Polar Biology (2025) 48:106
https://doi.org/10.1007/s00300-025-03422-5

RESEARCH

One for all and all for one: an efficient blood sampling strategy to assess penguin health and immune status

Silvia Simonetti[1] · Niccolò Fattorini[2,3] · Verónica D'Amico[4] · Ilaria Corsi[1] · Silvia Olmastroni[1]

Received: 8 April 2025 / Revised: 28 August 2025 / Accepted: 3 September 2025
© The Author(s) 2025

Abstract

Antarctic penguins are considered sentinels of ecosystem health. They face multiple anthropogenic threats, including global climate change and local stressors such as tourism, fisheries, and pollution. Monitoring their physiological responses to stress is essential for assessing the Antarctic population and ecosystem health. On the other hand, it is equally important to minimize impact of handling stress, i.e., reducing time of restraint, and if possible, reducing the number of birds sampled. Given the valuable insights provided by hematological parameters in monitoring and assessing individual physiological conditions, there is growing interest in refining analytical methods (i.e., the amount of blood) and counting techniques to minimize analysis time and the number of samples required. We here refined the analysis of Erythrocytes Nuclear Abnormalities (ENAs), for genomic instability, and total White Blood Cells (WBCs) for immune status, in Adélie penguin (*Pygoscelis adeliae*) blood smears. The improved method, using Giemsa staining, effectively distinguishes nuclear and cytoplasmic components in both erythrocytes and leukocytes, enabling the reliable enumeration of ENAs and the clear identification of different WBC classes. A statistical sample size-rarefaction analysis also informed on the number of individuals required, optimizing sampling effort while maintaining data precision. More specifically, the sums of ENAs and WBCs can be precisely estimated for all the three study colonies with a sample size of about 20–25 individuals, while the parameters micronucleus and two-lobed nucleus show greater uncertainty. For most other parameters, the estimates stabilize with approximately 15–20 samples. Our findings provide a foundation for developing optimized sampling protocols for improved hematological analysis in penguins, offering a practical tool for assessing their health and immune status; therefore, the adoption of the presented method should be supported as a cost-effective and less invasive tool at the population level for long-term conservation studies.

Keywords Blood smears · *Pygoscelis adeliae* · Erythrocytes nuclear abnormalities · White blood cells · Field sampling effort · Stress ecology

Introduction

Antarctic seabirds are threatened by several anthropogenic stressors originating from both global (i.e., climate changes) and local sources (i.e., tourism, fishery, pollution) that are challenging their health and survival (Croxall et al. 2002; Ropert-Coudert et al. 2019; Bestley et al. 2020). Despite evolutionary robustness and resilience toward multiple stressors, Antarctic penguin populations are currently at considerable risk in the Southern Ocean, being vulnerable to shifts in terrestrial and marine environments, to which the novel impacts from climate change (Trathan et al. 2014) and wildlife diseases (i.e., avian influenza: Aguado et al. 2024; Banyard et al. 2024) add up. Widely accepted as sentinels of the Southern Ocean ecosystem health, Antarctic penguins reflect spatial and temporal changes in their habitat (i.e., breeding, foraging, and trophic webs) and thus have been used in long-term ecological research programs for monitoring and conservation purposes (Ainley 2002; Boersma 2008; Southwell et al. 2017).

✉ Silvia Simonetti
silvia.simonetti2@unisi.it

[1] Department of Physical, Earth and Environmental Sciences, University of Siena, Via Mattioli 4, 53100 Siena, Italy

[2] Department of Life Sciences, University of Siena, Via Mattioli 4, 53100 Siena, Italy

[3] National Biodiversity Future Center, Palermo, Italy

[4] Centro Para El Estudio de Sistemas Marinos (CESIMAR) - CCT CONICET CENPAT, Brown 2915, U9120ACF Puerto Madryn, Chubut, Argentina

Published online: 27 September 2025

Over the years, sampling techniques and methodologies used in scientific studies have been improved in order to reduce disturbance and invasive manipulation and prevent unnecessary handling of penguins and their colonies (Le Maho et al. 2014; Turcu et al. 2023). Issues consequent of animal handling potentially causing injury, spread of non-native pathogens (i.e., human-borne or exotic diseases), and distress have been raised by the scientific community over the years (Kerry and Riddle 2009; Fraser et al. 2013; Coetzee and Chown 2016).

Blood collection has been used by far as a suitable monitoring tool to assess penguin population genetics, diet, exposure to pollutants, and, in general, stress, i.e., from pathogens and parasites (Maceda-Veiga et al. 2015). All these considered, information on the health status of the Antarctic penguin population is crucial for evaluating their ability to cope with environmental changes. Blood analysis has great potential to provide insights into the overall functioning of the organism. Hormone levels, isotope signatures, and genetic analysis can be immediately affected by handling stress and often require complex preservation methods (Maceda-Veiga et al. 2015). On the other hand, hematological parameters from blood smear, such as erythrocyte and leukocyte profiles, remain stable after sampling (Davis 2005) and involve simpler storage procedures (Maceda-Veiga et al. 2015). Analyzing blood smears is considered among the most cost-effective and informative non-lethal methods to assess individuals' health compared to blood analysis (Maceda-Veiga et al. 2015), capable of depicting physiological conditions, responses to stress, diseases, and genotoxicity (Menéndez-Blázquez et al. 2021; Kophamel et al. 2022). They have been successfully incorporated into several monitoring programs involving penguin populations to obtain a useful baseline knowledge of unique physiological parameters, as a reference to define spatiotemporal trends in physiological changes and detect underlying perturbations (De Mas et al. 2015; Colominas-Ciurò et al. 2017; Palacios et al. 2018). Guidelines have been proposed on procedural aspects of blood collection to minimize potential negative effects, including training researchers to properly conduct this procedural step by limiting stress and isolating penguins from the colony or nest (Wells et al. 2023). Approval from national Antarctic programs is required before sampling penguin blood, and the minimally harmful sampling methods should be adopted to minimize stress, pain, and long-term impact on birds.

Erythrocytes Nuclear Abnormalities (ENAs) and White Blood Cells (WBCs), commonly assessed from smears, indicate genomic instability (ENAs; Heddle et al. 1983; Hayashi 2016) and immune/stress status (WBCs; Davis et al. 2008). Differential WBC counts can help distinguish a cellular immune response resulting from disease or inflammation (Davis et al. 2008), and the ratio between heterophils and lymphocytes (H/L ratio) is considered a reliable and accurate physiological indicator of the stress response (Gross and Siegel 1983; Müller et al. 2011). Different studies on Antarctic seabirds aim to assess physiological differences and variability, cell immune response, health status, and stress levels through the analyses of either ENAs or WBCs (as in Menéndez-Blázquez et al. 2021; De Mas et al. 2015, and references within). Giemsa staining has been by far the most widely used for ENA or WBC identification (De Mas et al. 2015; Colominas-Ciurò et al. 2017; Palacios et al. 2018) due to its ability to provide good contrast between nuclei and cytoplasm (Campbell 2015). However, improper preparation can lead to overstaining and inconsistent results, thus limiting the advantages of the staining method (WHO 2016). Giemsa stain was also combined with Wright dye (Gallo et al. 2019; Vleck et al. 2000) to enhance nuclear and cytoplasmic detail, but this combination is specifically used for WBC differentiation. Other staining methods, i.e., Wright-Rosenfeld (Gallo et al. 2019) and May-Grunwald (Hawkey et al. 1989), were commonly used for only WBC counting in avian blood smears. Conversely, they are not suitable for ENA analysis unless used in combination with Giemsa staining, and they are only used for limited investigations and never applied routinely despite the great advantages (Kursa and Bezrukov 2008). Additionally, although a less-known stain method, Tinción 15 (D'Amico et al. 2014, 2016), has been proven to contrast nucleated cells, making it useful for ENA and WBC identification.

Considering the range of available staining techniques, refining methodologies, and establishing a standardized approach would be crucial to improving the efficiency and accuracy of blood parameter assessments in Antarctic penguins. A standardized laboratory method for optimal staining would implement smears reading, possibly reducing the number of samples and/or the amount of blood required for analysis. Furthermore, reducing the number of individuals required to collect biological samples, thus limiting the number of captures and handling and, ultimately, the time spent in the field to conduct these activities, would increase the cost-effectiveness of penguin monitoring. Although advances have been made in minimizing the fieldwork effort for assessing individual characteristics such as sex (e.g., Polito et al. 2012; Fattorini and Olmastroni 2021), there is limited information on immune-hematological parameters.

In our previous study on the drivers of immune-hematological parameters of Adélie penguins, we used Giemsa staining (Olmastroni et al. 2024) since it was the most commonly adopted dye, ensuring an easier result comparison with past studies. Here, we aim to refine the methodology behind that work, which was developed based on past studies analyzing Antarctic penguin blood smear (reported above) and validated to enhance existing techniques by integrating ENA and WBC counts. Currently, the possibility of adapting

the method to simultaneously evaluate ENAs and immune cells would significantly reduce analysis time. This approach enables the extraction of dual information on genomic stability and immune function from just a few drops of blood, making it a more efficient alternative to conducting separate analyses. Finally, we investigated the possibility of obtaining precise measurements of immune-hematological parameters under simulated scenarios of reduced sample size to optimize monitoring procedures.

Materials and methods

Blood smear samples analyzed in the present study have been collected from adult Adélie penguins (*Pygoscelis adeliae*, Hombron & Jacquinot, 1841) breeding in three colonies (Edmondson Point—EdPo, Adelie Cove—AdCo, Inexpressible Island—InIs) in Mid Victoria Land, Ross Sea (Antarctica), during 2017–18 Antarctic summer as a part of a wider research monitoring program. Details on the study area, sampling, and field methods have already been described in our previous studies (Olmastroni et al. 2019, 2024). Descriptive statistics (e.g., mean and standard deviation) for ENAs and WBCs are presented in Olmastroni et al. (2024). Here, we detailed the novel methodological approach applied for staining and counting ENAs and WBCs.

Blood smears analysis

Smear samples were dried at ambient temperature and stored at 4 °C in a plastic box until analysis. All samples were processed in the Ecotoxicology and Biological Monitoring research Lab of the University of Siena (Italy). Slides were fixed with cold (-20 °C) methanol (100% v/v) for 3 min, dried at room temperature (20 °C), and stained with a solution of 6% Giemsa in Milli-Q (pH 7.26), previously filtered with cellulose-acetate membrane syringe filters, pore size 0.22 µm (Sartorius). Smears were then rinsed in Milli-Q water, covered with a coverslip, and observed under a light microscope (Olympus BX51) equipped with a Camera. The analysis of ENAs and WBCs was carried out as follows. Initially, slides were examined at 40 × magnification, progressively scanning the x and y axes to identify monolayer fields with a similar density of well-stained erythrocytes.

The manual counting of nuclear abnormalities was performed on 10,000 erythrocytes using 100 × (oil immersion) magnification according to Zúñiga-González et al. (2000). Total number of ENAs was calculated for each individual and classified according to De Mas et al. (2015) as: micronucleus (MN), lobed nucleus (LN), tailed nucleus (TN), two-lobed nucleus (TL), budding nucleus (BN), nucleus with cavity (WC), kidney-shaped nucleus (KN), unknown nuclear malformation (UN).

The percentage of WBC types was determined by two different counts, both performed at 100 × (oil immersion) magnification: (i) the number of leukocytes in each class out of 100 immune cells (as usually assessed in previous works; D'Amico et al. 2016; Palacios et al. 2018) and (ii) the percentage of leukocytes in each class out of the total immune cells observed among 10,000 erythrocytes. Leukocytes were classified according to Samour (2006) as: heterophils (H), eosinophils (E), lymphocytes (L), basophils (B), and monocytes (M). The H/L ratio was evaluated from the WBCs profile, according to Davis et al. (2008). Furthermore, the total number of WBCs was assessed with two different counts: (i) among 10,000 erythrocytes at 100 × (oil immersion) magnification (estimated simultaneously ENAs count) and (ii) counted in 10 adjacent visual fields at 40 × magnification (as usually assessed in previous works; D'Amico et al. 2016; Palacios et al. 2018; Menéndez-Blázquez et al. 2021) to compare the two counting methods.

The time required to count 10,000 erythrocytes is approximately 2.5 h. Assessing the total number of WBCs simultaneously with the ENAs count does not require additional time, whereas counting WBCs in 10 adjacent visual fields requires at least an additional 30 min.

We analyzed 87 smears from as many individual penguins. However, the leukocyte composition could not be determined for one sample. Since the statistical evaluation for each parameter is independent from one another, we did not exclude the sample without the leukocyte profile. Therefore, the sample size was n = 87 for all parameters, and n = 86 for parameters concerning the leukocyte composition.

Our evaluation assumed a robust technical precision in all counts, meaning that, for each parameter, the number of cells in each smear was counted with no measurement error. All samples were analyzed by the same researcher to avoid potential inter-observer variability and bias in counts, and the researcher manually counted cells, which reasonably allowed us to meet this assumption. Eventually, as we evaluated one smear per individual penguin, our counts also assumed that each sample was representative of the immune-hematological levels of the relevant individual (Davis et al. 2008; Menéndez-Blázquez et al. 2021).

Statistical analysis: rarefaction to inform required sample size

We performed a general sample size-rarefaction analysis on each immune-hematological indicator measured to evaluate how the uncertainty in the mean response changed with reduced number of sampled penguins. We considered all the 23 parameters (i.e., the total number of ENAs/10,000 erythrocytes, the percentages of MN, LN, TN, TL, BN, WC, KN, and UN, the total number of WBCs/10,000 erythrocytes, the total number of WBCs/10 fields, as well as the

percentages of H, E, L, B, M and the H/L ratio counted both on 10,000 erythrocytes and 100 WBCs) (Table S1). Since the original number of collected samples differed between the three study colonies, and because researchers are usually interested in assessing site-specific mean responses as baseline levels separately across study sites, we conducted rarefactions separately for each colony. Thus, we arbitrarily established different absolute levels at which the sample size was reduced, depending on the original, colony-specific number of individuals sampled (EdPo: 59 to 5 individuals sampled, by 2; AdCo: 25 to 5, by 2; InIs: 17 to 5, by 2).

For each indicator, we exploited the well-known statistical properties of the sample mean when computed from a sample of *n* independent and identically distributed observations—usually referred to as a random sample (e.g., Lehmann 1983)—to estimate the relative standard error (RSE) as the familiar measure of uncertainty of mean responses. RSE indeed reflects the average discrepancy of the sample mean to the true mean, expressed in percentage of the true mean. Because it is well known that in the case of random samples $RSE = \sigma/(\mu\sqrt{n})$, where μ and σ are respectively the unknown mean and standard deviation of each indicator, we exploited the mean response $\bar{x}$ and the standard deviation s obtained in the whole sample to estimate the relevant RSE for each level of sample size n using $s/(\bar{x}\sqrt{n})$. Although the level of acceptable uncertainty depends on the context, generally, estimated values of RSEs < 5% would suggest high precision in mean responses, RSEs between 5 and 15% would indicate moderate precision, and RSEs > 15% would reflect low precision (Shechtman 2013). The analysis thus simulates a situation in which sampling effort in the field is reduced progressively from the actual sample size, up to samples of a few smears collected, to inform at which sample size researchers would be able to assess mean responses with an acceptable level of precision. We performed rarefaction analysis by applying the above RSE formula through an ad hoc code in R version 4.3.0 (R Core Team 2023) provided in the Supplementary Material.

Statistical analysis: concordance between leukocyte counting methods

For each immunological indicator, we tested separately the bias, the linear association, the repeatability, and the inequality between the measurements obtained through the two counting methods. These represent different properties of bivariate data distributions that may provide alternative information about the concordance of the two counting methods; therefore, using appropriate statistical indices, we assessed all of them.

The bias represents the potential general differences in the values of each indicator measured with the two counting techniques. Thus, detecting an overall and significant difference in the measured parameter would indicate the presence of bias among the counting methodologies. We evaluated it using the paired Wilcoxon test. We used such a nonparametric test to minimize assumptions about the data distribution, employing a Monte Carlo significance value based on 99,999 random reassignments of values to the two treatments (counting techniques), within each paired individual.

The linear association reflects the degree to which the measurements made by the two techniques are linearly associated, and we tested it for comparison with a previous study (Menéndez-Blázquez et al. 2021). We used Pearson's correlation coefficient to assess linear association, with significance computed using a two-tailed t-test with n-2 degrees of freedom.

The repeatability represents either the consistency (aka additivity) or absolute agreement (aka identity) between two measurements (Schuck 2004; Koo and Li 2016). We used the intraclass correlation coefficient (ICC; Shrout and Fleiss 1979) and relevant 95% confidence interval to measure repeatability between the two counting techniques, particularly consistency in our case. Among the available versions of ICC (cf. Shrout and Fleiss 1979; Koo and Li 2016; Liljequist et al. 2019), we employed the so-called ICC's Model 3, namely the two-way mixed-effects model, following the Shrout and Fleiss's formula specification and relevant equation for confidence interval (Shrout and Fleiss 1979). This model of ICC assumes fixed 'raters', rather than raters sampled from a normal distribution. We used this model as it is recommended whenever the selected 'raters' (i.e., our counting techniques) are the only 'raters' of interest (Shrout and Fleiss 1979; Koo and Li 2016). Hence, this ICC model is the freest from assumptions about 'raters', only reflecting the reliability of the two specific 'raters' involved in the reliability experiment. We determined both the $ICC_{individual}$, and the ICC_{mean} to investigate consistency both in terms of the reliability of a single 'rater' and the mean value of the two 'raters', respectively (Koo and Li 2016). These ICC types correspond respectively to the ICC(3,1) and ICC(3,k), as *per* Shrout and Fleiss's notation (1979). The ICC can range from 0 to 1. Although any classification is arbitrary to some degree, and ICC value should be better evaluated based on the specific context (Liljequist et al. 2019), we followed the more conservative thresholds by Koo and Li (2016). Thus, we considered ICC < 0.50 as poor, 0.50 ≤ ICC < 0.75 as moderate, 0.75 ≤ ICC < 0.90 as good, and ICC ≥ 0.90 as excellent repeatability. We also extended the repeatability analysis under a mixed model framework to estimate and decompose the variability of the observed measurements explained by the counting technique from that due to the sample identity (Stoffel et al. 2017). If the data variability explained by the sample identity is high and significant, then the variability in measurements within each sample is low (or, in other words,

the measured parameter is consistent within each sample), in turn indicating weak or no effect of the counting technique. For each immunological parameter separately, we specified the mixed effect model reflecting the ICC design we used (i.e., assuming 'fixed raters'), whereby each immunological parameter (response variable) is modeled against the counting technique ('rater') as a fixed effect while entering the sample identity as a random intercept. Therefore, following Stoffel et al. (2017), we estimated the fraction of the total variance explained by the fixed effect, i.e., by the counting technique (corresponding to the marginal R^2 of the mixed model; Nakagawa & Schielzeth 2013), and the fraction of the total variance explained by the random effect, i.e., by the sample identity (corresponding to the conditional R^2 minus the marginal R^2 of the mixed model; Nakagawa & Schielzeth 2013), as well as their uncertainties (confidence intervals) via parametric bootstrapping ($n = 1000$ bootstraps).

Finally, the inequality mirrors the degree of mismatch of two variables and, although being strongly related to the concept of repeatability, it may provide an alternative indicator of the reliability between 'raters' (cf. Pan and Yin 2023). We used Theil's inequality index (U), also known as Theil's forecasting accuracy coefficient (Theil 1966), to assess the inequality of the measurements obtained with the different counting techniques. This index is normalized within the 0–1 interval, and it is commonly used to measure prediction error between observed and predicted values in forecast models. Lower values of U indicate lower prediction errors and thus greater equality of measurements (Theil 1966).

We conducted the above analyses by setting the significance level at $\alpha = 0.01$. We performed the paired Wilcoxon tests and the ICC analyses using the software Past 4.13 (Hammer et al. 2001). We ran the Pearson correlation analyses and calculated the Theil's indices respectively through the *cor.test* function and by applying the relevant Theil's formula in R version 4.3.0 (R Core Team 2023). Finally, we estimated the variability explained by sample identity and counting technique within the mixed model framework using the R package rptR (Stoffel et al. 2017).

Results and discussion

Blood smears analysis

The established Giemsa staining method effectively distinguishes nuclear and cytoplasmic components in both erythrocytes and leukocytes, enabling the reliable enumeration of ENAs and the clear identification of different WBC classes. This technique allows a clear distinction between erythrocytes and WBCs even at 40 × magnification, where different WBC classes can also be identified (Fig. 1a). At 100 × magnification, the nuclei of the erythrocytes can be seen even more clearly, and a more precise differentiation of the leukocyte classes is possible (Fig. 1b).

Figure 2 illustrates that ENAs are readily identified at 100 × magnification and were classified as MN (Fig. 2a), LN (Fig. 2b), TN (Fig. 2c), TL (Fig. 2d), BN (Fig. 2e), WC (Fig. 2f), KN (Fig. 2g), and UN (Fig. 2h), according to our previous studies (Olmastroni et al. 2019, 2024) and based on De Mas et al. (2015).

Notably, the optimal staining for ENAs does not compromise the quality of WBC staining, allowing for clear differentiation of H (Fig. 3a), E (Fig. 3b), L (Fig. 3c), B (Fig. 3d), and M (Fig. 3e). In addition, the characteristic granules of granulocytic leukocytes (H, E, B) can be recognized clearly

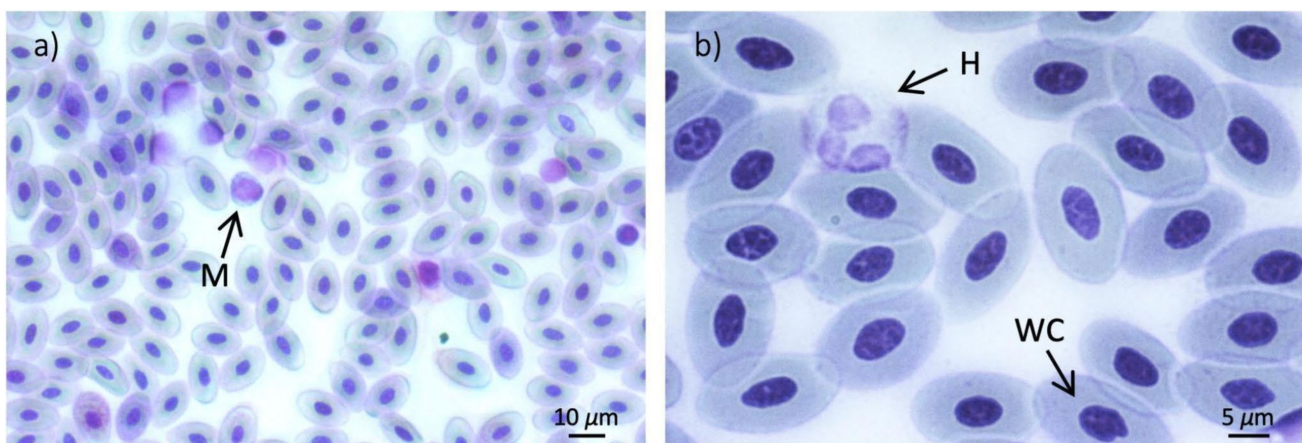


Fig. 1 Microscopic observation of blood smear of Adélie penguin (*Pygoscelis adeliae*) at different magnifications: **a**, a low magnification (40x) shows the overall field with blood cells (all erythrocytes except for one monocyte pointed by an arrow), **b**, a higher magnification (100 × with oil immersion) highlights the details of the cells, such as erythrocytes and WBCs (as the heterophil pointed by the left arrow), revealing any morphological abnormalities (as the nucleus with cavity pointed by the right arrow)

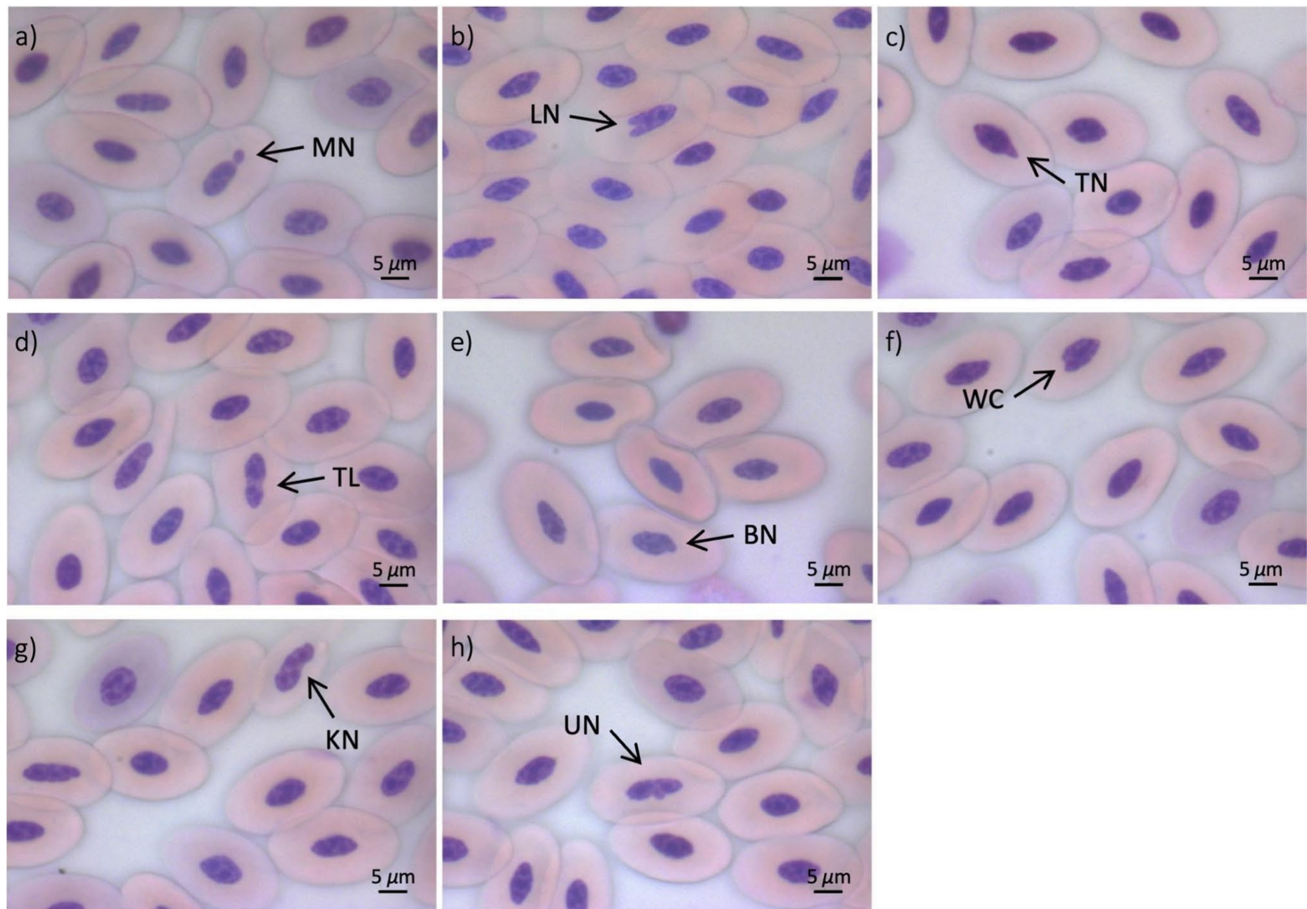


Fig. 2 Erythrocytes Nuclear Abnormalities (ENAs) in Adélie penguin blood smears: **a** micronucleus, **b** lobed nucleus, **c** tailed nucleus, **d** two-lobed nucleus, **e** budding nucleus, **f** nucleus with cavity, **g** kidney-shaped nucleus, **h** unknown nuclear malformation. All pictures were taken at 100× magnification

(Fig. 3a, b, d), further enhancing their classification into the three different categories.

The results obtained through our optimization of Giemsa staining are easily comparable with most of the data available in the literature, as Giemsa stain is usually the most commonly used in studies on the immune-hematological parameters of Antarctic seabirds (De Mas et al. 2015; Colominas-Ciurò et al. 2017; Palacios et al. 2018; Menéndez-Blázquez et al. 2021, and references therein). Moreover, the use of Giemsa allows for faster analysis compared to employing two different stains, as proposed by Kursa and Bezrukov (2008), which are required for the clear identification of both erythrocytes and leukocytes. This approach reduces both processing times and associated costs.

Given the wide range of methodologies available for blood smear staining, the implementation of a standardized protocol using a common dye as Giemsa and only one smear instead of two for the analysis of ENAs and WBCs represents a significant advancement in the assessment of hematological parameters. Such a standard enhances diagnostic efficiency and accuracy while simultaneously reducing the blood volume required for analysis and reducing the number of animals sampled.

Sample size-rarefaction to inform field sampling effort

A previous study conducted in the same areas showed that the variability in mean responses of some immune-hematological parameters considered to be major proxies for stress, namely total number of ENAs/10,000 erythrocytes, total number of WBCs/10,000 erythrocytes, percentage of MNs, and H/L ratio, increased according to a simulated reduction in sample size (Olmastroni et al. 2024). Nevertheless, that evaluation was conditioned on the specific sampled data thus, the extent to which the uncertainty of mean responses changes with reduced sampling effort remains unknown in general situations, e.g., for means computed from different sample data regarding these and other immune-hematological parameters.

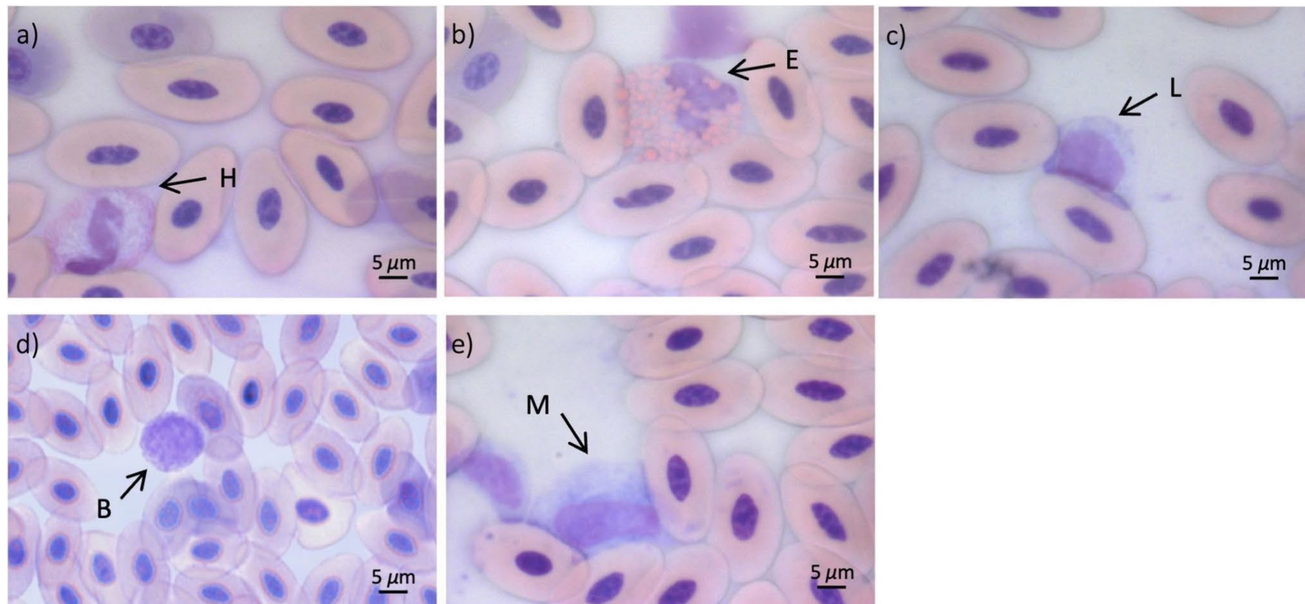


Fig. 3 Representative leukocyte types in Adélie penguin blood smears: **a** heterophils, **b**, eosinophils, **c** lymphocytes, **d** basophils, **e** monocytes. Pictures taken at 100 × magnification

Here, we evaluated the precision of the mean responses for each blood indicator to assess the reliability of the counts. The precision of the mean responses decreased non-linearly with decreasing sample size (Fig. 4).

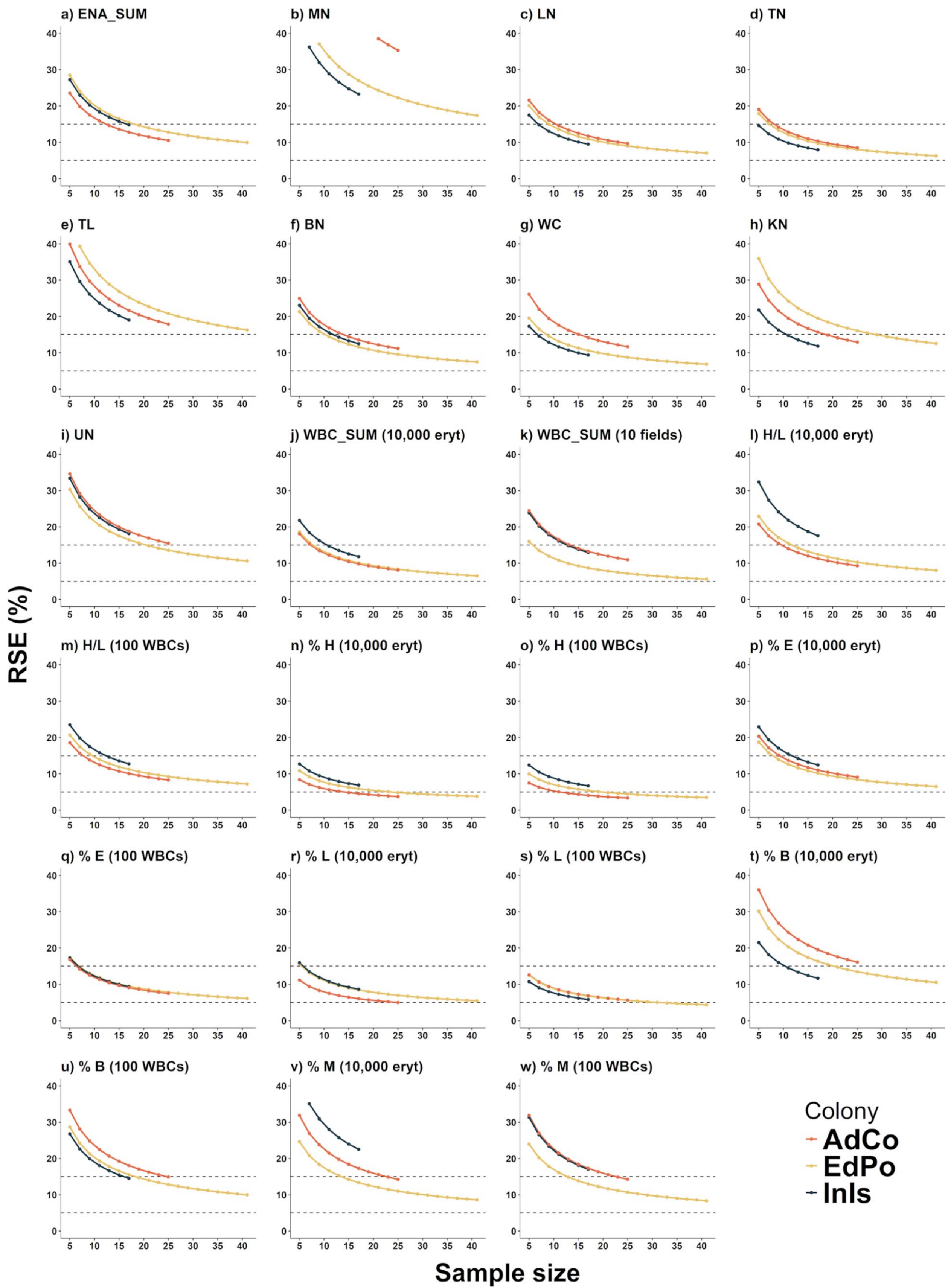
In general, the estimates of hematological parameters are characterized by low error values for almost all parameters considered. This is evident from the convergence of the curves (Fig. 4) toward low RSE values as the sample size increases. This means that increasing the number of samples provides marginal benefits to the precision of the estimates. This suggests that it is possible to reduce the number of sampled individuals while still obtaining a high degree of precision of hematological parameter estimates, mitigating handling and stress to the colony, i.e., at the population level (Zurlo et al. 1996; Wells et al. 2023).

However, it is worth noting that less frequent parameters with higher variability, such as MN and TL, show less precise estimates, especially with small sample sizes. This suggests that these parameters are less stable or more influenced by external factors and require a larger number of samples to obtain reliable estimates. More specifically, the sums of ENAs and WBCs are precisely estimated for all three colonies with a sample size of about 20–25, while the parameters MN and TL show greater uncertainty. For most other parameters, the estimates stabilize with approximately 15–20 samples (Fig. 4).

Our findings showed that a strategic sampling approach, targeting around 20 individuals, can yield accurate hematological data for most markers. This reduction is crucial to optimize the sampling effort and minimize the number of penguins handled during sample collection: although the handling impact on each individual would not change, a lower number of target individuals possibly reduces the overall stress and the risk of injury at the population level. This is particularly relevant in the current context of emerging threats to Antarctic penguins, including climate change (Trathan et al. 2014) and the increasing risk of wildlife diseases such as avian influenza (Aguado et al. 2024). However, the specific research objective is paramount in determining optimal sample size. For instance, robust genotoxicity studies focusing on MN require a substantially larger sample to ensure adequate statistical power. Conversely, investigations primarily targeting leukocyte counts can achieve reliable results with a reduced sampling effort.

Concordance between WBCs counting methods

Generally, the two counting techniques to assess WBCs provided similar values for most immunological parameters. The total WBC number assessed by both counting leukocytes in 10 fields and among 10,000 erythrocytes showed a moderate linear association and a moderate-good repeatability for the values (Table 1; Fig. 5a), albeit relatively lower than previous results by Menéndez-Blázquez et al. (2021). Additionally, an acceptable level of precision was estimated with a sampling effort of, on average, approximately 9 blood smear samples for both methods (Fig. 4j-k). However, it is worth noting that the number of leukocytes counted in 10 fields significantly differed from those counted among 10,000 erythrocytes, slightly biasing


a) ENA_SUM
b) MN
c) LN
d) TN
e) TL
f) BN
g) WC
h) KN
i) UN
j) WBC_SUM (10,000 eryt)
k) WBC_SUM (10 fields)
l) H/L (10,000 eryt)
m) H/L (100 WBCs)
n) % H (10,000 eryt)
o) % H (100 WBCs)
p) % E (10,000 eryt)
q) % E (100 WBCs)
r) % L (10,000 eryt)
s) % L (100 WBCs)
t) % B (10,000 eryt)
u) % B (100 WBCs)
v) % M (10,000 eryt)
w) % M (100 WBCs)
RSE (%)
Sample size
Colony
AdCo
EdPo
InIs

◂**Fig. 4** Variation in the percentage relative standard error (% RSE) of the mean colony-specific response (green: EdPo; red: AdCo; blue: InIs) for all the 23 immune-hematological parameters measured in our study according to a reduction in sample size. Horizontal dashed lines indicate thresholds commonly suggesting high (RSE < 5%) and moderate precision (5% ≤ RSE < 15%). For parameters with cases exceeding RSEs > 40% (%MN, %TL, %M/10,000 erythrocytes), lines are cut out from the figure for better representation

the mean value downwards (-22%; Table 1; Fig. 5a). In this case, the variance explained by the counting technique was much higher than for the other parameters, albeit being low in general terms (approximately 7%; Table 1). This could be due to the uneven density of erythrocytes of the analyzed blood smears for each field, which should not guarantee the presence of 10,000 erythrocytes in 10 fields.

Values of WBC composition determined by counting the leukocyte number in each class out of 100 WBCs showed a strong positive correlation with those determined with the other method (counting leukocytes out of 10,000 erythrocytes, during ENAs count), with Pearson's r values between 0.86 and 0.95 (Table 1; Fig. 5c-n), nearing perfect linear association. Furthermore, a good–excellent repeatability (with ICCs spanning 0.85–0.97 and sample identity explaining from 85%-to-95% variance in measurements, with a negligible variance explained by the counting technique), as well as low inequality (with Theil's U indices between 0.04 and 0.12), were found (Table 1), thus approaching perfect consistency, and equality. Indeed, mean values of WBC indicator composition did not generally differ between the two counting methods except for % basophils, which was slightly underestimated when counting them out of 100 WBCs (Table 1; Fig. 5i-j). Both counting methods demonstrated a similar level of acceptable precision for each class of WBCs and the H/L ratio (Fig. 4l-w). It is worth noting that Menéndez-Blázquez et al. (2021) demonstrated that 10 microscope fields observed at 40 × magnification contain approximately 10,000 erythrocytes. In light of this, the present findings suggest that especially the leukocyte profile (i.e., the percentage of each WBC class) can be rapidly and reliably obtained by analyzing 10 microscope fields at 40 × magnification, effectively reducing analysis time when the focus is solely on leukocyte assessment.

Taken together, these results demonstrate the effectiveness of directly counting WBCs during the ENAs count (over 10,000 erythrocytes) and highlight the redundancy of repeating the count in different ways, enabling a reduction in analysis duration. Thus, in a simultaneous count of ENAs and WBCs, the use of a WBCs count of 10,000 erythrocytes at 100 × (oil immersion) magnification instead of counting them in 10 fields at 40 × magnification can significantly

Table 1 Mean ± standard deviation of immunological parameters measured by the two counting methodologies (WBCs in 10,000 erythrocytes vs WBCs in 10 fields, for total number of WBCs, or in 100 WBCs, for other parameters), showing their general differences as assessed by the Wilcoxon paired test, as well as the linear association (Pearson's r and p-value, in parenthesis), repeatability (individual and mean ICCs and relevant 95% confidence interval, in brackets), variance fractions explained by sample identity and counting methodology (and relevant 95% confidence interval, in brackets), and inequality (Theil's U index) between techniques

Parameter	Mean ± St. Dev.*	Mean ± St. Dev **	Test statistic and p-value	Pearson's correlation	$ICC_{individual}$	ICC_{mean}	% Variance explained ***	% Variance explained ****	U
Total number of WBCs	78.95 ± 33.79	61.25 ± 28.15	$W = 3087.5$ ($p < 0.001$)	$r = 0.61$ ($p < 0.001$)	0.61 [0.45; 0.72]	0.75 [0.62; 0.84]	56.0 [42.6; 67.7]	7.6 [3.7; 13.1]	0.21
% H	42.24 ± 10.04	43 ± 9.53	$W = 2250.0$ ($p = 0.038$)	$r = 0.95$ ($p < 0.001$)	0.95 [0.92; 0.97]	0.97 [0.96; 0.98]	94.8 [91.9; 96.5]	0.2 [0; 0.6]	0.04
% E	20.13 ± 8.97	20.44 ± 7.83	$W = 1978.5$ ($p = 0.389$)	$r = 0.93$ ($p < 0.001$)	0.92 [0.89; 0.95]	0.96 [0.94; 0.97]	92.4 [88.5; 94.9]	0 [0; 0.3]	0.07
% L	26.52 ± 8.63	25.85 ± 7.01	$W = 2141.5$ ($p = 0.169$)	$r = 0.87$ ($p < 0.001$)	0.85 [0.78; 0.90]	0.92 [0.88; 0.95]	85.1 [78.1; 90.1]	0.2 [0; 1.0]	0.08
% B	5.97 ± 4.00	5.50 ± 3.69	$W = 2402.5$ ($p = 0.006$)	$r = 0.92$ ($p < 0.001$)	0.92 [0.88; 0.95]	0.96 [0.94; 0.97]	91.7 [87.4; 94.6]	0.4 [0.1; 1.2]	0.12
% M	5.14 ± 3.49	5.21 ± 3.23	$W = 1825.5$ ($p = 0.572$)	$r = 0.92$ ($p < 0.001$)	0.92 [0.87; 0.94]	0.96 [0.93; 0.97]	91.6 [87.1; 94.5]	0 [0; 0.3]	0.11
H/L ratio	1.86 ± 1.03	1.85 ± 0.86	$W = 2081.0$ ($p = 0.123$)	$r = 0.86$ ($p < 0.001$)	0.85 [0.78; 0.90]	0.92 [0.88; 0.95]	85.1 [78.3–89.9]	0 [0–0.4]	0.12

*WBCs in 10,000 erythrocytes

**WBCs in 10 fields, for the total number of WBCs, or in 100 WBCs, for other parameters

***variance explained by the sample identity

****variance explained by the counting technique

The sample size is $n = 87$ individuals for total WBCs and $n = 86$ individuals for the other parameters

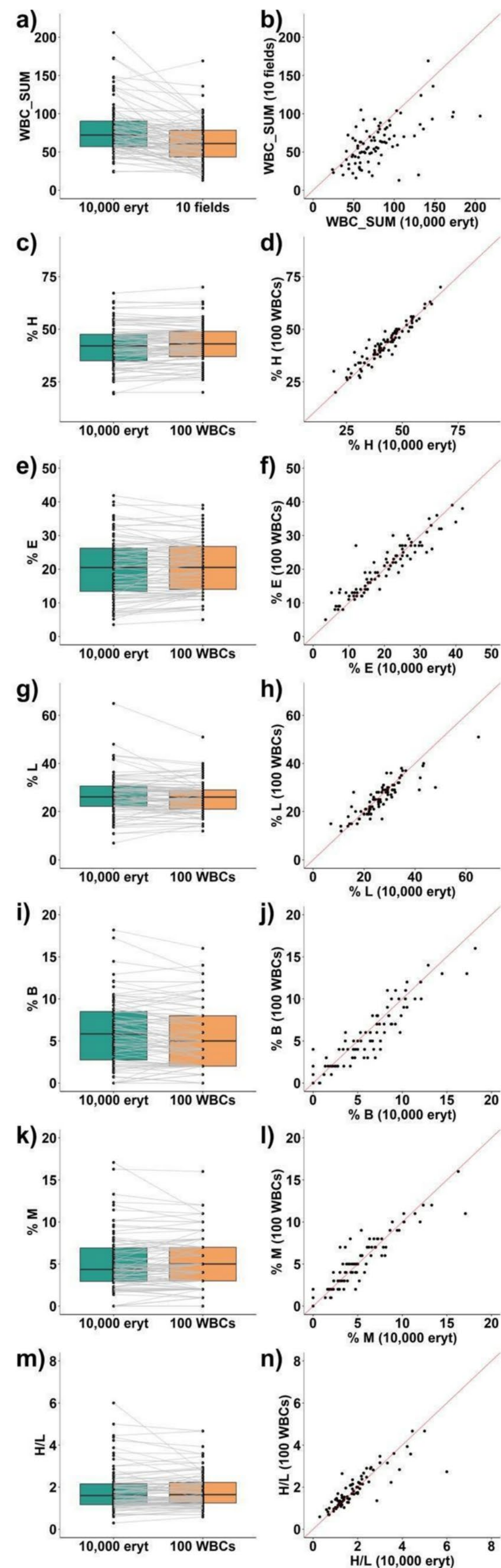


Fig. 5 **a-b** Total number of WBCs; **c-d** % H; **e–f** % E; **g-h**: % L; **i-j** % B; **k-l** % M; **m–n** H/L ratio. The left panel (**a**, **c**, **e**, **g**, **i**, **k**, **m**) shows differences between values of paired samples measured with the two counting methods (black dots connected by gray lines); a modified boxplot showing the median (horizontal thick line) and interquartile range (box limits) of measured values is also plotted for each indicator and counting method. The right panel (**b**, **d**, **f**, **h**, **j**, **l**, **n**) shows the relationship between values of the same indicator measured with the two counting methods (WBCs in 10,000 erythrocytes vs WBCs in 10 fields, for the total number of WBCs, or in 100 WBCs, for other parameters). The diagonal, 1:1 red line reflects perfect repeatability (variance explained by the sample identity = 100%) and the lowest inequality (Theil's U index = 0) between measurements

save time when analyzing a large number of penguin blood smears.

Conclusion

In this study, we tested an alternative methodology for blood smear cell counts by adapting an existing approach to simultaneously assess ENAs and WBCs in Adélie penguins. A refined staining and counting method improved efficiency and reduced the number of samples needed to obtain a high degree of precision of hematological parameter estimates. WBC counts within ENA analysis (10,000 erythrocytes) were generally comparable to traditional methods, reducing both smear number and analysis time without losing accuracy. The only parameter that failed to reach the optimal accuracy level was the total WBC count, which showed merely moderate repeatability between the two counting methods (among 10,000 erythrocytes vs in 10 subsequent fields). Rarefaction analysis indicates that reliable estimates for most parameters can be obtained with approximately 20 individuals per parameter, significantly reducing blood smears needed and minimizing fieldwork efforts. This approach also helps standardize the sample size needed for robust analysis. However, larger sample sizes are still required for specific studies focusing on micronucleus assessments to ensure statistical robustness. Additionally, our investigation could be improved through an assessment of the technical precision and repeatability of the immune-hematological parameters estimated from multiple smears on the same individual penguin.

Studying the stress response of key organisms is crucial for monitoring the Antarctic population and ecosystem health, also in light of the growing concerns over emerging diseases that may harm Antarctic wildlife (Grimaldi et al. 2015; Aguado et al. 2024). The data presented herein provide a basis for the development of optimized sampling and analysis protocols for the monitoring of hematological

parameters in penguins. Emphasizing the 3Rs principle, particularly Reduction, this methodology minimizes disturbance while maintaining analytical precision. We suggest its adoption to enhance Antarctic ecosystem monitoring, offering a cost-effective and less invasive tool for long-term conservation studies.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00300-025-03422-5.

Acknowledgements The authors are grateful to Dr. Lucia Burrini for her invaluable help during fieldwork. We would also like to thank Dr. Claudia Faleri for her assistance in the acquisition of the microscope images.

Author contributions IC and SO designed the study and secured funding, NF and SO collected biological samples, SS developed the adjusted staining method and conducted microscopic count analysis, NF performed statistical analysis, all authors analysed the data, SS and NF curated and visualised the data, and all authors contributed to writing the manuscript.

Funding Open access funding provided by Università degli Studi di Siena within the CRUI-CARE Agreement. This research was funded by Ministry of University and Research (MUR) and the National Research Council Italy (CNR) through Programma Nazionale di Ricerche in Antartide (PNRA) with project #PNRA2016_0004 PenguinERA.

Data availability Data will be made available on request.

Declarations

Conflict of interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Ethical approval All applicable international, national, and /or institutional guidelines for the care and use of animals were followed. All procedures performed in studies involving animals followed the ethical standards of SCAR's Code of Conduct for the Use of Animals for Scientific Purposes in Antarctica (https://scar.org/~documents/policy/antarctic-treaty/atcm-xxxiv-and-cep-xiv-2011/atcm34-ip053?layout=default). Sampling activities were carried out in the framework of the project #PNRA2016_0004 PenguinERA and under PNRA Permit for sampling: "Olmastroni campionamento/interferenza: XXXIII/C5" and under PNRA. Permit for carrying out activities in ASPA n. 165: "Olmastroni ASPA165: XXXIII/A5".

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Aguado B, Begeman L, Günther A, Iervolino M et al (2024) Searching for high pathogenicity avian influenza virus in Antarctica. Nat Microbiol 9:3081–3083. https://doi.org/10.1038/s41564-024-01868-7

Ainley DG, 2002. The Adélie Penguin. In: bellwether of climate change. Columbia University Press, New York

Banyard AC, Bennison A, Byrne AMP et al (2024) Detection and spread of high pathogenicity avian influenza virus H5N1 in the Antarctic region. Nat Commun 15:7433. https://doi.org/10.1038/s41467-024-51490-8

Bestley S, Ropert-Coudert Y, Bengtson Nash S et al (2020) Marine ecosystem assessment for the Southern Ocean: birds and marine mammals in a changing climate. Front Ecol Evol 8:566936. https://doi.org/10.3389/fevo.2020.566936

Boersma PD (2008) Penguins as marine sentinels. BioSci 58(7):597–607. https://doi.org/10.1641/B580707

Campbell TW (2015) Avian hematology and cytology, 3rd edn. Iowa State University Press, Ames

Coetzee BWT, Chown SL (2016) A meta-analysis of human disturbance impacts on Antarctic wildlife. Biol Rev 91:578–596. https://doi.org/10.1111/brv.12184

Colominas-Ciuro R, Santos M, Coria N, Barbosa A (2017) Reproductive effort affects oxidative status and stress in an Antarctic penguin species: an experimental study. PLoS ONE 12(5):e0177124. https://doi.org/10.1371/journal.pone.0177124

Croxall JP, Trathan PN, Murphy EJ (2002) Environmental change and Antarctic seabird populations. Science 297:1510–1514. https://doi.org/10.1126/science.1071987

D'Amico VL, Coria N, Palacios MG et al (2014) Physiological differences between two overlapped breeding Antarctic penguins in a global change perspective. Polar Biol 39:57–64

D'Amico VL, Bertellotti M, Benzal J et al (2016) Leukocyte counts in different populations of Antarctic *Pygoscelid* penguins along the Antarctic peninsula. Polar Biol 39:199–206. https://doi.org/10.1007/s00300-015-1771-3

Davis AK (2005) Effect of handling time and repeated sampling on avian white blood cell counts. J Field Ornithol 76(4):334–338. https://doi.org/10.1648/0273-8570-76.4.334

Davis AK, Maney DL, Maerz JC (2008) The use of leukocyte profiles to measure stress in vertebrates: a review for ecologists. Funct Ecol 22(5):760–772. https://doi.org/10.1111/j.1365-2435.2008.01467.x

De Mas E, Benzal J, Merino S et al (2015) Erythrocytic abnormalities in three Antarctic penguin species along the Antarctic Peninsula: biomonitoring of genomic damage. Polar Biol 38:1067–1074. https://doi.org/10.1007/s00300-015-1667-2

Fattorini N, Olmastroni S (2021) Pitfalls and advances in morphometric sexing: insights from the Adélie penguin *Pygoscelis adeliae*. Polar Biol 44(8):1563–1573. https://doi.org/10.1007/s00300-021-02893-6

Fraser WR, Patterson-Fraser DL, Ribic CA, Schofield O, Ducklow H (2013) A nonmarine source of variability in Adélie penguin demography. Oceanography 26(3):207–209. https://doi.org/10.5670/oceanog.2013.64

Gallo L, Vanstreels RET, Cook RA et al (2019) Hematology, plasma biochemistry, and trace element reference values for free-ranging adult magellanic penguins (*Spheniscus magellanicus*). Polar Biol 42:733–742. https://doi.org/10.1007/s00300-019-02467-7

Grimaldi WW, Seddon PJ, Lyver PO, Nakagawa S, Tompkins DM (2015) Infectious diseases of Antarctic penguins: current status and future threats. Polar Biol 38:591–606. https://doi.org/10.1007/s00300-014-1632-5

Gross WB, Siegel HS (1983) Evaluation of the heterophil/lymphocyte ratio as a measure of stress in chickens. Avian Dis 27(4):972–979. https://doi.org/10.2307/1590198

Hammer Ø, Harper DAT, Ryan PD (2001). PAST: paleontological statistics software package for education and data analysis. PE 4:1–9

Hawkey CM, Horsley DT, Keymer IF (1989) Haematology of wild penguins (Sphenisciformes) in the Falkland Islands.". Avian Pathol 18(3):495–502. https://doi.org/10.1080/03079458908418621

Hayashi M (2016) The micronucleus test—most widely used in vivo genotoxicity test—. Genes Environ 38:18. https://doi.org/10.1186/s41021-016-0044-x

Heddle JA, Hite M, Kirkhart B et al (1983) The induction of micronuclei as a measure of genotoxicity: a report of the Us Environmental Protection Agency gene-tox program. Mutat Res 123(1):61–118. https://doi.org/10.1016/0165-1110(83)90047-7

Kerry KR, Riddle M (2009) Health of Antarctic wildlife: a challenge for science and policy. Springer, Berlin

Koo TK, Li MY (2016) A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 15:155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Kophamel S, Illing B, Ariel E et al (2022) Importance of health assessments for conservation in noncaptive wildlife. Conserv Biol 36:e13724. https://doi.org/10.1111/cobi.13724

Kursa M, Bezrukov V (2008) Health status in an Antarctic top predator: Micronuclei frequency and white blood cell differentials in the South Polar skua (*Catharacta maccormicki*). PolF 77(1):1–5

Le Maho Y, Whittington J, Hanuise N et al (2014) Rovers minimize human disturbance in research on wild animals. Nat Methods 11:1242–1244. https://doi.org/10.1038/nmeth.3173

Lehmann EL (1983) Theory of point estimation. Wiley, New York

Liljequist D, Elfving B, Skavberg Roaldsen K (2019) Intraclass correlation–a discussion and demonstration of basic features. PLoS ONE 14(7):e0219854

Maceda-Veiga A, Figuerola J, Martínez-Silvestre A et al (2015) Inside the redbox: applications of haematology in wildlife monitoring and ecosystem health assessment. Sci Total Environ 514:322–332. https://doi.org/10.1016/j.scitotenv.2015.02.004

Menéndez-Blázquez J, Soto F, Negrete J et al (2021) Leukocyte counts in blood smears of Antarctic seals and penguins: a new less time-consuming method. Polar Biol 44(11):2195–2198. https://doi.org/10.1007/s00300-021-02950-0

Müller C, Jenni-Eiermann S, Jenni L (2011) Heterophils/Lymphocytes-ratio and circulating corticosterone do not indicate the same stress imposed on Eurasian kestrel nestlings. Funct Ecol 25(3):566–576. https://doi.org/10.1111/j.1365-2435.2010.01816.x

Nakagawa S, Schielzeth H (2013) A general and simple method for obtaining R2 from generalized linear mixed-effects models. Methods Ecol Evol 4:133–142

Olmastroni S, Pompeo G, Jha AN et al (2019) Erythrocytes nuclear abnormalities and leukocyte profile of the immune system of Adélie penguins (*Pygoscelis adeliae*) breeding at Edmonson Point, Ross Sea, Antarctica. Polar Biol 42:1343–1352. https://doi.org/10.1007/s00300-019-02522-3

Olmastroni S, Simonetti S, Fattorini N et al (2024) Living in a challenging environment: monitoring stress ecology by non-destructive methods in an Antarctic seabird. Sci Total Environ 922:171249

Palacios MJ, Valera F, Colominas-Ciuró R, Barbosa A (2018) Cellular and humoral immunity in two highly demanding energetic life stages: reproduction and moulting in the chinstrap penguin. J Ornithol 159:283–290

Pan T, Yin Y (2023) An application of Theil indexes for the interrater reliability: a comparison with intraclass correlations. Chinese/English J Educ Meas Eval 4(2):1

Polito MJ, Clucas GV, Hart T, Trivelpiece WZ (2012) A simplified method of determining the sex of *Pygoscelis* penguins using bill measurements. Mar Ornithol 40:89–94

R Core Team (2023) R: A Language and environment for statistical computing. R Foundation for Statistical Computing, Vienna

Ropert-Coudert Y, Chiaradia A, Ainley D et al (2019) Happy feet in a hostile world? The future of penguins depends on proactive management of current and expected threats. Front Mar Sci 6:248. https://doi.org/10.3389/fmars.2019.00248

Samour J (2006) Diagnostic value of hematology. Clinical Avian Med 2:587–610

Schuck P (2004) Assessing reproducibility for interval data in health-related quality of life questionnaires: which coefficient should be used? Qual Life Res 13:571–585

Shechtman O (2013) The coefficient of variation as an index of measurement reliability. In: Doi S, Williams G (eds) Methods of clinical epidemiology. Springer Series on Epidemiology and Public Health, Berlin

Shrout PE, Fleiss JL (1979) Intraclass correlations: uses in assessing rater reliability. Psychol Bull 86:420–428

Southwell C, Emmerson L, Takahashi A et al (2017) Large-scale population assessment informs conservation management for seabirds in Antarctica and the Southern Ocean: a case study of Adélie penguins. GECCO 9:104–115. https://doi.org/10.1016/j.gecco.2016.12.004

Stoffel MA, Nakagawa S, Schielzeth H (2017) rptR: repeatability estimation and variance decomposition by generalized linear mixed-effects models. Methods Ecol Evol 8:1639–1644. https://doi.org/10.1111/2041-210X.12797

Theil H (1966) Applied economic forecasting. North Holland Publishing Co., Amsterdam

Trathan PN, García-Borboroglu P, Boersma D et al (2014) Pollution, habitat loss, fishing, and climate change as critical threats to penguins. Conserv Biol 29(1):31–41. https://doi.org/10.1111/cobi.12349

Turcu M-C, Paștiu AI, Bel L-V, Pusta DL (2023) Minimally invasive sampling methods for molecular sexing of wild and companion birds. Animals 13(21):3417. https://doi.org/10.3390/ani13213417

Vleck CM, Vertalino N, Vleck D, Bucher TL (2000) Stress, corticosterone, and heterophil to lymphocyte ratios in free-living Adélie penguins. Condor 102(2):392–400

Wells MR, Philips A, Lea MA, Carver S (2023) The potential of ecologists to enhance our understanding of seabird health. Mar Ornithol 51:11–22. https://doi.org/10.5038/2074-1235.51.1.1504

World Health Organization (WHO) (2016) Quality control of Giemsa stock solution and buffered water (No.WHO/HTM/GMP/MM/SOP/2016.03 c). World Health Organization

Zúñiga-González G, Torres-Bugarın O, Luna-Aguirre J et al (2000) Spontaneous micronuclei in peripheral blood erythrocytes from 54 Animal Species (Mammals, Reptiles and Birds): Part two. Mutat Res 467(1):99–103. https://doi.org/10.1016/s0165-1218(96)90056-7

Zurlo J, Rudacille D, Goldberg AM (1996) The three Rs: the way forward. Environ Health Perspect 104(8):878–880. https://doi.org/10.1289/ehp.96104878

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.