Blood cell differential count discretisation modelling to predict survival in adults reporting to the emergency room: a retrospective cohort study

Fumagalli, Riccardo Mario ; Chiarelli, Marco ; Cazzaniga, Massimo ; Bonato, Claudio ; D'Angelo, Luciano ; Cavalieri D'Oro, Luca ; Cerino, Mario ; Terragni, Sabina ; Lainu, Elisa ; Lorini, Cristina ; Scarazzati, Claudio ; Tazzari, Sara Elisabetta ; Porro, Francesca ; Aldé, Simone ; Burati, Morena ; Brambilla, William ; Nattino, Stefano ; Locatelli, Matteo ; Valsecchi, Daria ; Spreafico, Paolo ; Tantardini, Valter ; Schiavo, Gianpaolo ; Zago, Mauro Pietro ; Fumagalli, Luca Andrea Mario

DOI: https://doi.org/10.1136/bmjopen-2023-071937

BMJ Open

Blood cell differential count discretisation modelling to predict survival in adults reporting to the emergency room: a retrospective cohort study

Riccardo Mario Fumagalli,[1,2] Marco Chiarelli,[3] Massimo Cazzaniga,[1] Claudio Bonato,[4] Luciano D'Angelo,[1] Luca Cavalieri D'Oro,[5] Mario Cerino,[1] Sabina Terragni,[1] Elisa Lainu,[1] Cristina Lorini,[1] Claudio Scarazzati,[1] Sara Elisabetta Tazzari,[1] Francesca Porro,[1] Simone Aldé,[1] Morena Burati,[1] William Brambilla,[1] Stefano Nattino,[1,6] Matteo Locatelli,[1,7] Daria Valsecchi,[1] Paolo Spreafico,[1] Valter Tantardini,[1] Gianpaolo Schiavo,[1] Mauro Pietro Zago,[3] Luca Andrea Mario Fumagalli [3]

To cite: Fumagalli RM, Chiarelli M, Cazzaniga M, *et al*. Blood cell differential count discretisation modelling to predict survival in adults reporting to the emergency room: a retrospective cohort study. *BMJ Open* 2023;**13**:e071937. doi:10.1136/bmjopen-2023-071937

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2023-071937).

Received 23 January 2023
Accepted 02 October 2023

For numbered affiliations see end of article.

Correspondence to
Dr Luca Andrea Mario Fumagalli; lu.fumagalli@asst-lecco.it

ABSTRACT

Objectives To assess the survival predictivity of baseline blood cell differential count (BCDC), discretised according to two different methods, in adults visiting an emergency room (ER) for illness or trauma over 1 year.

Design Retrospective cohort study of hospital records.

Setting Tertiary care public hospital in northern Italy.

Participants 11 052 patients aged >18 years, consecutively admitted to the ER in 1 year, and for whom BCDC collection was indicated by ER medical staff at first presentation.

Primary outcome Survival was the referral outcome for explorative model development. Automated BCDC analysis at baseline assessed haemoglobin, mean cell volume (MCV), red cell distribution width (RDW), platelet distribution width (PDW), platelet haematocrit (PCT), absolute red blood cells, white blood cells, neutrophils, lymphocytes, monocytes, eosinophils, basophils and platelets. Discretisation cut-offs were defined by benchmark and tailored methods. Benchmark cut-offs were stated based on laboratory reference values (Clinical and Laboratory Standards Institute). Tailored cut-offs for linear, sigmoid-shaped and U-shaped distributed variables were discretised by maximally selected rank statistics and by optimal-equal HR, respectively. Explanatory variables (age, gender, ER admission during SARS-CoV2 surges and in-hospital admission) were analysed using Cox multivariable regression. Receiver operating curves were drawn by summing the Cox-significant variables for each method.

Results Of 11 052 patients (median age 67 years, IQR 51–81, 48% female), 59% (n=6489) were discharged and 41% (n=4563) were admitted to the hospital. After a 306-day median follow-up (IQR 208–417 days), 9455 (86%) patients were alive and 1597 (14%) deceased. Increased HRs were associated with age >73 years (HR=4.6, 95% CI=4.0 to 5.2), in-hospital admission (HR=2.2, 95% CI=1.9 to 2.4), ER admission during SARS-CoV2 surges (Wave I: HR=1.7, 95% CI=1.5 to 1.9; Wave II: HR=1.2, 95% CI=1.0 to 1.3). Gender, haemoglobin, MCV, RDW, PDW, neutrophils, lymphocytes and eosinophil counts were significant overall. Benchmark-BCDC model included basophils and platelet count (area under the ROC (AUROC) 0.74). Tailored-BCDC model included monocyte counts and PCT (AUROC 0.79).

Conclusions Baseline discretised BCDC provides meaningful insight regarding ER patients' survival.

STRENGTHS AND LIMITATIONS OF THIS STUDY

⇒ Discretisation of simple BCDC in adults seeking immediate medical attention in emergency rooms provided significant survival insight.
⇒ The study size is adequate for explorative purposes.
⇒ Retrospective methodology and a lack of clinical data make the study findings unsuitable for clinical use.
⇒ Confirmation and validation are needed in a study setting, including detailed clinical data.

INTRODUCTION

Immune responses to chronic and acute diseases may drive disease outcomes, but assessments are rarely standardised in clinical practice. Therefore, it is often restricted to measuring the total white cell count in peripheral blood, a low-cost, easily performed laboratory test. Few studies have reported the predictivity of blood cell count on patient survival under acute and non-acute conditions.[1–3] Reference values of blood cell differential counts (BCDCs) are calculated based on the general population, whereas fewer

details are known for acutely ill patients. In this population subset, the identification of differential blood cell cut-off values may be challenging considering the non-linear effects of continuous variables on survival.[4 5] Among non-linear relationships, U-shaped associations between continuous biological variables and outcomes are commonly observed in clinical and epidemiological studies, and several statistical methods have been tested to solve the issue.[6] To assess the predictivity of peripheral BCDCs on survival, we retrospectively discretised automated peripheral BCDCs. Blood samples were taken at the initial visit of adults seeking immediate medical attention to our emergency room (ER) for illness, injury or medical condition, during the year 2020. Patients with SARS-CoV2 were included to aid in the identification of survival biomarkers[7] across wide disease spectrum. Patients' BCDC values were discretized by laboratory reference benchmark values (benchmark method) and by selected optimal cut-off finding methods (tailored methods), namely, optimal equal HR (OEHR) method[6] and maximally selected rank statistic (MSRS) method.[8] Two models were built selecting among BCDC those variables associated with overall survival. Finally, we analysed and compared the area under the receiver operating characteristic curve (AUROC) of both models originated from the sum of variables obtained by both methods.

METHODS

This retrospective cohort analysis was conducted using data recorded at the ER of the Alessandro Manzoni Hospital, Lecco, Italy (600 beds, national health services (NHS) hospital) between 1 January and 31 December 2020. Eligible patients were aged >18 years, consecutively admitted to the ER seeking immediate medical attention for illness, injury or medical condition, and for whom a complete blood differential count collection was indicated by ER medical staff at first presentation. The recorded patient characteristics were age, sex and outcome after evaluation by the ER medical staff, namely, discharge at home or in-hospital admission to any ward. Moreover, as two waves of the SARS-CoV2 worldwide epidemic in 2020 affected our healthcare facility, ER-admitted patients were split into three groups based on registered epidemic surges in Regione Lombardia issued by the Italian Health Authority (Istituto Superiore di Sanità).[9] The first wave time period (Wave I) was set starting from 20 February until 1 June; second wave (Wave II) starting from 1 October until 31 December; time out of these dates was considered as off-waves. Time period grouping was then analysed as a survival predictor variable along with demographics (age and gender) and ER discharge at home or admission to any hospital ward. A complete BCDC was performed using the automated Sysmex XN-9000 analyser on peripheral blood samples taken at baseline. Survival was the referral outcome for explorative model development and was assessed on 30 June 2021 by a population registry office query through the NHS territorial service. Predictors were searched among the BCDC automated analysis assessments of haemoglobin (Hb), mean corpuscular volume (MCV), red cell distribution width (RDW), platelet distribution width (PDW), platelet haematocrit (PCT) and the absolute count of red blood cells (RBC), white blood cells (WBC), neutrophils (Neu), lymphocytes (Lym), monocytes (Mon), eosinophils (Eos), basophils (Bas) and platelets (PLT). Missing data were excluded, as only patients having BCDC records were evaluated. Statistical analysis of BCDC was steadily carried out using descriptive methods for distribution and dispersion (skewness and kurtosis). Explorative analyses of continuous variables and differences between live and dead patients were performed by Mann-Whitney U test.

The 'benchmark' reference model was set by discretisation of BCDC continuous values on our laboratory reference interval, established according to the C28-A3 guideline by the Clinical and Laboratory Standards Institute.[10] The 'tailored' discretisation was set as follows. The relationship between each continuous variable and log relative hazard was plotted using the penalised B-splines (psplines) technique[11] for fitting the non-linear effect of covariate in Cox models,[12] by minimising pitfalls associated with dichotomisation of biological variables.[4] Variables were treated differently according to their respective distribution profile. Linear and sigmoid-shaped variables were dichotomised by the MSRS method.[8] U-shaped variables were univariately discretised by cut-off point determination using the OEHR method.[6] Discretised explanatory BCDC variables were then analysed using a Cox multivariable regression along with demographics (age and gender), ER destination (in-hospital admission or discharge at home) and year time period (WaveI, Wave II and off-waves). Two models containing independently significant BCDC subsets (one for each discretisation method) were built. Benchmark score sum and tailored score sum were calculated counting BCDC factors beyond cut-off values, for each of the two discretisation methods. The whole population was stratified into five risk groups according to score sum quintiles, and survival curves were calculated using the Kaplan-Meier method, and then compared by log-rank test in either method. Finally, receiver operating curves (ROCs) for both discretisation methods were drawn using either score sum. ROCs were then compared using the DeLong test.[13] Ethical obligations were fulfilled by the Hospital Board, in compliance with national regulations regarding retrospective observational studies. Analyses were performed using R and Jamovi (R-based free software).[14 15]

Patient and public involvement

Patients or the public were not involved in the design, or reporting or dissemination plans of our research.

Table 1 Patient characteristics

Characteristic (n=11052)	*
Age	67 (51–81)
Gender	
Female	5284 (48%)
Male	5768 (52%)
Discharged/inward admitted	
Discharged at home	6489 (59%)
In-hospital admitted	4563 (41%)
Wave I/II/off-waves	
Off-waves	5575 (50%)
Wave I	2807 (25%)
Wave II	2670 (24%)
Follow-up days	306 (208–417)
Alive/dead	
Alive	9455 (86%)
Deceased	1597 (14%)

*Median (IQR), n (%).

RESULTS

In this study, 11 635 complete blood cell counts were registered in the Emergency General Department during 2020. Patients younger than 18 years and pregnant women in labour were excluded, as listed in the Paediatric and Obstetric Emergency Laboratory Services subsets. After further removal of duplicated tests and repeated subsequent admissions, 11 133 patients remained. Outliers exceeding 99% of BCDC element values were removed (81 patients) to exclude extreme outliers, most likely affected by haematological diseases. Finally, 11 052 patients were available for analysis. The median age was 67 years (IQR 51–81) and 48% of participants were female. Fifty-nine per cent of patients (n=6489) were discharged, and 41% (n=4563) were acutely admitted to hospital wards. After a median follow-up of 306 days (IQR 208–417 dd), 9455 patients (86%) were alive, and 1597 (14%) were deceased (table 1).

According to SARS-CoV2 surges, 2807 (25.4%) patients were admitted to the ER during Wave I, 2670 (24.2%) during Wave II and 5575 (50.4%) during off-wave time periods. Out of all 4563 patients acutely admitted to the hospital, 543 (12%) were deceased within 30 days of admission, 599 (13%) were deceased more than 30 days of admission, whereas 3421 (75%) were alive at a median time of 10 months. Out of all 6489 patients discharged from the emergency department, 204 (3%) were deceased within 30 days, 251 (4%) died later, whereas 6034 (93%) were alive after a median time of 10 months. Mortality at 1 month was 12%, 7% and 4%, and overall mortality at 10 months was 20%, 13% and 12% during Wave I, Wave II and off-waves, respectively (online supplemental table 1) The flow of ER-presenting patients by year (SARS_Cov2 Wave I/II/offwaves), through discharge or inward-hospital admission, to follow-up status (alive or deceased) is depicted as an alluvial diagram in online supplemental figure 1.

Descriptive statistics, skewness, kurtosis and Q–Q plots of BCDC variables are reported and plotted in the supplemental material, respectively (online supplemental material section 1). Most of the BCDC subsets were non-linear. Differences between the alive and deceased groups were explored using continuous variable analysis and compared using the Mann-Whitney U test. All BCDC variables, except for Mon count, significantly differed between the dead and alive groups. Age showed a linear relationship with the log relative hazard plot and was discretised in two intervals on a cut-off set at 73 years by MSRS (online supplemental material section 2). Demographics (gender and age), clinical (ER medical staff indication for discharge or inward admission) and year time period (Wave I/II/off-waves) were analysed by multivariable Cox regression and found to be all independent predictors in a survival model. As expected, the strongest survival predictors were elder age (>73 years, HR=5.57, 95% CI 4.92 to 6.29) and inward admission (HR=2.83, 95% CI 2.53 to 3.16), whereas male sex HR was 1.18 (95% CI 1.07 to 1.30). Year time period HR were 1.77 (95% CI 1.58 to 1.98) and 1.27 (95% CI 1.11 to 1.44) for Wave I and Wave II, respectively (online supplemental table 2, online supplemental figure 2). Concerning BCDC cut-off tailoring, the relationship between each continuous variable and log relative hazard was plotted using the penalised B-splines technique.[12] Neu count was the only variable (as age) showing linear shape; Hb, MCV and RDW showed sigmoid shapes and were dichotomised by the MSRS method.[8] WBC, Lym, Mon, Eos, Bas count, PLT, PCT and PDW showed a non-linear U-shape and were univariately discretised by the OEHR method.[6] Exemplary linear, sigmoid and U-shaped variables are plotted in online supplemental material section 2.

The results of BCDC variables discretisation into favourable and unfavourable value intervals, according to selected methods, are detailed in table 2.

Details and graphics on variable discretisation for both OEHR and MSRS techniques are reported in online supplemental material section 2. Concerning the benchmark discretisation method, explanatory BCDC variables were discretised into on-reference and off-reference intervals, according to the reference laboratory value and analysed by Cox multivariable survival analysis regression along with gender, age, discharge/inward and Wave I/II/off-wave benchmark BCDC model output consisted of Hb (HR=1.56, 95% CI 1.39 to 1.75), RDW (HR=1.57, 95% CI 1.42 to 1.74), MCV (HR=1.30, 95% CI 1.16 to 1.44), Neu (HR=1.40, 95% CI 1.26 to 1.55), Lym (HR=1.44, 95% CI 1.29 to 1.61), Eos (HR=1.30, 95% CI 1.17 to 1.46), Bas (HR=1.37, 95% CI 1.16 to 1.62), PLT (HR=1.32, 95% CI 1.17 to 1.48), PDW (HR=1.17, 95% CI 1.04 to 1.32), along with male gender (HR=1.28, 95% CI 1.15 to 1.41), age >73 years (HR=4.57, 95% CI 4.03 to 5.18), inward admission (HR=2.18, 95% CI 1.94 to 2.44), Wave I (HR=1.71,

Table 2 Baseline blood cell differential count laboratory reference intervals (benchmark) and unfavourable intervals by MSRS/OEHR (tailored) discretisation

Covariate	Distribution profile	Benchmark discretisation	Tailored univariate discretisation	
		Laboratory reference interval (Clinical and Laboratory Standards Institute)	MSRS unfavourable interval	OEHR unfavourable interval
		Lower–upper	Single cut-off	L–R cut-off
Age (years)	Linear		>73	
Haemoglobin (g/dL)	Linear	13.5–17	<11.7	
RBC (10^9/L)	Sigmoid	4.50–5.90	<3.79	
MCV (fl)	Sigmoid	80-96	>97	
RDW (%)	Sigmoid	12.2–15	>13.8	
WBC (10^9/L)	Asym-U	4.0–9.0		<0.36 to >15.43
Neutrophils (10^9/L)	Linear	1.60–6.80	>9.65	
Lymphocytes (10^9/L)	Asym-U	0.8–4.5		<1.04 to >6.42
Monocytes (10^9/L)	Asym-U	0.12–1.08		<0.25 to >1.32
Eosinophils (10^9/L)	Asym-U	0.05–0.65		<0.01 to >1.41
Basophils (10^9/L)	Asym-U	0.01–0.20		<0.01 to >0.12
PLT (10^9/L)	Asym-U	150-400		<131 to >426
PCT (%)	Asym-U	0.19–0.40		<0.14 to >0.39
PDW (%)	Symm-U	9.6–15.2		<8.9 to >16

Asym, asymmetrical; Hb, haemoglobin ; MCV, mean cell volume ; MSRS, maximally selected rank statistic ; OEHR, optimal-equal HR; PCT, platelet haematocrit ; PDW, platelet distribution width ; PLT, platelets; RBC, absolute count of red blood cells ; RDW, red cell distribution width; Symm, symmetrical ; WBC, white blood cells.

95% CI 1.53 to 1.92) and Wave II (HR=1.19, 95% CI 1.04 to 1.36)—as in online supplemental table 3 and plotted in online supplemental figure 3. Details and modelling steps are plotted in online supplemental material section 3.

Tailored BCDC discretisation model output selected by Cox multivariable survival analysis regression consisted of Hb (HR=1.41, 95% CI 1.26 to 1.57), RDW (HR=2.24, 95% CI 2.00 to 2.52), MCV (HR=1.45, 95% CI 1.29 to 1.63), Neu (HR=1.38, 95% CI 1.23 to 1.54), Lym (HR=1.33, 95% CI 1.20 to 1.49), Mon (HR=1.20, 95% CI 1.05 to 1.39), Eos (HR=1.65, 95% CI 1.46 to 1.85), PCT (HR=1.43, 95% CI 1.27 to 1.61), PDW (HR=1.17, 95% CI 1.02 to 1.35), along with male gender (HR=1.23, 95% CI 1.12 to 1.36), age >73 years (HR=3.82, 95% CI 3.37 to 4.34), inward admission (HR=1.95, 95% CI 1.74 to 2.18), Wave I (HR=1.67, 95% CI 1.49 to 1.87) and Wave II (HR=1.17, 95% CI 1.03 to 1.34). Details are listed in table 3; the plot is depicted in online supplemental figure 4.

Among BCDC significant by Cox analysis elements, the median number of BCDC unfavourable variables in alive and dead patients was two and three for the benchmark and two and four for the tailored discretisation method, respectively. The number of unfavourable BCDC variables was classified into five equal groups and cutpoints were computed (score sum). The 20th, 40th, 60th and 80th percentiles were 1, 2, 3, 4 points and 0, 1, 2, 3 points for benchmark and tailored methods, respectively (online supplemental table 4).

Score sum calibration yielded mortality rates in the I, II, III, IV and V subgroups of 4.6%, 10.5%, 16.3%, 25.4% and 42.3% for the benchmark and 2.8%, 7.6%, 16%, 28.3% and 46.3% for the tailored method, respectively. The relationship between dead and alive in subgroups of patients stratified according to quintiles of the score sum (tailored method) is shown in figure 1. All details concerning score sum calibration in benchmark and tailored models are reported in online supplemental material section 3. Survival curves of the whole population, stratified into five risk groups according to quintiles score sum (0–4 in benchmark and 1–5 in tailored), were calculated using the Kaplan-Meier method and compared by log-rank test. The event summary is detailed in online supplemental table 5.

A growing probability of death was observed from the I to the V quintile, as shown in figure 2 for the tailored model.

Survival curves and log-rank tests are reported in online supplemental material section 3. Finally, we explored the predictive ability of survival of BCDC discretised using the benchmark and tailored methods. An ROC was generated by ranking the number of unfavourable BCDC variables for each discretisation method. Both curves were predictive for survival, with AUROCs of 74% and 79% for the

Table 3 Hazards regression—tailored baseline blood cell differential count and gender, age, Wave I/II/off-waves and inward/discharged

Dependent: survival		All	HR (univariable)	HR (multivariable)
Sex	F	5284 (47.8)	–	–
	M	5768 (52.2)	1.16 (1.05–1.28, p=0.003)	1.23 (1.12–1.36, p<0.001)
Age	fav (<73 year)	6693 (60.6)	–	–
	unfav (>73 years)	4359 (39.4)	6.72 (5.95–7.58, p<0.001)	3.82 (3.37–4.34, p<0.001)
Wave I/Wave II/off-waves	Off-waves	5575 (50.4)	–	–
	Wave I	2807 (25.4)	1.71 (1.53–1.92, p<0.001)	1.67 (1.49–1.87, p<0.001)
	Wave II	2670 (24.2)	1.30 (1.14–1.49, p<0.001)	1.17 (1.03–1.34, p=0.019)
Discharged home /inward	discharged home	6489 (58.7)	–	–
	inward	4563 (41.3)	4.03 (3.61–4.49, p<0.001)	1.95 (1.74–2.18, p<0.001)
Hb_tailr (>11.7 g/dL)	fav	9297 (84.1)	–	–
	unfav	1755 (15.9)	3.59 (3.24–3.97, p<0.001)	1.41 (1.26–1.57, p<0.001)
RDW_tailr (<13.8%)	fav	7360 (66.6)	–	–
	unfav	3692 (33.4)	4.76 (4.29–5.29, p<0.001)	2.24 (2.00–2.52, p<0.001)
MCV_tailr (<97 fl)	fav	9737 (88.1)	–	–
	unfav	1315 (11.9)	2.76 (2.46–3.10, p<0.001)	1.45 (1.29–1.63, p<0.001)
Neu_tailr (<9.65×10^9/L)	fav	9118 (82.5)	–	–
	unfav	1934 (17.5)	2.11 (1.89–2.35, p<0.001)	1.38 (1.23–1.54, p<0.001)
Lym_tailr (>1.04 to <6.42×10^9/L)	fav	8068 (73.0)	–	–
	unfav	2984 (27.0)	3.04 (2.75–3.35, p<0.001)	1.33 (1.20–1.49, p<0.001)
Mon_tailr (>0.25 to <1.32×10^9/L)	fav	10 167 (92.0)	–	–
	unfav	885 (8.0)	2.36 (2.06–2.71, p<0.001)	1.20 (1.05–1.39, p=0.010)
Eos_tailr (>0.01 to <1.41×10^9/L)	fav	9304 (84.2)	–	–
	unfav	1748 (15.8)	3.06 (2.76–3.40, p<0.001)	1.65 (1.46–1.85, p<0.001)
PCT_tailr (>0.14 to <0.39%)	fav	9839 (89.0)	–	–
	unfav	1213 (11.0)	2.87 (2.56–3.22, p<0.001)	1.43 (1.27–1.61, p<0.001)
PDW_tailr (>8.9 to <16%)	fav	10 079 (91.2)	–	–
	unfav	973 (8.8)	1.84 (1.60–2.12, p<0.001)	1.17 (1.02–1.35, p=0.026)

Eos, eosinophil; fav, favourable; Lym, lymphocyte; MCV, mean cell volume; Mon, monocytes; Neu, neutrophil; PCT, plateletcrit; PDW, platelet distribution width; RDW, red cell distribution width; tailr, tailored; unfav, unfavourable.

benchmark (including Hb, MCV, RDW, Neu, Lym, Eos, Bas, PLT and PDW) and tailored model (including Hb, MCV, RDW, Neu, Lym, Eos, Mon, PCT and PDW), respectively (online supplemental figure 5). According to the DeLong test, the predictivity of the tailored method of BCDC discretisation performed better than that of the benchmark laboratory reference values (area under the curve: 79% vs 74%, pairwise comparison 95% CI −0.06 to -0.04, p<0.001). Further details are available in online supplemental material section 3.

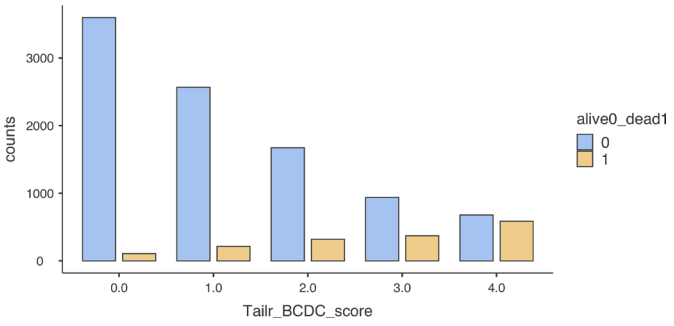

Figure 1 Counts of alive and deceased patients by increasing Tailored unfavourable BCDC score sum

Inclusion of demographic variables (age >73 years, male sex), ER outcome (inward admitted) and year time-period (Wave I=2 points and Wave II=1 point) in the models increased AUROCs predictivity on survival from 74% to 80% and 79% to 83% for the benchmark and tailored discretisation method, respectively (online supplemental material section 3). In order to test the study hypothesis in the SARS-CoV2-free population, namely, off-wave patients which counted 50% of all study population (n=5575), the same models were tested after removing Wave I and Wave II subgroups. Characteristics are listed in online supplemental material section 4. Multivariable Cox analysis of characteristics, discharge/

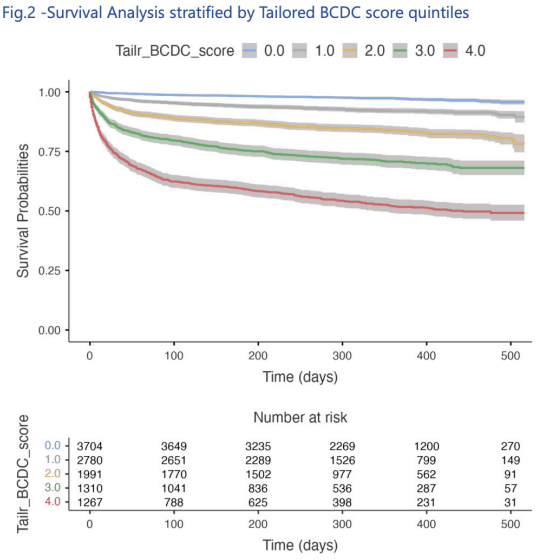

Figure 2 Survival analysis stratified by tailored BCDC score quintile. BCDC, blood cell differential count.

inward and BCDC discretisation methods yielded slighter models, but the most significant variables, relative HR and 95% CI values, remained stable. Among baseline characteristics and discharged/inward variables, the male gender no longer resulted significantly with respect to the whole population. Benchmark discretisation of MCV, Eos count (online supplemental table 6) and tailored discretisation of Mon count (online supplemental table 7) were excluded from the respective off-wave survival models. Plots are illustrated in online supplemental figure 6 and online supplemental figure 7, respectively. Further details are available in online supplemental material section 4.

DISCUSSION

An automated complete blood differential count represents a simple, low cost and easily performed assessment of a patient's general and immune status in emergency settings. However, its predictive ability for survival remains to be thoroughly assessed. This retrospective study on patients referring to our ER during 2020 showed that selected elements from a single baseline blood sample, taken at initial presentation for any illness or injury, were efficient in predicting survival, independently from demographic parameters (age and sex), from the ER medical staff decision to discharge or hospitalise the patient and from the 2020 year period of ER presentation (during SARS-CoV2 epidemics Wave I, Wave II or during off-wave periods). The laboratory reference values defined on general population intervals were narrow compared with those computed using tailored methods on HR. The HRs of each significant BCDC variable and AUROC overall predictivity were lower considering the interval was based on the general population instead of targeted at survival in a general population subset, such as ER-presenting patients. Concerning the tailored discretisation method, the OEHR methodology[6] was very helpful for the discretisation of U-shaped variables, which are biologically relevant in health and disease. This method targeted value intervals that were not purely based on frequency in the general population for each BCDC element. Additionally, it was more methodologically fitted and did not rely on a fixed a priori threshold but was specifically based on HR. Furthermore, the OEHR method proved useful for the discretization of variables characterised by physiologically low cell counts within BCDC (Eos and Bas counts), which are flawed by intrinsic inaccuracy in the automated counts.[16 17] Clinically significant tailored thresholds largely consisted of either absence ($<0.01\times10^9$/L) or consistent presence of Eos or Bas in peripheral blood. Both methods for value discretisation exhibited an acceptable performance on survival predictivity, although the purpose is explorative. Moreover, it is clear that the increasing number of unfavourable BCDC variables at baseline is associated with increased mortality risk, as demonstrated by both models. Beyond statistical methodology implications, this would represent biological insight into marrow as well as spleen and lymphatic tissue response. Both of these processes are often hurt during acute illness or trauma, so impairment extent and severity are likely linked with poor outcomes. The major study limitation was the lack of data regarding patient performance status, symptoms, vital signs and disease diagnosis, all of which were required to assess the relevance of BCDC for outcome predictions. Such a defect may flaw the study findings beyond the fact that the study purpose is explorative. To outflank the defect, we introduced inward admission among explanatory variables. This is usually considered an outcome separate from survival in clinical studies. In this study setting, such variables could roughly surrogate a more severe clinical condition and/or clinically relevant instrumental findings, resulting in the medical staff's decision to admit the patient to the hospital. Further analysis of the detailed clinical characteristics dataset is needed; as a first step, nested retrospective studies on patient subgroups, whose clinical course is better defined, are worth. Another concern relates to the COVID-19 pandemic, which affected the number of patients referring to the ER, hospitalisation rates and the mortality rate in 2020. The Italian region of Lombardia, where our hospital is located, underwent two SARS-CoV2 waves during 2020, resulting in two peaks in hospital admissions of approximately 13 000 patients in early April and 10 000 in November,[9] out of a population of 10 million people (45% over 50 years old). The influx of patients with COVID-19 in our hospital during 2020 might have skewed the results of this study. Consequently, the discretised BCDC value intervals identified as favourable or unfavourable in this population may be unsuitable if applied in a different timeframe. Indeed, we split the year time into three periods so that we could reasonably assume that, at least in the off-wave period, the population accrued in the ER was mostly free from COVID-19, thus providing an adequate sample population for testing the

study hypothesis about BCDC predictivity on survival either within a COVID-19-free sample or together with a larger, mixed COVID-19 population (Waves I and II). Likely useful insights emerged from this retrospective monocentric study. First, thoroughly investigating BCDC variables in patients addressing the ER demonstrated usefulness for outcome predictions, as we found for counts of Neu, Lym, Mon, Eos, Bas and for values of RDW, MCV and PDW. Surprisingly, general use variables (ie, total WBC count and PLT count) were excluded from the multivariable Cox analysis models, likely due to both collinearity and lower predictive ability with respect to the other BCDC subsets. The clinical implications of lower predictive ability of WBC and PLT found in this study should be considered cautiously; anyway in our ER population, combinations of low Hb, high RDW, MCV, neutrophilia, lymphopenia, eosinopenia and/or altered PDW and PCT were frequently encountered and strongly reflected increased mortality risks. In longitudinal studies, lymphopenia is associated with increased mortality risk for all causes.[18 19] Lymphocytopenia in severely ill patients[20] was a better predictor of bacteraemia in comparison to the total leucocyte and Neutrophil count in a cohort of 21 372 cases. Lymphopenia was a predictor of illness severity and short-term mortality risk in a cohort of 58 260 patients admitted to the hospital.[21] Lymphopenia in chronic diseases is an unfavourable independent variable, including in cancer.[22–24] In prospective studies, lymphopenia was associated with an increased risk of all cause-specific mortality[19] and was additive to traditional risk factors.[18] Extreme lymphocytosis was noticed in haematological disease and in patients with central nervous system bleeding or head trauma with poor prognosis.[1] Concerning RDW and lymphopenia, a recent study on 1641 SARS-CoV2 hospitalised patients demonstrated that high RDW at baseline and subsequent RDW elevation increased in-hospital mortality. This finding is not restricted to COVID-19, as it was described in a retrospective study on 1715 chronic hepatitis C virus patients undergoing a 5-year follow-up.[25] Lymphopenia and elevated RDW were associated with long-term mortality risk in 15 179 patients undergoing coronary angiography in both acute and non-acute settings.[26] Eosinopenia was reported as a marker of poor outcome in lung disease[27] and critically ill patients.[28] Absolute eosinopenia was associated with clinically poor outcomes in first-wave COVID-19 pneumonia.[7 29] In experimental rabbit models, lymphocytes are redistributed from peripheral blood to lymphatic tissue after cortisol administration, for instance, following major surgery. *Escherichia coli* endotoxemia and surgery were accompanied by lymphocytopenia and increased cortisol.[30–32] Conversely, lymphocytes are redistributed from the spleen and bone marrow to peripheral blood, lungs and liver after epinephrine infusion.[33] From a pathophysiological perspective, lymphopenia, including eosinopenia, neutrophilia and increased RDW[34] may be markers and drivers of an ineffective response to major health-disturbing events. Our study represents a straightforward explorative attempt to accurately assess the acute patient immune status at presentation, enabling the detection of at-risk patients, mostly overt at short-term (1–3 months) but significant over longer time (significant at 10 months). Further investigation should include prognostic stratification and evenly tailored diagnostic and therapeutic pathways. Such a model cannot be applied in a clinical setting but represents a basis for further clinical research addressing whether baseline BCDC can generate reliable and specific biomarkers for premature detection of specific acute conditions. Finally, long-term therapeutic research should explore whether it is possible to actively manage host immunity by reproducing an effective response to improve the overall outcome.

CONCLUSION

Values from automated peripheral BCDCs taken at baseline in adult patients visiting our ER during 2020 were discretised using laboratory reference values (benchmark) or OEHR and MSRS (tailored). Variables were adequately efficient and robust in predicting 1-year survival, independently from demographics (age and sex), from year time period (SARS-Cov2 Wave I, Wave II and off-waves) and ER medical staff decision to discharge or hospitalise patients. The tailored discretisation of Hb, MCV, RDW, Neu, Lym, Mon, Eo, PCT and PDW yielded more accurate survival predictions than the benchmark laboratory reference interval in our cohort of patients. Further studies are warranted to validate these findings and explore whether specific BCDC patterns can predict the outcomes of single acute diseases or conditions.

Author affiliations
[1]Dipartimento Emergenza Accettazione, Pronto Soccorso, Ospedale Alessandro Manzoni, Lecco, LC, Italy
[2]Klinik für Angiologie, UniversitätsSpital Zürich, Zurich, Switzerland
[3]Dip.Chirurgico, Chirurgia Urgenza, Ospedale Alessandro Manzoni, Lecco, LC, Italy
[4]Dipartimento Servizi Clinici, Ospedale Alessandro Manzoni, Lecco, LC, Italy
[5]UOC Epidemiologia, Agenzia per la Tutela della Salute Brianza, Monza, Lombardia, Italy
[6]Scuola Spec. Medicina Emergenza-Urgenza, Università degli Studi di Milano, Milano, Lombardia, Italy
[7]Polo formativo, Agenzia per la Tutela della Salute Brianza, Monza, Lombardia, Italy

Acknowledgements We are very grateful to Alessandro Manzoni Hospital units of 'Pronto Soccorso' and 'Dipartimento Servizi Clinici' graduate, nursing and technical staff; to Dr M Tavola, and G Baraldo for their wise hints; to Dr M Biffi and Dr C Curioni for their factual help; to Dr G Nattino (Unit of Causal Inference in Epidemiology, Istituto Mario Negri IRCCS) for invaluable advice. We would like to thank Editage (www.editage.com) for English language editing.

Contributors RMF, M.Cazzaniga, M.Chiarelli, LAMF: conception and design of the work, acquisition, analysis, and interpretation of data, drafting the work and revising it critically for intellectual content. CB, LD'A, LCD'O and MPZ: design of the work, acquisition and interpretation of data and critical revision for intellectual content. M.Cerino, ST, EL, CL, CS, SET, FP, SA, MB, WB, SN, ML, DV, PS, VT and GS: provided and cared for study patients; contributed substantially to the design of the work, collection, acquisition and interpretation of data and critical revision for intellectual content. LAMF is responsible for the overall content as the guarantor. All authors gave final approval to the submitted manuscript and ensured that questions related to the accuracy or integrity of any part of the work were appropriately investigated and resolved. All authors declare that no one did receive support for the present manuscript (eg, funding, provision of study materials, medical writing, article

processing charges, etc) present or past; nor grants or contracts from any entity; nor royalties or licenses; nor consulting fees; nor payment or honoraria for lectures, presentations, speakers bureaus, manuscript writing or educational events; nor payment for expert testimony; nor support for attending meetings and/or travel; nor patents are planned, issued or pending; nor participation on a Data Safety Monitoring Board or Advisory Board; nor leadership or fiduciary role in other board, society, committee or advocacy group, paid or unpaid; nor stock or stock options; nor receipt of equipment, materials, drugs, medical writing, gifts or other services, nor other financial or non-financial interests.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Disclaimer The views expressed in the manuscript are Authors' own.

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Ethics approval This study concerns retrospective human data, indeed Institutional Board exempted this study according to EU and Italian regulations on good clinical practice (law decree n 211 issued 24 june 2003) concerning retrospective observational studies. Retrospective observational studies on de-identified data are exempted from Ethical Committee assessment and formal approval. Patients admitted in public (NHS) hospital are required to read and sign a form concerning storage and use of anonymized data on health records, as in our case. Hospital Board exemption for this study was stated in 'Deliberazione n 566 14 Luglio 2021, Atti del Direttore Generale anno 2021, ASST Lecco Regione Lombardia'. PDF full copy is loaded under 'supplemental materials for Editors only'.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available in a public, open access repository. Extra data can be accessed via the Dryad data repository at http://datadryad.org/ with the doi: 10.5061/dryad.dncjsxm5g.

ORCID iD

Luca Andrea Mario Fumagalli http://orcid.org/0000-0002-4670-7277

REFERENCES

1 Kho AN, Hui S, Kesterson JG, *et al*. Which observations from the complete blood cell count predict mortality for hospitalized patients? *J Hosp Med* 2007;2:5–12.
2 Shimoni Z, Froom P, Benbassat J. Parameters of the complete blood count predict in hospital mortality. *Int J Lab Hematol* 2022;44:88–95.
3 Chiarelli M, Achilli P, Tagliabue F, *et al*. Perioperative lymphocytopenia predicts mortality and severe complications after intestinal surgery. *Ann Transl Med* 2019;7:311.
4 Altman DG, Royston P. The cost of dichotomising continuous variables. *BMJ* 2006;332:1080.
5 Royston P, Altman DG, Sauerbrei W. Dichotomizing continuous predictors in multiple regression: a bad idea. *Stat Med* 2006;25:127–41.
6 Chen Y, Huang J, He X, *et al*. A novel approach to determine two optimal cut-points of a continuous predictor with a u-shaped relationship to hazard ratio in survival data: simulation and application. *BMC Med Res Methodol* 2019;19:96.
7 Cazzaniga M, Fumagalli LAM, D'angelo L, *et al*. Eosinopenia is a reliable marker of severe disease and unfavourable outcome in patients with COVID-19 pneumonia. *Int J Clin Pract* 2021;75:e14047.
8 Therneau Terry M, Thomas L. Survival: survival analysis. 2019. Available: https://CRAN.R-project.org/package=survival
9 Istituto Nazionale Fisica Nucleare, COVIDSTAT, Available: https://covid19.infn.it
10 CLSI. Defining, establishing, and Verifying reference intervals in the clinical laboratory; approved guideline -Third edition. In: *CLSI document EP28-A3c*. Wayne, PA: Clinical and Laboratory Standards Institute, 2008.
11 Perperoglou A, Sauerbrei W, Abrahamowicz M, *et al*. A review of spline function procedures in R. *BMC Med Res Methodol* 2019;19:46.
12 Eilers PHC, Marx BD. Flexible smoothing with B-Splines and penalties. Statististical Science 1996;11:89–121.
13 DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988;44:837–45.
14 R Core Team. R: A language and environment for statistical computing. 2020. Available: https://cran.r-project.org
15 The Jamovi project [jamovi. (Version 1.6)]. Computer Software; 2021. Available: https://www.jamovi.org
16 Amundsen EK, Urdal P, Henriksson CE. Poor accuracy of the eosinophil count of hematology instruments limits the use of eosinopenia as a marker of bacterial infection or increased mortality. *Int J Lab Hematol* 2015;37:e119–22.
17 Amundsen EK, Henriksson CE, Holthe MR, *et al*. Is the blood basophil count sufficiently precise, accurate, and specific?: three automated hematology instruments and flow cytometry compared. *Am J Clin Pathol* 2012;137:86–92.
18 Zidar DA, Al-Kindi SG, Liu Y, *et al*. Association of lymphopenia with risk of mortality among adults in the US general population. *JAMA Netw Open* 2019;2:e1916526.
19 Warny M, Helby J, Nordestgaard BG, *et al*. Incidental lymphopenia and mortality: a prospective cohort study. *CMAJ* 2020;192:E25–33.
20 Wyllie DH, Bowler ICJW, Peto TEA. Relation between lymphopenia and bacteraemia in UK adults with medical emergencies. *J Clin Pathol* 2004;57:950–5.
21 Andreu-Ballester JC, Pons-Castillo A, González-Sánchez A, *et al*. Lymphopenia in hospitalized patients and its relationship with severity of illness and mortality. *PLoS One* 2021;16:e0256205.
22 Stanley KE. Prognostic factors for survival in patients with inoperable lung cancer. *J Natl Cancer Inst* 1980;65:25–32.
23 Lissoni P, Fumagalli L, Paolorossi F, *et al*. Changes in lymphocyte number during cancer chemotherapy and their relation to clinical response. *Int J Biol Markers* 1999;14:115–7.
24 Lissoni P, Brivio F, Fumagalli L, *et al*. Efficacy of cancer chemotherapy in relation to the pretreatment number of lymphocytes in patients with metastatic solid tumors. *Int J Biol Markers* 2004;19:135–40.
25 Damjanovska S, Davitkov P, Gopal S, *et al*. High red cell distribution width and low absolute lymphocyte count associate with subsequent mortality in HCV infection. *Pathog Immun* 2021;6:90–104.
26 Zafrir B, Hussein S, Jaffe R, *et al*. Lymphopenia and mortality among patients undergoing coronary angiography: long-term follow-up study. *Cardiol J* 2022;29:637–46.
27 Holland M, Alkhalil M, Chandromouli S, *et al*. Eosinopenia as a marker of mortality and length of stay in patients admitted with exacerbations of chronic obstructive pulmonary disease. *Respirology* 2010;15:165–7.
28 Abidi K, Belayachi J, Derras Y, *et al*. Eosinopenia, an early marker of increased mortality in critically ill medical patients. *Intensive Care Med* 2011;37:1136–42.
29 Ndieugnou Djangang N, Peluso L, Talamonti M, *et al*. Eosinopenia in COVID-19 patients: a retrospective analysis. *Microorganisms* 2020;8:1929.
30 Toft P, Hokland M, Hansen TG, *et al*. Changes in lymphocyte subpopulations and adhesion/activation molecules following endotoxemia and major surgery. *APMIS* 1995;103:261–6.
31 Toft P, Tønnesen E, Svendsen P, *et al*. Redistribution of lymphocytes after cortisol administration. *APMIS* 1992;100:154–8.
32 Toft P, Svendsen P, Tønnesen E, *et al*. Redistribution of lymphocytes after major surgical stress. *Acta Anaesthesiol Scand* 1993;37:245–9.
33 WHITE C, LING TH, KLEIN AM. The effect of administration of epinephrine on the leukocyte counts of normal subjects. Blood. 1950 Aug;5(8):723-31. PMID: 15426653. 1950.
34 Turcato G, Cervellin G, Bonora A, *et al*. Red blood cell distribution width improves reclassification of patients admitted to the emergency department with acute decompensated heart failure. *J Med Biochem* 2018;37:299–306.