Research Article

Clinical Profiles of Selected Biomarkers Identifying Infection and Cancer Patients: A Gorzów Hospital Example

Katarzyna Brzeźniakiewicz-Janus,1 Marcus Daniel Lancé,2 Andrzej Tukiendorf3,5, Miroslaw Franków,1 Joanna Rupa-Matysek,4 Edyta Wolny-Rokicka6,5 and Lidia Gil4

1Department of Hematology, Multi-Specialist Hospital Gorzów Wielkopolski, Faculty of Medicine and Health Science, University of Zielona Góra, Gorzów Wielkopolski, Poland
2Department of Anesthesiology, Intensive Care Unit and Perioperative Medicine, Hamad Medical Corporation, Doha, Qatar
3Department of Public Health, Wrocław Medical University, Wrocław, Poland
4Department of Hematology and Bone Marrow Transplantation, Poznań University of Medical Sciences, Poznań, Poland
5Department of Surgery and Oncology, Multi-Specialist Hospital Gorzów Wielkopolski, Faculty of Medicine and Health Science, University of Zielona Góra, Gorzów Wielkopolski, Poland

Correspondence should be addressed to Andrzej Tukiendorf; andrzej.tukiendorf@gmail.com

Received 22 June 2019; Accepted 6 August 2019; Published 2 September 2019

Academic Editor: Giuseppe Murdaca

Copyright © 2019 Katarzyna Brzeźniakiewicz-Janus et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Introduction. Many pathobiological processes that manifest in a patient’s organs could be associated with biomarker levels that are detectable in different human systems. However, biomarkers that promote early disease diagnosis should not be tested only in personalized medicine but also in large-scale diagnostic evaluations of patients, such as for medical management. Objective. We aimed to create an easy algorithmic risk assessment tool that is based on obtainable “everyday” biomarkers, identifying infection and cancer patients. Patients. We obtained the study data from the electronic medical records of 517 patients (186 infection and 331 cancer episodes) hospitalized at Gorzów Hospital, Poland, over a one and a half-year period from the 1st of January 2017 to the 30th of June 2018. Methods and Results. A set of consecutive statistical methods (cluster analysis, ANOVA, and ROC analysis) was used to predict infection and cancer. For in-hospital diagnosis, our approach showed independent clusters of patients by age, sex, MPV, and disease fractions. From the set of available “everyday” biomarkers, we established the most likely bioindicators for infection and cancer together with their classification cutoffs. Conclusions. Despite infection and cancer being very different diseases in their clinical characteristics, it seems possible to discriminate them using “everyday” biomarkers and popular statistical methods. The estimated cutoffs for the specified biomarkers can be used to allocate patients to appropriate risk groups for stratification purposes (medical management or epidemiological administration).

1. Introduction

Many pathobiological processes that manifest in a patient’s organs could be associated with biomarker levels that are detectable in different human systems (e.g., reactive oxygen species in the systemic circulation [1] or apolipoprotein E in Alzheimer’s disease [2]). The accessibility of biomarkers that promote early disease diagnosis in both symptomatic and asymptomatic patients, such as infection and cancer, could lead to the application of personalized medicine in many serious diseases. However, a biomarker should be tested in epidemiological diagnostics for its qualifications. An ideal biomarker should be reliable objectively distinguishing between normal biologic and pathological processes (regardless of methodological difficulties and expenses). In a large-scale diagnostic evaluation of patients for medical management, for example, a biomarker should also be standardized (accurately measured in different laboratories), timely (not time-consuming), practical (comprehensible), and inexpensive (cost-effective) [2].
There has been increasing interest in identifying disease-related biomarkers to predict pathogenic processes. However, there is still minimal information about how biomarkers relate to disease progression, severity, or response to therapy.

For instance, in cardiovascular disease, it remains uncertain whether the association between low vitamin D status and arterial disease is causal or whether it is just a side factor. The results of the study in [3] demonstrated that low vitamin D status was a risk factor for the severity of arterial disease. Similarly, triglyceride-cholesterol (TG-C) imbalance across lipoprotein subclasses predicts diabetic kidney disease and mortality in type 1 diabetes. In this report, low C, low TG-C ratio, and high TG-C ratio subclasses represented three phenotypes associated with increasing patient mortality (<3%, 6%, and 40%, respectively) [4]. Long-term total and cardiovascular mortality in patients undergoing coronary angiography was associated with mean LDL particle diameter (large >16.8 nm, intermediate 16.5–16.8 nm, and small <16.5 nm) [5]. The study authors conclude that both large and small LDL diameters were independently associated with increased risk mortality of all causes compared with LDL of intermediate size [5]. However, circulating lipids were also associated with alcoholic liver cirrhosis and represented potential biomarkers for risk assessment [6]. In this study, 6 of the 25 lipid classes and subclasses were significantly associated with alcoholic liver cirrhosis. Based on the multivariate classification models, the authors established that the addition of lipid measurements to the clinical characteristics of patients resulted in improved ability to estimate the severity of liver cirrhosis [6]. These data were confirmed by [7], who suggested that metabolomic profiles based on diacyl-phosphatidylcholines, lysophosphatidylcholines, ether lipids, and sphingolipids were a new class of biomarkers for excess alcohol intake and had potential for future epidemiological and clinical studies. A study of viral haemorrhagic fever [8] highlights that the most common clinical and laboratory profiles are very helpful for diagnosis of dengue viral infections. In the population of 102 dengue patients, elevated levels of CRP (mg/L) 453 7.56 9.18 3.2 0.01 46.8 were very helpful for diagnosis of dengue viral infections. Moreover, the study authors conclude that both large and small LDL diameters were independently associated with increased risk mortality of all causes compared with LDL of intermediate size [5]. However, circulating lipids were also associated with alcoholic liver cirrhosis and represented potential biomarkers for risk assessment [6]. In this study, 6 of the 25 lipid classes and subclasses were significantly associated with alcoholic liver cirrhosis. Based on the multivariate classification models, the authors established that the addition of lipid measurements to the clinical characteristics of patients resulted in improved ability to estimate the severity of liver cirrhosis [6]. These data were confirmed by [7], who suggested that metabolomic profiles based on diacyl-phosphatidylcholines, lysophosphatidylcholines, ether lipids, and sphingolipids were a new class of biomarkers for excess alcohol intake and had potential for future epidemiological and clinical studies. A study of viral haemorrhagic fever [8] highlights that the most common clinical and laboratory profiles are very helpful for diagnosis of dengue viral infections. In the population of 102 dengue patients, elevated levels of AST in 46 (45.1%) and ALT in 18 (17.6%) patients were found to be among the most common clinical manifestations of the infection. This finding could alert physicians to the likelihood of dengue virus infections in the study area. In other words, a single biomarker could be indicative of several pathologies that do not necessarily have a direct relationship. Depending on the combinations of markers, different diseases could be detected by the biomarkers.

In this paper, we show how a combined statistical methodology based on “everyday” diagnostic investigations contributes to differentiating infection from cancer patients at an early stage. With our blinded data analysis approach, a new diagnostic filter could easily categorize the patients into appropriate disease categories. We used infection and cancer patients following a priori clinical assumptions that these diseases are very different in their clinical characteristics. For this reason, we believe they will contribute to more reliable cutoffs of the possible biomarkers to distinguish early symptoms, and for stratification with a few biomarkers, the chance of classifying patients into the wrong category will be diminished.

### Table 1: Description statistics of selected biomarkers.

<table>
<thead>
<tr>
<th>Biomarker</th>
<th>n</th>
<th>Mean</th>
<th>SD</th>
<th>Me</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>MPV (fL)</td>
<td>517</td>
<td>10.2</td>
<td>1.03</td>
<td>10.1</td>
<td>8.0</td>
<td>14.0</td>
</tr>
<tr>
<td>WBC (10^9/L)</td>
<td>515</td>
<td>12.8</td>
<td>12.3</td>
<td>10.5</td>
<td>0.12</td>
<td>200</td>
</tr>
<tr>
<td>Neutrophils (10^9/L)</td>
<td>221</td>
<td>7.81</td>
<td>7.03</td>
<td>6.21</td>
<td>0.01</td>
<td>47.7</td>
</tr>
<tr>
<td>Lymphocytes (10^9/L)</td>
<td>219</td>
<td>3.16</td>
<td>12.6</td>
<td>1.35</td>
<td>0.05</td>
<td>174</td>
</tr>
<tr>
<td>Monocytes (10^9/L)</td>
<td>219</td>
<td>1.19</td>
<td>3.15</td>
<td>0.75</td>
<td>0.01</td>
<td>33.6</td>
</tr>
<tr>
<td>RBC (ml/mmm3)</td>
<td>517</td>
<td>4.16</td>
<td>0.85</td>
<td>4.27</td>
<td>1.10</td>
<td>6.37</td>
</tr>
<tr>
<td>Haematocrit (%)</td>
<td>517</td>
<td>35.2</td>
<td>6.5</td>
<td>35.7</td>
<td>9.7</td>
<td>53.3</td>
</tr>
<tr>
<td>Haemoglobin (g/dL)</td>
<td>517</td>
<td>12</td>
<td>2.35</td>
<td>12.1</td>
<td>3.2</td>
<td>18.9</td>
</tr>
<tr>
<td>MCV (fL)</td>
<td>517</td>
<td>85.4</td>
<td>6.89</td>
<td>85.8</td>
<td>67.0</td>
<td>105</td>
</tr>
<tr>
<td>MCH (pg/cell)</td>
<td>517</td>
<td>29.1</td>
<td>2.6</td>
<td>29.2</td>
<td>19.1</td>
<td>37</td>
</tr>
<tr>
<td>MCHC (g/dL)</td>
<td>517</td>
<td>34.0</td>
<td>1.51</td>
<td>34.1</td>
<td>28</td>
<td>38.9</td>
</tr>
<tr>
<td>PLT (10^9/L)</td>
<td>517</td>
<td>279</td>
<td>152</td>
<td>260</td>
<td>24</td>
<td>1997</td>
</tr>
<tr>
<td>CRP (mg/L)</td>
<td>453</td>
<td>7.56</td>
<td>9.18</td>
<td>3.2</td>
<td>0.01</td>
<td>46.8</td>
</tr>
</tbody>
</table>

2. Materials

Altogether, 517 individual MPV measurements, one for each patient (F = 55%, aged = 53.4 ± 27.3, ranged 1–96), were available for the 186 diagnosed with infection (36%) and 331 (64%) cancer (new and old) diagnoses at the Multi-Specialist Hospital Gorzów Wielkopolski from the 1st of January 2017 to the 30th of June 2018.

Measurements of the blood biomarkers were performed in the hospital laboratory unit using a Sysmex XN-2000 (Sysmex Corporation, Japan) analytical system with EDTA-KE/2.7 mL samples. Additionally, serum Z/4.9 mL samples were used to measure CRP. The descriptive statistics for MPV and other selected morphological biomarkers are given in Table 1.

The Bioethical Committee of Poznań University of Medical Sciences approved the study, and it was conducted in accordance with the Declaration of Helsinki (No. KB-1028/17).

At this point, we realized that the use of retrospective data may be the most concerning limitation of our study. Although the data used in the analysis were mostly complete (n = 517), some unorganized or incomplete medical records could bias inferential statistics, especially for neutrophils, lymphocytes, and monocytes.

3. Methods and Results

The computation was performed with the R statistical platform [9].

To avoid highly multidimensional data analysis due to examination of several biomarkers at the same time, our investigation of the clinical profiles of the diseases, we started from the mean platelet volume (MPV), a marker derived from common whole blood count. The choice of the MPV biomarker was justified by the fact that the immune system is linked to platelet function because platelets are involved in the so-called innate immunity [10]. Platelet count and MPV could be indicative of an infection. On the other side, cancer is a consumptive disease that frequently affects bone marrow function by suppressing it. In addition,
treating cancer with chemotherapy induces bone marrow depression in all cell lines, which finally results in altered platelet parameters.

In the statistical analysis, a cluster analysis was first performed using the simple idea described by [11]. Based on patient age and sex and their MPV, a classification tree (dendrogram) of patients was created [12] and is presented in Figure 1.

Description: in Figure 1, particular patients (coded by “P” numbers) from up to down are hierarchically aggregated in separated branches that are represented by individual leaves of the dendrogram. Four main families (clusters “1,” “2,” “3,” and “4”) of patients can be distinguished in the created classification tree based on patient age and sex and their MPV.

Following the “P” numbers, the identification of patients was performed, and descriptive statistics of the established clusters were conducted using one-way ANOVAs with adjusted Bonferroni-corrected p values (Table 2).

From the results reported in Table 2, we can see the sex separation between the clusters (1 and 2 = males and 3 and 4 = females) and the estimated p values indicate highly representative clusters of patients. The plots of means for the analysed risk factors (age, sex, MPV, and cancer) are presented in the corresponding Figure 2.

To complete the one-way ANOVAs, multiple comparisons between clusters (post hoc Tukey’s HSD and Dunn’s test p values) for age and MPV were performed (Table 3).

It can be seen in Table 3 that clusters 1 and 3 and 2 and 4 are nearly “identical” in terms of the analysed clinical factors (age, MPV, and cancer). For the remaining biomarkers, additional one-way ANOVAs were conducted (with Bonferroni adjustment) to check for differences among the clusters (Table 4).

From the statistical estimates reported in Table 4, we can see that the diseases (infection and cancer) and the seven selected biomarkers (RBC, haematocrit, haemoglobin, MCV, MCH, MCHC, and CRP) statistically significantly different among the established clusters of patients. For these results, plots of means are presented in the corresponding Figure 3.

Finally, and separately for males (clusters 3 and 4) and females (clusters 1 and 2), ROC analysis [13] was conducted to estimate the cutoffs for the statistically significant classifiers (biomarkers) predicting infection and cancer in the analysed group of patients (Table 5).

From the results reported in Table 5, we can see that age and all selected biomarkers were statistically significant classifiers of the clusters within male and female patients. The most predictive were age for males (39) and females (27), which segregated patients with 100% precision. Next, MCV had AUCs of 84% and 87%, respectively. MPV in males had the poorest AUC (63%), as did MCHC in females (62%). Based on these values, we can establish profiles of infection and cancer patients with reference to estimated threshold values (Table 6).
4. Discussion

In addition to classic solutions, modern statistical techniques are becoming more and more popular in personalized medicine and important for clinical trials. In the study of biomarkers in which regression modelling fails, cluster analysis...
Figure 3: Continued.
decision tree analysis, principal component analysis, network analysis, and receiver operating characteristic curves are progressively gaining greater research significance.

The current study shows how application of modern statistical ideas could add to the classical approaches by screening and combining biomarkers distinguishing between different diseases such as infection and cancer. In clinical practice, patients are not diagnosed with biomarkers only but those combined with symptoms, medical history, and imaging (e.g., X-ray and ultrasound). Principally, there should be serious suspicion or signs, such as fever or redness, before blood tests are ordered. However, sometimes the symptoms are not very specific. In that sense, a predefined search strategy could help identify risk factors for disease on an individual basis.

As we assumed, the most incomplete data bioindicators (i.e., neutrophils, lymphocytes, and monocytes) did not play a role in our biomarker analysis. However, we showed that basing on age and the remaining “everyday” biomarkers (i.e., MPV, RBC, haematocrit, haemoglobin, MCV, MCH, MCHC, and CRP) provided statistically reliable cutoffs of their levels for distinguishing early symptoms of the analysed diseases for stratification purposes. After a patient has been allocated to a “risk group,” more specific investigations could be initiated. Taking several successive bioindicators together will improve the precision of the classification.

Table 5: ROC analysis of patients (with cutoffs, AUC, and 95% CI).

<table>
<thead>
<tr>
<th>Sex</th>
<th>Clinical factor</th>
<th>Cutoff</th>
<th>Males</th>
<th>AUC</th>
<th>95% CI</th>
<th>Females</th>
<th>AUC</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>Cutoff</td>
<td>AUC</td>
<td>95% CI</td>
<td>Cutoff</td>
<td>AUC</td>
<td>95% CI</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>39</td>
<td>100%</td>
<td>100%-100%</td>
<td>27</td>
<td>100%</td>
<td>100%-100%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MPV (fL)</td>
<td>10.65</td>
<td>63%</td>
<td>56%-70%</td>
<td>9.55</td>
<td>68%</td>
<td>60%-76%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RBC (mln/mm³)</td>
<td>4.50</td>
<td>81%</td>
<td>75%-87%</td>
<td>4.31</td>
<td>80%</td>
<td>74%-86%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Haematocrit (%)</td>
<td>33.5</td>
<td>66%</td>
<td>59%-73%</td>
<td>33.7</td>
<td>65%</td>
<td>58%-72%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Haemoglobin (g/dL)</td>
<td>12.0</td>
<td>69%</td>
<td>63%-76%</td>
<td>11.5</td>
<td>67%</td>
<td>60%-74%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCV (fL)</td>
<td>84.1</td>
<td>84%</td>
<td>78%-89%</td>
<td>82.3</td>
<td>87%</td>
<td>82%-92%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCH (pg/cell)</td>
<td>27.2</td>
<td>74%</td>
<td>68%-81%</td>
<td>28.6</td>
<td>80%</td>
<td>74%-86%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCHC (g/dL)</td>
<td>34.0</td>
<td>70%</td>
<td>63%-77%</td>
<td>33.6</td>
<td>62%</td>
<td>55%-69%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CRP (mg/L)</td>
<td>2.63</td>
<td>84%</td>
<td>78%-89%</td>
<td>4.38</td>
<td>74%</td>
<td>66%-81%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 6: Clinical profiles of infection and cancer for selected clinical factors.

<table>
<thead>
<tr>
<th>Clinical factor/disease</th>
<th>Infection</th>
<th>Cancer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>Younger</td>
<td>Older</td>
</tr>
<tr>
<td>MPV (fL)</td>
<td>Lower</td>
<td>Higher</td>
</tr>
<tr>
<td>RBC (mln/mm³)</td>
<td>Higher</td>
<td>Lower</td>
</tr>
<tr>
<td>Haematocrit (%)</td>
<td>Higher</td>
<td>Lower</td>
</tr>
<tr>
<td>Haemoglobin (g/dL)</td>
<td>Higher</td>
<td>Lower</td>
</tr>
<tr>
<td>MCV (fL)</td>
<td>Lower</td>
<td>Higher</td>
</tr>
<tr>
<td>MCH (pg/cell)</td>
<td>Lower</td>
<td>Higher</td>
</tr>
<tr>
<td>MCHC (g/dL)</td>
<td>Higher</td>
<td>Lower</td>
</tr>
<tr>
<td>CRP (mg/L)</td>
<td>Lower</td>
<td>Higher</td>
</tr>
</tbody>
</table>

[14], decision tree analysis [15], principal component analysis [16], network analysis [17], and receiver operating characteristic curves [18] are progressively gaining greater research significance.

The current study shows how application of modern statistical ideas could add to the classical approaches by screening and combining biomarkers distinguishing between different diseases such as infection and cancer. In clinical practice, patients are not diagnosed with biomarkers only but those combined with symptoms, medical history, and imaging (e.g., X-ray and ultrasound). Principally, there should be serious suspicion or signs, such as fever or redness, before blood tests are ordered. However, sometimes the symptoms are not very specific. In that sense, a predefined search strategy could help identify risk factors for disease on an individual basis.

As we assumed, the most incomplete data bioindicators (i.e., neutrophils, lymphocytes, and monocytes) did not play a role in our biomarker analysis. However, we showed that basing on age and the remaining “everyday” biomarkers (i.e., MPV, RBC, haematocrit, haemoglobin, MCV, MCH, MCHC, and CRP) provided statistically reliable cutoffs of their levels for distinguishing early symptoms of the analysed diseases for stratification purposes. After a patient has been allocated to a “risk group,” more specific investigations could be initiated. Taking several successive bioindicators together will improve the precision of the classification. In other
words, the more indicators are fed into the analysis, the stronger the results get, which will reduce the probability of misclassification.

At this point, it seems necessary to comment on the classification of patients by their age. It is only the preselection of patients in our study that would cause erroneous allocation of “young cancer” cases and “old infection” patients. However, if the remaining seven bioindicators were more specific, this disadvantage could be overcome. We realize that the validation of patients does not have to be just perfect. Having several biomarkers at our disposal, however, we can determine a possible risk group with high probability.

Such an approach speeds up finding a diagnosis and avoids unnecessary investigations that may afflict patients. Both could save time, increase patient comfort and well-being, and save money by reducing ineffective studies. Additionally, such an epidemiological solution could be an easy and effective tool to place patients on appropriate clinical treatment pathways, which finally could impact economic results. Moreover, this tool could be important after exhausting conventional tests and statistics to predict and diagnose an underlying, unspecified disease for an individual patient. Again, combining a set of markers derived from a cluster search could help identify the correct disease and treatment. For example, the mean platelet volume is a general marker that has been associated with all kinds of diseases. Although there are still some methodological concerns, the strength of MPV as a diagnostic marker may increase in value using modern statistics. As each marker per se is not strong enough to distinguish clusters, it helps to combine these preselected markers and search then for correlations with specific diseases.

At the end of this discussion, we would like to emphasize one more important conclusion coming from our research, i.e., that not only very unique and precise diagnostic devices and tests need to be obligatory for a meaningful assessment of patient health. We can use standard and inexpensive biomarkers as sufficient analytical tools with strong and invaluable diagnostic power.

5. Conclusions

Based on the gathered material, the clinical and statistical methods, and the obtained results, the following conclusions can be drawn from our study:

(i) Despite infection and cancer being very different diseases in their clinical characteristics, it seems possible to discriminate them using “everyday” biomarkers and popular statistical methods

(ii) The estimated cutoffs of the specified biomarkers can be used to allocate patients to the appropriate “risk group” for stratification purposes

(iii) We believe that filtering by a few biomarkers could diminish the chance of classification of patients into the wrong categories of the disease

(iv) The presented methodology can be of use in a large-scale diagnostic evaluation of patients, such as for medical management or epidemiological administration

(v) Standard diagnostic tests may be sufficient to allocate patients to an increased risk group without the need for unique and expensive analytical methods

(vi) The diagnostic qualification of disease cases may depend on the assumed number of clinical criteria met in the algorithm

(vii) The established clinical norms for the biomarkers should undergo scientific verification and comparison in other populations

Abbreviations

ALT: Alanine transaminase
ANOVA: Analysis of variance
AST: Aspartate transaminase
CI: Confidence interval
CRP: C-reactive protein
HSD: Honest significant difference
LDL: Low-density lipoprotein
MCH: Mean corpuscular haemoglobin
MCHC: Mean corpuscular haemoglobin concentration
MCV: Mean corpuscular volume
MPV: Mean platelet volume
PLT: Platelets
RBC: Red blood cell
ROC: Receiver operating characteristic
TG-C: Triglyceride-cholesterol
WBC: White blood cell.

Data Availability

Personal information on the study patients has been anonymized, and the dataset used in the statistical analysis is completely available for further clinical trials upon request.

Conflicts of Interest

The authors have no conflicts of interest to declare.

Authors’ Contributions

All authors have had access to the data and all drafts of the manuscript. KB-J, MDL, and AT designed the study. KB-J, MDL, and AT managed and analysed the data. KB-J, MDL, AT, MF, JR-M, and EW-R wrote the draft of the manuscript. MF, JR-M, EW-R, and LG reviewed the manuscript. All authors read and approved the final manuscript.

Acknowledgments

Our special thanks to Zuzanna Walkowiak, MSc, Multi-Specialist Hospital Gorzów Wielkopolski, Poland, for data collection.
References


