Use of exploratory factor analysis to ascertain the correlation between the activities of rheumatoid arthritis and infection by human parvovirus B19

Natalja Kakurina a,*, Anda Kadisa b, Aivars Lejnieks c, Helena Mikazane b, Svetlana Kozireva c, Modra Murovska c

a Daugavpils Regional Hospital, Daugavpils, Latvia
b Department of Internal Diseases, Riga Stradins University, Riga, Latvia
c August Kirchenstein Institute of Microbiology and Virology, Riga Stradins University, Riga, Latvia

ABSTRACT

Background and objective: We evaluated a possible correlation between the clinical activities of rheumatoid arthritis (RA) and human parvovirus B19 (B19) infection using exploratory factor analysis (EFA).

Materials and methods: RA patients were organized into two groups: 100 patients in the main group and 97 in the RA(DAS28) group. Four subgroups were defined from the main group according to the presence or absence of certain infection-specific markers: group I comprised 43 patients who had IgG antibodies against B19; group II, 25 patients with active B19 infection (B19-specific IgM antibodies and/or plasma viremia); group III, 19 patients with latent/persistent B19 infection (virus-specific sequences in peripheral blood leukocytes’ DNA with or without B19-specific IgG antibodies), and group IV, 13 patients without infection markers. The RA(DAS28) group was divided into four subgroups similarly to the main group: group I, 35; group II, 31; group III, 19; and group IV, 12 patients. Disease-specific clinical values in both groups were analyzed employing EFA, and the RA(DAS28) group was additionally assessed using Disease Activity Score (DAS28).

Results: RA activity was higher in patients who had markers of B19 infection. The highest activity of RA in both study groups was in patients with latent/persistent infection. In the RA(DAS28) group, according to DAS28, the highest activity of RA was in patients with active B19 infection.

Conclusions: Using EFA and DAS28, a correlation between the clinical activity of RA and B19 infection was confirmed. These data suggest that EFA is applicable for medico-biological studies.

© 2015 Lithuanian University of Health Sciences. Production and hosting by Elsevier Urban & Partner Sp. z o.o. All rights reserved.
1. Introduction

Rheumatoid arthritis (RA) is a chronic autoimmune disorder characterized by symmetrical, destructive polyarthritis and simultaneously by systemic inflammation that has a long-term impact on various organs, vessels and the hematopoietic system [1]. The exact cause of the rheumatoid process is not known. However, certain infections are considered to trigger autoimmunity and therefore are being studied in connection with RA.

Human parvovirus B19 (B19) has been observed to cause the initial immune response (i.e. the initial reaction preceding RA development in potentially receptive hosts) [2,3]. B19 infection is quite common in children and adults. It is usually asymptomatic but, if symptoms occur, they resemble those of the common cold. The infection presents with various manifestations that are well-known and covered extensively in the literature [4,5]. Evidence implicating B19 in RA causation is conflicting [6,7], but several reports have clearly confirmed its role in the pathogenesis of RA [8,9]. Thus, one can assume that infection by B19 has (at least in part) a role in disease activity that is directly (or sometimes implicitly) indicated by various clinical and laboratory data (both segregate and aggregate).

The present study was carried out to evaluate a possible correlation between the clinical activity of RA and human B19 infection using a statistical method known as exploratory factor analysis (EFA). EFA is efficient at reducing the dimensionality of the initial variables, eliminating outliers and retaining “significant factors.” In this case, the significant factors were the activity factors of RA. EFA was based partially on the data acquired within our previous study on the prevalence and clinical significance of parvovirus infection in RA patients [10].

Most clinical trials rely exclusively on the Disease Activity Score (DAS) [11,12] but we also explored another approach. One of the significant factors from the output of the EFA, a “common factor,” was assumed to be the aggravating factor of disease activity due to the statistical significance of the disease-specific clinical parameters it incorporated. The common factor was derived during the EFA so, unlike the DAS, it was native. However, to observe the objectivity of our study, the results of both approaches were evaluated and compared.

2. Materials and methods

The study protocol was approved by the Ethics Committee of Riga East Clinical University Hospital (Riga, Latvia). Each patient gave informed, voluntary, written consent to participate in this study.

2.1. Patients

The study featured two groups of RA patients: the main group and RA(DAS28) group. The main group was a cohort of 100 patients (72 women and 28 men; range, 21–81 years; mean, 55 years). The RA(DAS28) group comprised 97 RA patients (85 women and 12 men; range, 19–82 years; mean, 56 years). All patients fulfilled the current classification criteria for RA set by the American College of Rheumatology (ACR) [13]. Patients were hospitalized in the Rheumatology Department of Riga East Clinical University Hospital.

Initially, to exclude therapy as a co-factor of the activity of B19 infection, patients in the main group were divided into two subgroups [14]. The first group represented patients who had received various disease-modifying anti-rheumatic drugs and/or corticosteroids (68 RA patients). The second group comprised patients who had received only analgesics and/or non-steroidal anti-inflammatory drugs (32 RA patients). Elimination of another possible co-factor, disease duration, was essential, so 28 patients with early RA (i.e. with disease duration <2 years) were segregated from the main cohort.

Later, four subgroups were defined from the main group according to the presence or absence of certain infection-specific markers. The first subgroup (group I) comprised 43 patients who had suffered B19 infection and had IgG antibodies against B19. Group II comprised 25 patients with active B19 infection. They had IgM antibodies against B19 and/or B19 genomic sequences in plasma DNA samples. Group III comprised 19 patients with latent/persistent B19 infection. They had B19 genomic sequences in DNA isolated from peripheral blood leukocytes with or without IgG antibodies against B19. Group IV comprised 13 patients without infection markers and who were later termed “virus-negative patients”. The RA(DAS28) group was (similar to the main group) divided into several subgroups according to the presence or absence of infection-specific markers. Thus, group I had 35 patients, group II had 31 patients, group III had 19 patients, and group IV had 12 patients.

2.2. Clinical and laboratory characteristics

All laboratory examinations were carried out using state-certified and standardized laboratory methods with appropriate equipment and reagents [14]. Each manifestation of the disease was evaluated by adhering to the corresponding identification standards [15]. Most RA-specific clinical parameters were selected based on our previous work [10] (though only some retained their utility for EFA). We evaluated the erythrocyte sedimentation rate (ESR; mm/h; Westergren method), tender and swollen joint counts (ACR 66/68-joint count), duration of morning stiffness (in h), level of C-reactive protein (CRP; mg/L); hemoglobin level (g/L), platelet count \((10^9/L)\), and lymphocyte count \((10^9/L)\).

Additionally, DAS28 was calculated using the standard formula for patients in the RA(DAS28) group. DAS28 \(\leq 2.6\) signified remission of RA; DAS28 > 2.6 \(\leq 3.2\) indicated low disease activity, DAS28 > 3.2 \(\leq 5.1\) indicated moderate disease activity, and DAS28 > 5.1 indicated high disease activity.

2.3. Methods of viral diagnostics

IgG and IgM antibodies against B19 in plasma samples were identified using the VP2 enzymatic immunoassays developed by Biotrin (Dublin, Ireland) according to manufacturer protocols. Viral DNA was confirmed to be present by the nested
polymerase chain reaction assay method. The procedures mentioned above were conducted as described previously [10].

2.4. Statistical analysis

Assorted generally accepted descriptive and conclusive statistical methods were applied during the study to suit specific tasks and/or data types using SPSS v13.0 (SPSS, Chicago, IL, USA). For instance, statistical hypotheses were tested using the Fishers exact test, chi-square test, and ANOVA. The validity of EFA was justified using the Kaiser–Meyer–Olkin (KMO) index of sampling adequacy which, in this case, was 0.661 for the main group and 0.541 for the RA(DAS28) group (the KMO index should be >0.5 for satisfactory factor analyses to proceed) [16–18].

Finally, variance tests were carried out at a probability level of 95% or $P = 0.050$. Therefore, $P \leq 0.050$ was the rationale to reject the null hypotheses.

3. Results

Drug therapy was excluded as a possible co-factor of infection activity because initial patient segregation from the main group and its analysis by the chi-square test did not reveal significance. Disease duration was also considered to be not relevant because there were no significant differences in the numbers of patients with early RA and patients with disease duration $>2$ years. Therefore, the only suitable method for patient segregation was patient grouping by infection markers.

As part of EFA, to reduce the large number of variables into a smaller number that would explain variation and correlations among observations, two significant factors emerged: F1 and F2. F1 aggregated the ESR, CRP level, duration of morning stiffness, tender and swollen joint counts, and hemoglobin level. F2 incorporated the platelet count and lymphocyte count (Fig. 1).

The output of EFA was subsequently rotated to allow positioning of specific factors so that each factor comprised a few (but highly loaded) variables. Two clinical parameters, the ESR and CRP level, displayed the highest contribution among the initial values of F1 (i.e., had the highest factor loading) whereas the lymphocyte count had the highest factor loading in F2. F1 caused 33% dispersion of the initial parameters, whereas F2 caused only 16%. Moreover, F2 proved to be excessively complex to interpret from a clinical viewpoint. Therefore, for statistical and practical reasons, F1 was selected as the common factor for further analyses, and F2 (along with the parameters it incorporated) was rendered redundant.

F1 was standardized ($z_{F1}$) and converted to an ordinal scale with three distinctive points using the formula

$$z_{F1} \pm 0.5 \times \sigma_{z_{F1}}$$

Subsequently, using the SPSS software algorithms combined with the mentioned formula, and with the mean value of $z_{F1}$ at 0 according to the formula, three distinct groups were formed: (a) the group with a low level of the rheumatoid process activity – the calculated value less than $-0.5$; (b) the group with an average level of the rheumatoid process activity – the calculated value in between $-0.5$ and $+0.5$ (greater than $-0.5$ and lower than $+0.5$); (c) the group with a high level of the rheumatoid process activity – the calculated value greater than $+0.5$. Finally, the average values of the above-mentioned factors were determined.
According to all the parameters, each level was unique and probably different \( (P = 0.004 \text{ for joints and } P < 0.0001 \text{ for all other positions}) \) (Table 1).

Judging by the results of the ANOVA test and considering all other relevant indicators, patients with a high level of the common factor differed significantly from patients with average and low levels. Significant differences were also discovered among patients within all levels judging by values of the ESR and hemoglobin level. Fisher exact test, chi-square test and ANOVA confirmed that each RA patient with a high level of the common factor was also positive for infection markers, and this was significant \( (P = 0.024) \) (Fig. 2).

Several other significant differences were observed while further implementing the concept of the common factor into the study. For instance, analyzing patients previously segregated by the presence of infection markers, significant differences were identified between group III and group IV \( (P = 0.01) \) as well as group I and group IV \( (P = 0.032) \) (Fig. 3).

Furthermore, there were significantly more patients with a high level of the common factor in group III than among virus-negative patients (group IV). There were greater numbers of patients with a high level of the common factor in group I than among virus-negative patients (group IV) and the difference was significant. Ultimately, the difference in the numbers of patients with a high level of the common factor between group II and group IV was not taken into consideration because it was not significant \( (P = 0.055) \), but it could be regarded as a statistical tendency.

The RA(DAS28) group, similar to the main group, was also subjected to EFA. However, the distribution of the initial clinical parameters among the resulting significant factors was slightly different for this group. The first of the significant factors, F1c, aggregated the ESR, CRP level and duration of morning stiffness. However, the second significant factor, F2c, incorporated the platelet count, lymphocyte count and hemoglobin level (Fig. 4). As in the case of the main group, the final output of EFA was rotated so that each factor comprised a few high-loaded variables. This time, only the CRP level displayed the highest contribution among the initial values of the first factor (F1c): 0.783. The highest loaded variable of the second significant factor (F2c) was also different: it was the hemoglobin level and it equaled 0.885. F2c proved to be excessively complex so F1c was selected as the common factor for further analyses and F2c (not unlike F2 in the main group), along with the parameters it incorporated, was rendered redundant. Three thresholds of the common factor were identified as patients stratified by the degree of disease activity (Table 2). Again, according to all parameters, each level was unique and probably different \( (P < 0.001 \text{ for all other positions}) \).

A significant difference \( (P = 0.016) \) in the F1c levels was observed between group I and group III. RA patients with latent/persistent B19 infection had a high F1c level in 46% of cases, whereas RA patients who suffered B19 infection had a high F1c level in only 9% of cases (Fig. 5).

The results of EFA for the main group enabled derivation of a formula to calculate the value of the aggregated RA activity factor F1 for each patient from the RA(DAS28) group individually. A direct and significant correlation was observed between factor F1 and DAS28 (Fig. 6) (Spearman rank correlation coefficient, 0.720; \( P < 0.01 \)): higher DAS28 values corresponded to higher F1 values.

Unlike the main group, the RA(DAS28) group was also assessed using a widely employed measure: DAS28. According to the results of the assessment, RA was in remission in 7 patients from the RA(DAS28) group. For another 7 patients, the disease had low activity. Mean and high levels of disease

![Fig. 2 – Distribution of virus-negative patients and patients with infection markers among the three levels of the common factor F1 for the main group.](image)

![Fig. 3 – Distribution of F1 levels in each subgroup for the main group.](image)
activity were observed in 45 and in 38 patients, respectively. The distribution yielded no significant differences according to the chi-square test.

The mean DAS28 levels were established for each of the 4 subgroups: group I, 4.59 (which corresponded to an average activity level of RA); group II, 5.24 (high level); group III, 4.36 (average level); and group IV, 4.79 (average level) (Fig. 7). There were two significant differences according to the t-test: between group I and group II and between group II (P = 0.046) and group III (P = 0.024). The mean levels of DAS28 differed significantly between patients with active infection and patients with latent/persistent infection, and patients who had infection.

4. Discussion

The exact etiopathogenesis of RA is not clear despite the many studies devoted to it. Certain viral infections (including B19) are considered initializing factors for RA [8,19–23]. Molecular mimicry between host and viral proteins seems to be the main mechanism involved in the induction of autoimmunity. Lunardi et al. identified a peptide that shares homology with the B19 VP1 protein and with human cytokeratin. Moreover, this peptide shares similarity with the transcription factor GATA1 that plays essential parts in megakaryopoiesis and erythropoiesis [19]. In several studies, no correlation between B19 infection and disease activity was observed [24,25]. Nevertheless, those studies explored the effect of B19 infection on other diseases, such as juvenile dermatomyositis and systemic sclerosis, but not RA. Lehman et al. showed that despite intensive therapy >50% of children with juvenile polyarticular arthritis associated with persistent B19 infection had chronic inflammation years after the onset of disease, and that the presence of B19 genomes triggers disease severity [26,27]. Therefore it is not possible to comprehensively compare the results of those studies with the data acquired in our study.

Initially, DAS28 was not employed in our study because our study began in 2001 and at the time DAS28 was not used very often. Later, the introduction of DAS28 in the study group was considered inefficient because it would result in division of the

---

**Fig. 4 – Schematic representation of the significant factors F1c and F2c, initial parameters and their contributions (as numerical values) for the RA(DAS28) group.**

---

**Table 2 – Three levels of the common factor F1c for the RA (DAS28) group.**

<table>
<thead>
<tr>
<th></th>
<th>Low mean ± SE</th>
<th>Average mean ± SE</th>
<th>High mean ± SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Duration of MS, h</td>
<td>1.1 ± 0.2</td>
<td>2.5 ± 0.2</td>
<td>5.1 ± 0.8</td>
</tr>
<tr>
<td>CRP, mg/L</td>
<td>4.4 ± 0.9</td>
<td>13.7 ± 2.4</td>
<td>50.7 ± 9.5</td>
</tr>
<tr>
<td>ESR*, mm/h</td>
<td>11 ± 1.6</td>
<td>30.5 ± 2.5</td>
<td>49.7 ± 6.1</td>
</tr>
</tbody>
</table>

MS, morning stiffness; CRP, C-reactive protein; ESR, erythrocyte sedimentation rate.

*Westergren method.

**Fig. 5 – Distribution of F1c levels in each subgroup of RA patients for the RA(DAS28) group.**
research material. Hence, we enrolled a new group [RA(DAS28) group] of 97 randomly selected RA patients in whom DAS28 was used to detect disease activity. DAS28 is calculated using a standard formula. However, the mathematical parameter F1, which reflects the main parameters of RA activity, is derived using factor analysis. A significant correlation was confirmed between the values of EFA and DAS28. Inclusion of the RA (DAS28) group with DAS28 revealed a connection between RA activity and B19 infection.

The larger range of our study allowed us to analyze a possible correlation between the clinical activity of RA and B19 infection using two distinct tools: EFA and DAS28. However, a certain inconsistency was revealed while processing and comparing the data acquired from these methods: EFA results for both groups suggested that the highest activity of the disease was in patients with latent/persistent B19 infection, but the data obtained using DAS28 suggested it was in patients with active B19 infection. The source of the inconsistency can be traced to one of the clinically subjective sub-parameters of DAS28, the Visual Analog Scale, in which the self-reported wellbeing of patients with active infection was worse. Some of the values of the common factor might also be considered to be subjective, but there is no doubt about their contribution to the clinical course of RA.

Ultimately, our statistical and comparative analyses enabled outlining the general direction and requirements for further research. For example, recruitment of a larger study cohort may lead to the discovery of how the activity phase of B19 infection influences the clinical course of RA. It may also clarify the importance of latent/persistent B19 infection for predicting development of the disease.

5. Conclusions

EFA proved to be applicable for medico-biological studies. Using EFA and DAS28, a correlation between the clinical activity of RA and B19 infection was confirmed. RA activity was higher in patients who had markers of B19 infection. The highest activity of RA in both study groups was in patients with latent/persistent infection. In the RA(DAS28) group, according to DAS28, the highest activity of RA was in patients with active B19 infection.

Conflicts of interest

The authors declare that they have no competing interests.
Acknowledgments

This study was supported in part by the grant 09.1112 of the Latvian Council of Sciences and National Program in Biomedicine (2009–2013).

REFERENCES