3–33.0), 18.6	0.1018
MMV (Mean monocyte volume)	178.6 $\pm$ 14.6 (152.0–221.6), 176.6	175.2 $\pm$ 14.0 (153.3–226.6), 171.8	0.0804
MMC (Mean monocyte conductivity)	121.1 $\pm$ 10.2 (104.5–152.6), 119.6	122.3 $\pm$ 9.0 (96.1–148.9), 121.8	0.2135

Table 1 continued

Parameter	Group 1 (n = 42)	Group 2 (n = 83)	p value*
MMS (Mean monocyte scatter)	85.4 ± 5.9 (72.8–101.8), 85.0	83.7 ± 6.3 (66.3–98.3), 83.8	0.0949
MMV-SD (Mean monocyte volume-SD)	24.8 ± 4.6 (12.9–36.5), 24.5	22.8 ± 4.6 (16.9–33.9), 21.1	0.0016
MMC-SD (Mean monocyte conductivity-SD)	7.7 ± 3.2 (4.5–19.1), 6.9	6.7 ± 2.9 (3.6–16.4), 5.6	0.0107
MMS-SD (Mean monocyte scatter-SD)	11.6 ± 1.7 (7.0–14.9), 11.5	11.2 ± 1.5 (8.0–17.2), 11.0	0.1582

p-values in bold italics were statistically significant

All values are in mean ± SD (range), median

SD standard deviation

*Mann–Whitney U test for significant differences between multiple groups. Age is not available for group 2. However, all subjects (patients or controls) in this study are older than 13 years

check if a sufficiently discriminatory algorithm could still be generated even in the absence of these parameters (Table 2, approach 3). With this approach too, the models attained 92.9% sensitivity with 100% specificity (Table 3).

And finally, in order to correlate with previous studies that relied exclusively on VCS indices with performance inferior to our VCS + CBC approach, two models were constructed based solely on the VCS indices. One model comprised of neutrophil, lymphocyte and monocyte parameters while the other one was constrained only to the lymphocyte parameters (Table 2, approach 4). These two models were less accurate compared to previous sets misclassifying 14 (11%) and 23 (18%) cases respectively (Table 3).

Validation Cohort Results

Validation cohort 1 Since, this set comprised of only healthy controls, the predictions of the various models when applied to these samples was reflective of percentage of cases correctly classified as benign (or specificity for malignancy). As shown in Table 4, there was a marked reduction in the percentages of correctly classified cases for model 1–1, 1–2, 3–1 and 3–2. Specificity was similar or better for the other models as compared to their respective performances in the derivation cohort.

Validation cohort 2 Among these 1465 consecutive routine samples run 2 years later, evaluation of hospital and laboratory records revealed 5 cases with hematological malignancies: two CLL and three chronic myeloid leukemia in chronic phase (CML-CP). An additional seven non-clonal cases with absolute lymphocytosis (LY# range $5.04\text{--}7.26 \times 10^9/\text{L}$) were found. Most models showed drastic reductions in specificities (i.e. were unstable over time, Table 4). The exceptions were models 1–4, 2–3, 4–1 and 4–2 with specificities > 80% even after 2 years. Of these, models 4–1 and 4–2 correctly identified six out of seven cases of non-clonal lymphocytosis.

Discussion

Despite the ready availability of automated WBC analysis worldwide, very few studies have explored their applications in the diagnosis of lymphoproliferative disorders. Prior papers attempted to distinguish between benign and malignant lymphocytosis based on the lymphocytes' VCS parameters alone. Silva et al. [4] reported certain trends as well as cut-off values for lymphocyte volume, conductivity and their SD that aided differentiation of CLL from healthy controls. Algorithmic approaches derived from their cut-off values achieved 70–78% sensitivity and 62–100% specificity [4].

Table 2 Logistic regression models/equations generated using various combinations of parameters (rationale explained in the Results section)**Approach 1: Models using all parameters available****Model 1–1**

$$\text{Activity} = -120.05 - 1.61*[\text{MCH}] + 3.11*[\text{MCHC}] + 1.18*[\text{MPV}] + 0.35*[\text{MO \%}] - 2.09*[\text{BA \%}] + 1.06*[\text{LY\#}] + 0.374*[\text{MNS}] + 0.72*[\text{MNV-SD}] - 1.86*[\text{MNS-SD}]$$

Model 1–2

$$\text{Activity} = -32.85 - 0.83*[\text{MCH}] + 0.97*[\text{MCHC}] + 0.35*[\text{MPV}] + 0.14*[\text{MO \%}] + 0.52*[\text{LY \#}] + 0.126*[\text{MNS}]$$

Model 1–3

$$\text{Activity} = 14.51 - 0.65*[\text{MCH}] + 0.15*[\text{MO \%}] + 0.49*[\text{LY \#}]$$

Model 1–4

$$\text{Activity} = [\text{LY \#}]$$

Approach 2: Models excluding TLC. LY % AND LY#**Model 2–1**

$$\text{Activity} = -59.24 - 0.78*[\text{MCH}] + 1.13*[\text{MCHC}] + 0.38*[\text{RDW}] - 0.18*[\text{NE \%}] + 1.57*[\text{NE \#}] + 0.163*[\text{MNS}] + 0.071*[\text{MMOV}] + 0.044*[\text{MLV}]$$

Model 2–2

$$\text{Activity} = -34 - 0.73*[\text{MCH}] + 1.11*[\text{MCHC}] + 0.49*[\text{RDW}] - 0.16*[\text{NE \%}] + 1.5*[\text{NE \#}] + 0.056*[\text{MMOV}]$$

Model 2–3

$$\text{Activity} = 15.9 - 0.52*[\text{MCH}] - 0.12*[\text{NE \%}] + 0.74*[\text{NE \#}]$$

Approach 3: Models excluding ne % in addition to TLC. LY % and LY#**Model 3–1**

$$\text{Activity} = -52.1 - 0.76*[\text{MCH}] + 1.69*[\text{MCHC}] - 2.66*[\text{BA \%}] + 42.14*[\text{BA\#}] + 0.153*[\text{MNS}] + 0.648*[\text{MNV-SD}] - 1.07*[\text{MNS-SD}] - 0.74*[\text{MNC-SD}]$$

Model 3–2

$$\text{Activity} = -26.6 - 0.64*[\text{MCH}] + 1.44*[\text{MCHC}] - 2.71*[\text{BA \%}] + 32.82*[\text{BA\#}] + 0.23*[\text{MNV-SD}] - 0.69*[\text{MNS-SD}]$$

Approach 4: Models using VCS parameters only**Model 4–1**

$$\text{Activity} = -12.15 + 0.10*[\text{MNS}] + 0.44*[\text{MNV-SD}] - 1.15*[\text{MNC-SD}] - 0.12*[\text{MMOS}] + 0.23*[\text{MMOV-SD}] - 0.71*[\text{MMOS-SD}] + 0.44*[\text{MLV-SD}]$$

Model 4–2

$$\text{Activity} = -2.63 + 0.46*[\text{MLV-SD}] - 0.26*[\text{MLS-SD}] - 0.11*[\text{MLC-SD}]$$

Haschke-Becher et al. [5] attempted to differentiate CLL from healthy controls based on MLV-SD and MLC-SD to attain sensitivity and specificity of 88.6 and 84.4% respectively. Another study tested a similar approach using lymphocytic VCS parameters to distinguish pathological cases of lymphocytoses from healthy controls with variable sensitivity (62–85%) and specificity (72–99%) [6]. Recently, the Sysmex XN series HA was shown to detect monoclonal T-granular lymphocytic expansions with 67.5% sensitivity and 98.2% specificity, and also distinguish viral mononucleosis syndromes from other diagnostic groups with 63.6% sensitivity and 97.5% specificity [13]. It is of course true that the differential counts from a counter may not be perfect as atypical lymphocytes may get counted as monocytes and basophils too, in some counters, fall in an overlap area. However, the aim of such studies, including this one, is to evaluate the automated parameters, with a view to reduce human intervention to a minimum.

The current study aimed to explore the enhanced efficacy of all the available VCS parameters, i.e. from lymphocytes, monocytes and neutrophils as well as complete CBC data so as to predict the nature of clonal *versus* non-clonal lymphoproliferations. This approach seemed promising based on our previous data [8], although our sample size of the oncology cohort is admittedly small, and the results might have been specific to CLL, aplastic anemia and HIV patients, rather than being generalizable across all lymphocytoses. We also tested only adult samples and not pediatric ones for logistic reasons, although the latter too would have been interesting since atypical lymphocytes are frequently seen in young children. Another shortcoming was that only 21 cases in group 2 had $\text{LY\#} > 4.0 \times 10^9/\text{L}$. A more informative cohort would have been patients with reactive lymphocytosis with TLC reaching CLL levels, but in practice such patients are much less commonly encountered.

Table 3 Performance characteristics of the various logistic regression models generated using four different approaches (details in Results section)

Model no.	AUC	Cut-off	Sensitivity	Specificity	% Cases classified correctly	No. of cases with erroneous diagnosis	Actual Diagnosis of erroneous cases
Models with all descriptors included							
1–1	0.993	– 0.7	95.2	100	98.4	2	CLL (1), HCL (1)
1–2	0.979	– 0.4	92.9	100	97.6	3	CLL (1), HCL (1), SMZL (1)
1–3	0.976	– 1.122	95.2	95.2	95	5	CLL (1), SMZL (1), HAV (1), AA (2)
1–4	0.89	4.98	85.7	97.6	93.6	8	HCL (2), SMZL (1), PLL (1), MCL (1), FL (1), HAV (1), HIV (1)
Models without WBC, LY %, LY#							
2–1	0.975	– 0.28	95.2	100	98.4	2	CLL (1), SMZL (1)
2–2	0.970	– 0.6	92.8	97.6	96	5	CLL (1), SMZL (1), MCL (1), AA (1), HIV (1)
2–3	0.948	0	85.7	95.2	92	10	CLL (2), SMZL (1), MCL (1), HCL (1), PLL (1), HAV (1), HIV (1), AA (2)
Models without WBC, LY %, LY#, NE %							
3–1	0.979	0.2	92.9	100	97.6	3	HCL (2), CLL (1)
3–2	0.960	– 0.4	90.5	96.4	94.4	7	HCL (2), CLL (2), AA (2), HIV (1)
Models with VCS parameters only							
4–1	0.938	– 0.7	92.9	86.7	89	14	MCL (1), CLL (2), AA (11)
4–2	0.844	– 0.56	76.2	85.5	82	23	SMZL (1), PLL (1), MCL (2), FL (1), HCL (1), CLL (4), AA (11), HIV (2)

AA aplastic anemia, *FL* follicular lymphoma, *HAV* hepatitis A virus infection, *HCL* hairy cell leukemia, *HIV* human immunodeficiency virus, *MCL* mantle cell lymphoma, *PLL* prolymphocytic leukemia, *SMZL* splenic marginal zone lymphoma

Table 4 Performance characteristics of validation cohorts

Model no.	Derivation cohort specificity (n = 125), %	Validation set 1 (n = 130), specificity, %	Validation set 2 (n = 1465), specificity, %
1–1	100	44	33.6
1–2	100	75	54.8
1–3	95.2	94	58
1–4	97.6	100	99.3
2–1	100	95	69
2–2	97.6	100	78.1
2–3	95.2	100	84
3–1	100	35	35.1
3–2	96.4	55	45.7
3–3	100	93	57.3
4–1	86.7	76	86
4–2	85.5	87	90.2

Sensitivity is not calculated for validation cohort 1 due to all 130 samples being normal, and for validation cohort 2 since the numbers involved were too large to investigate for subtle lymphoproliferations. Only cases called clonal by models 1–4, 2–3, 4–1 and 4–2 were explored in detail

With the initial derivation cohort dataset, several models were generated based on different approaches, with very good predictive values for most of them. However, the regression equations/models generated included many parameters (for e.g., MCH, MCHC, MPV, MO %, BA %, MNS, MNS-SD and MNV-SD in model 1-1) whose inclusion was not explained by the disease biology. This, and encouragement by the learned peer reviewers, prompted us to validate these models on healthy persons (validation cohort 1) as well as consecutive routine samples of varied etiology (validation cohort 2). This approach, which yielded surprising results, should be mandatory for all future studies on positional parameters.

The performance characteristics of various models deteriorated from derivation cohort to validation cohort 1 and further in cohort 2. The most consistent performances were of models 1-4, 4-1 and 4-2 (Table 4). Model 1-4 unsurprisingly, comprises of only the LY# which itself is a major defining criterion for CLL whereas model 4-2 comprises of only lymphocytic positional parameters, whose stable performance can be readily explained based on the biology of disease. However, model 1-4 would have called all seven non-clonal cases with LY# in the range of $5.04\text{--}7.26 \times 10^9/\text{L}$ as clonal whereas the model 4-2 correctly picked all but one of them as non-clonal. Regarding specific lymphoproliferation subsets, of the four models that did not misclassify hairy cell leukemia (1-2, 2-1, 2-2, 4-1), three incorporated monocytic VCS parameters, while none of the other models had those. This is possibly due to the positional characteristics of hairy cells, that make them akin to monocytes rather than lymphocytes.

With regards to the performance characteristics, the specificity of model 4-2 being above 85% in all the sets may be acceptable to clinical labs. This model, even after excluding the TLC, LY % and LY# still yielded a reasonably good performance and hence might (in the future) successfully identify even cases of monoclonal B-cell lymphocytosis that typically have normal range absolute lymphocyte count [10]. In a routine CBC run, this model can be integrated in a LIS so as to serve as a preliminary screening tool to select cases for smear review and further ancillary test evaluation whenever lymphocytosis is encountered. Future studies can evaluate CLL and non-CLL cases with low absolute lymphocyte counts with these models.

The inclusion of various non-lymphocyte indices/parameters in other models might only represent an over-fitted condition since these models performed poorly when tested on routine samples. The reasons for such drastic change in performances may be selection bias as most of the patients in group 2 of derivation cohort were marrow aplasia and HIV infection, while the validation cohort 2 were unselected cases. Another reason might be

instrument-specific stability of VCS parameters and their dependence on critical -routine maintenance and calibration (beyond optimization and controls) that is not yet feasible for a routine high-volume clinical laboratory. These research-use-only parameters may therefore require more rigorous standardisation in the near future.

Compliance with Ethical Standards

Conflict of interest The authors declare that they have no conflict of interests.

Ethical Approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed Consent We request waiver of the requirement for informed consent on this study as it was performed on routine specimens sent for clinically indicated complete blood counts. The VCS data was obtained without any further costs, need for extra sample and without compromising the results of the clinical analysis. No data in this paper reveals the identity of the patients, and no interventions were made based on the experimental results.

References

1. Lecompte TP, Bernimoulin MP (2015) Novel parameters in blood cell counters. *Clin Lab Med* 35:209–224
2. Sharma P, Bhargava M, Sukhachev D et al (2014) LH750 hematology analyzers to identify malaria and dengue and distinguish them from other febrile illnesses. *Int J Lab Hematol* 36:45–55
3. Bhargava M, Saluja S, Sindhuri U et al (2014) Elevated mean neutrophil volume + CRP is a highly sensitive and specific predictor of neonatal sepsis. *Int J Lab Hematol* 36:e11–e14
4. Silva M, Fourcade C, Fartoukh C et al (2006) Lymphocyte volume and conductivity indices of the haematology analyser Coulter GEN.S in lymphoproliferative disorders and viral diseases. *Clin Lab Haematol* 28:1–8
5. Haschke-Becher E, Vockenhuber M, Niedetzky P et al (2008) A new high-throughput screening method for the detection of chronic lymphatic leukemia and myelodysplastic syndrome. *Clin Chem Lab Med* 46:85–88
6. Jean A, Boutet C, Lenormand B et al (2011) Combination of cellular population data and CytoDiff analyses for the diagnosis of lymphocytosis. *Clin Chem Lab Med* 49:1861–1868
7. Lin TS, Awan FT, Byrd JC (2013) Chronic lymphocytic leukemia. In: Hoffman R, Benz EJ, Silberstein LE et al (eds) *Hematology: basic principles and practice*, 6th edn. Elsevier, Philadelphia, pp 1170–1191
8. Sharma P, Rastogi P, Naseem S, Chand I, Sachdeva MUS, Varma N (2016) Cell population data and a novel LPD-index aid diagnosis of lymphocytosis by automated hematology analyzers. *Int J Lab Hematol* 38(Suppl 2):95–96
9. Bossuyt PM, Reitsma JB, Bruns DE et al (2003) Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *BMJ* 326:41–44
10. Hallek M, Cheson BD, Catovsky D et al (2008) Guidelines for the diagnosis and treatment of chronic lymphocytic leukemia: a report from the International workshop on chronic lymphocytic

- leukemia updating the national cancer institute-working group 1996 guidelines. *Blood* 111:5446–5456
11. Swerdlow S, Campo E, Harris N (2008) WHO classification of tumours of haematopoietic and lymphoid tissues. IARC press, Lyon
 12. Hosmer DW Jr, Lemeshow S, Sturdivant RX (2013) Applied logistic regression, 3rd edn. Wiley, Hoboken
 13. Furundarena JR, Uranga A, Sainz MR, González C, Uresandi N, Argoitia N, Araiz M (2017) Usefulness of the lymphocyte positional parameters in the Sysmex XN haematology analyser in lymphoproliferative disorders and mononucleosis syndrome. *Int J Lab Hematol*. <https://doi.org/10.1111/ijlh.12726>