

RESEARCH

Open Access



# Joint index vector: a novel assessment measure for stratified medicine in patients with rheumatoid arthritis

Susumu Nishiyama<sup>1\*</sup> , Tetsuji Sawada<sup>2</sup>, Jinju Nishino<sup>3</sup> and Shigeto Tohma<sup>4</sup>

\*Correspondence:

newcity@mtd.biglobe.ne.jp

<sup>1</sup> Rheumatic Disease Center,  
Kurashiki Medical Center, 250  
Bakuro-cho, Kurashiki-shi,  
Okayama 710-8522, Japan  
Full list of author information  
is available at the end of the  
article

## Abstract

**Objective:** To predict the next-year status in patients with rheumatoid arthritis using big data.

**Methods:** Joint index (JI) of upper/large (UL), upper/small (US), lower/large (LL), and lower/small (LS) was calculated as the sum of tender and swollen joint counts divided by the number of evaluable joints in each region of interest. Joint index vector  $V(x, y, z)$  was defined as  $x = JI_{UL} + JI_{US}$ ,  $y = JI_{LL} + JI_{LS}$ , and  $z = JI_{UL} + JI_{LL} - JI_{US} - JI_{LS}$ . Low disease activity was defined as  $|V_{xy}| (= \sqrt{x^2 + y^2}) \leq 0.1$ . Patients with  $|V_{xy}| > 0.1$  were further classified into three groups: evenly affected (EVN):  $|z| \leq 0.2$ , small joint dominant (SML):  $z < -0.2$ , and large joint dominant (LAR):  $z > 0.2$ . To predict the next-year  $V(x, y, z)$  of each patient, a transformation matrix was computed from the mean vectors of the EVN, SML, and LAR groups and their translation vectors.

**Results:**  $|V_{xy}|$  was correlated with Simplified Disease Activity Index (SDAI) ( $r = 0.82$ ).  $Z$  of mean vector increased as the disability index of the Health Assessment Questionnaire (HAQ-DI) and the Steinbrocker class worsened. The LAR group had the worst HAQ-DI and the second highest SDAI after those in the SML group. Positive predictive value and likelihood ratio in predicting the LAR group were 58.7% and 5.9, respectively. Likelihood ratio was greater with treatment, at 7.2, 7.4, and 8.6 when targeted patients were treated with methotrexate, biologics, and both drugs, respectively.

**Conclusions:** Patients with high disease activity and poor functional state were predicted with high probability using joint index vectors.

**Keywords:** Data management, Decision analysis, Rheumatoid arthritis, Stratified medicine

## Introduction

Although it is difficult to forecast which joints will be involved or intact in patients with rheumatoid arthritis (RA) in the long term, the ability to predict the next-year distribution of affected joints for individual patients would be useful for choosing the correct therapeutic option.

A number of factors affect the development of RA, and several markers including genetic information have been investigated as potential means of distinguishing patients with preferable outcomes from others; however, confirmed evidence regarding personalized therapy is still limited [1, 2]. Recalling the basic fact that joint inflammation is an essential aspect of RA, it is quite natural and reasonable to look closely into which joints are thus affected. Since it is hard to predict single joint involvement separately, we tried to forecast the proportions of affected joints in four joint categories: upper/small, upper/large, lower/small, and lower/large. After treatment with biological disease-modifying anti-rheumatic drugs (bDMARDs), there was a significant relation in the change in the proportions of affected joints between the upper/small and upper/large categories as well as between the lower/small and the lower/large categories, while the response to bDMARDs in the upper limbs was independent of that in the lower limbs [3]; therefore, upper joints and lower joints should be evaluated separately. We have previously reported that upper/small joints affected activity-related HAQ, whereas large-joint involvement was associated with an increase in both activity-related and damage-related HAQ [4]. These findings indicate that we should discriminate large joints from small joints as well as upper joints from lower joints.

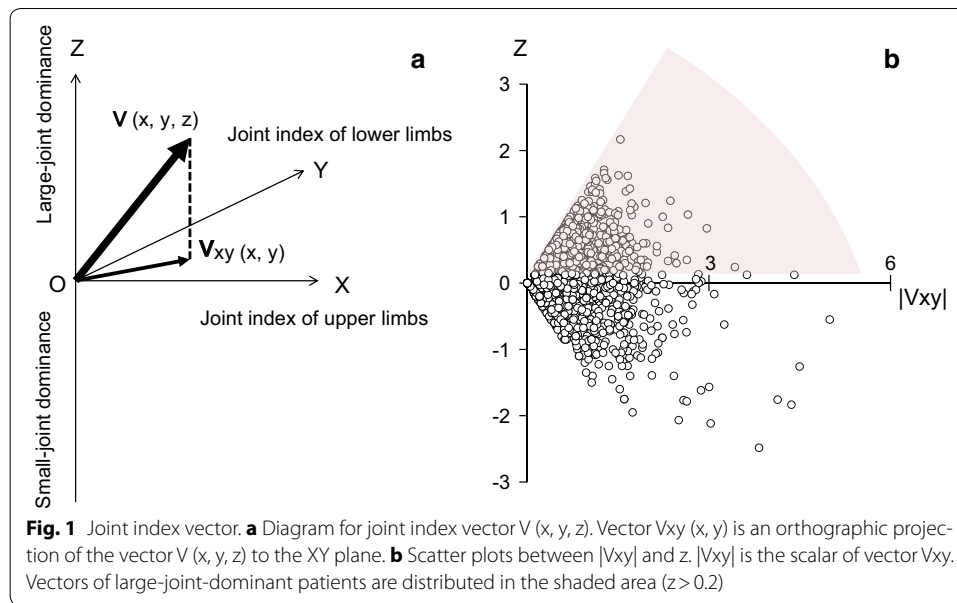
The joint index vector incorporates both the upper- and lower-joint index and the large- or small-joint dominance status in each patient. Based on this calculation, patients with high disease activity and poor functional state were predicted with high probability.

## Methods

The National Database of Rheumatic Diseases in Japan (*NinJa*) is a nationwide database that was assembled to illuminate the current status and issues of patients with RA in Japan and has been continuously updated since 2002. RA patients are registered annually from institutes in Japan, and data on 15,341 individuals from 51 institutes were available as of 2016. The *NinJa* project was reviewed and approved by the ethics committee at each participating institution and all patients participating in the study provided informed consent.

Excluding patients with missing values for Simplified Disease Activity Index (SDAI) [5] or the disability index of the Health Assessment Questionnaire (HAQ-DI) [6, 7], 11,013 patients were analyzed. Serial registration data from patients without orthopedic surgery between 2013 and 2014 ( $n=10,206$ ) and between 2015 and 2016 ( $n=10,118$ ) were used for validation of forecasting next-year joint involvement (Additional file 1).

The joint index (JI) was calculated as described previously [3]. Briefly, joints were divided into four categories based on functional aspect: upper/large (UL: shoulder, sternoclavicular, elbow, and wrist joints), upper/small (US: proximal interphalangeal and metacarpophalangeal joints), lower/large (LL: hip, knee, ankle, and tarsometatarsal joints), and lower/small (LS: metatarsophalangeal joints). The JI for each of these categories is the sum of the number of tender and swollen joints divided by the number of evaluable joints in that category. JI is within the range of 0–2.



Next, the joint index vector  $V(x, y, z)$  was calculated as  $x = JI_{UL} + JI_{US}$ ,  $y = JI_{LL} + JI_{LS}$ , and  $z = JI_{UL} + JI_{LL} - JI_{US} - JI_{LS}$ , where  $JI_{UL}$ ,  $JI_{US}$ ,  $JI_{LL}$ , and  $JI_{LS}$  indicate the joint indices of the upper/large, upper/small, lower/large, and lower/small joint categories, respectively. Thus,  $X$  and  $Y$  axes represent the  $J$ I of the upper limbs and the lower limbs, respectively, while the  $Z$  axis indicates large- or small-joint dominance of articular involvement (Fig. 1a).  $X$  and  $Y$  are within the range of 0–4 and  $z$  is within the range of  $-4$  to  $4$ .  $V_{xy}(x, y)$  is the orthographic projection of the vector  $V(x, y, z)$  to the  $XY$  plane.  $|V_{xy}|$  is the scalar of  $V_{xy}(x, y)$  calculated as the square root of  $(x^2 + y^2)$  (Fig. 1a).

$|V_{xy}| \leq 0.1$  was defined as low disease activity (LDA). Patients with  $|V_{xy}| > 0.1$  were further divided according to  $z$  value, such that patients with high disease activity and  $|z| \leq 0.2$  were considered evenly affected (EVN), those with  $z < -0.2$  were considered small joint dominant (SML) and those with  $z > 0.2$  were considered large joint dominant (LAR). Figure 1b shows the scatter plot of  $|V_{xy}|$  and  $z$ . Large-joint-dominant patients ( $z > 0.2$ ) are plotted in the shaded area of the upper part of the diagram. The following algorithm is the procedure to calculate the joint index vector. You can try this program as EXCEL macro using sample data (Additional file 2).

**Algorithm 1:** Procedure to calculate Joint index vector from individual affected joints

---

```

Sub JointIndexVector()
'Dim JointIndex_UL, JointIndex_US, JointIndex_LL, JointIndex_LS, X, Y, Z, Vxy As Single
Dim n As Integer
' Worksheet "Database" consists of ID and joint evaluation data
'Status of each joint is as follows:
' 0: neither tender nor swollen joint
' 1: swollen joint
' 10: tender joint
' 11: Tender and swollen joint
' -11: Not evaluable joint (e.g. joint replacement)
'

Worksheets("JointIndexVector").Activate
Range("A1").Select
ActiveCell.Offset(0, 0) = "ID"
ActiveCell.Offset(0, 1) = "UpperLarge" 'shoulder, sternoclavicular, elbow, hand
ActiveCell.Offset(0, 2) = "UpperSmall" 'MCP, PIP
ActiveCell.Offset(0, 3) = "LowerLarge" 'Hip, knee, ankle, Lisfranc
ActiveCell.Offset(0, 4) = "LowerSmall" 'MTP
ActiveCell.Offset(0, 6) = "X"
ActiveCell.Offset(0, 7) = "Y"
ActiveCell.Offset(0, 8) = "Z"
ActiveCell.Offset(0, 9) = "Vxy"
n = 1 'patient ID
Worksheets("Database").Activate
Do
    Range("B1").Offset(n, 0).Select
    JointIndex_UL = TenderJointIndex(8) + SwollenJointIndex(8)
    Range("J1").Offset(n, 0).Select
    JointIndex_US = TenderJointIndex(20) + SwollenJointIndex(20)
    Range("AD1").Offset(n, 0).Select
    JointIndex_LL = TenderJointIndex(8)
    Range("AF1").Offset(n, 0).Select
    JointIndex_LL = JointIndex_LL + SwollenJointIndex(6)
    Range("AL1").Offset(n, 0).Select
    JointIndex_LS = TenderJointIndex(10) + SwollenJointIndex(10)
    X = JointIndex_UL + JointIndex_US
    Y = JointIndex_LL + JointIndex_LS
    Z = JointIndex_UL + JointIndex_LL - JointIndex_US - JointIndex_LS
    Vxy = Sqr(X^2 + Y^2)
    Worksheets("JointIndexVector").Activate
    Range("A1").Select
    ActiveCell.Offset(n, 0) = n
    ActiveCell.Offset(n, 1) = JointIndex_UL
    ActiveCell.Offset(n, 2) = JointIndex_US
    ActiveCell.Offset(n, 3) = JointIndex_LL
    ActiveCell.Offset(n, 4) = JointIndex_LS
    ActiveCell.Offset(n, 6) = X
    ActiveCell.Offset(n, 7) = Y
    ActiveCell.Offset(n, 8) = Z
    ActiveCell.Offset(n, 9) = Vxy
    n = n + 1
    Worksheets("Database").Activate
Loop Until Range("A1").Offset(n, 0) = ""
End Sub

```

---

**Algorithm 2:** Function to calculate Tender joint index and Swollen joint index

---

```

Function TenderJointIndex(count)
    Dim i As Integer
    Dim TenderJointCount, EvaluableJointCount As Integer
    ,
    TenderJointCount = 0
    EvaluableJointCount = 0
    For i = 0 To count - 1
        If (ActiveCell.Offset(0, i) >= 0) Then
            TenderJointCount = TenderJointCount + (ActiveCell.Offset(0, i) ¥ 10)
            EvaluableJointCount = EvaluableJointCount + 1
        End If
    Next i
    TenderJointIndex = TenderJointCount / EvaluableJointCount
End Function

Function SwollenJointIndex(count)
    Dim i, temp As Integer
    Dim SwollenJointCount, EvaluableJointCount As Integer
    ,
    SwollenJointCount = 0
    EvaluableJointCount = 0
    For i = 0 To count - 1
        If (ActiveCell.Offset(0, i) >= 0) Then
            SwollenJointCount = SwollenJointCount + ActiveCell.Offset(0, i) Mod 10
            EvaluableJointCount = EvaluableJointCount + 1
        End If
    Next i
    SwollenJointIndex = SwollenJointCount / EvaluableJointCount
End Function

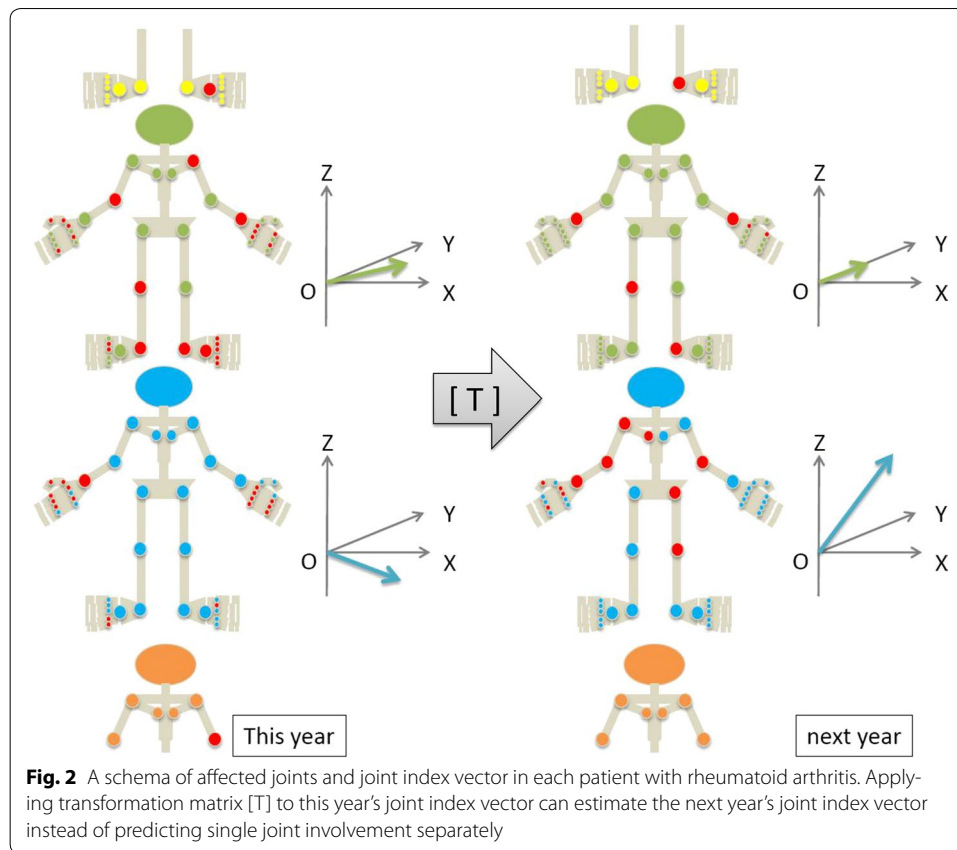
```

---

Although it is hard to predict single joint involvement separately, applying transformation matrix [T] to this year's joint index vector can estimate the next year's joint index vector (Fig. 2). Transformation matrix [T] was computed using serial registration data from the *NinJa* database from 2013 to 2014. If [A] is a  $3 \times 3$  matrix, with three mean vectors of the EVN, SML, and LAR groups in 2013, and [B] is a  $3 \times 3$  matrix, with three mean vectors in 2014 originating from patients in the EVN, SML, and LAR groups in 2013, then the matrix equation is as follows:

$$[A][T] = [B].$$

Therefore,



$$[T] = [A]^{-1}[B].$$

The estimated vector of each patient in 2016 ( $eV_{2016}$ ) was calculated by applying the transformation matrix to each person's vector in 2015 ( $V_{2015}$ ) as  $eV_{2016} = V_{2015} [T]$ .

### Statistical analysis

Pearson product-moment correlation coefficient was used to examine the correlation between  $|V_{xy}|$  and SDAI. The Student's  $t$ -test was used to compare the two groups.

### Results

Figure 3 shows the mean values of  $|V_{xy}|$  and  $z$  stratified by physical function, disease duration, and disease activity.  $z$  value elevated as Steinbrocker class progressed (Fig. 3a) and as HAQ-DI increased (Fig. 3b). Although a tendency of  $|V_{xy}|$  to increase over time was also observed, this increase was nonlinear compared to the increase in  $z$  value as physical function deteriorated (Fig. 3a, b).  $|V_{xy}|$  and  $z$  stayed close to their original values regardless of disease duration (Fig. 3c).  $|V_{xy}|$  increased as the disease activity rose, while  $z$  value did not (Fig. 3d). The correlation coefficient between  $|V_{xy}|$  and SDAI was 0.82 (Fig. 4). The mean (standard error) of  $|V_{xy}|$  in patients with low disease activity ( $SDAI \leq 11$ ) was 0.116 (0.002).

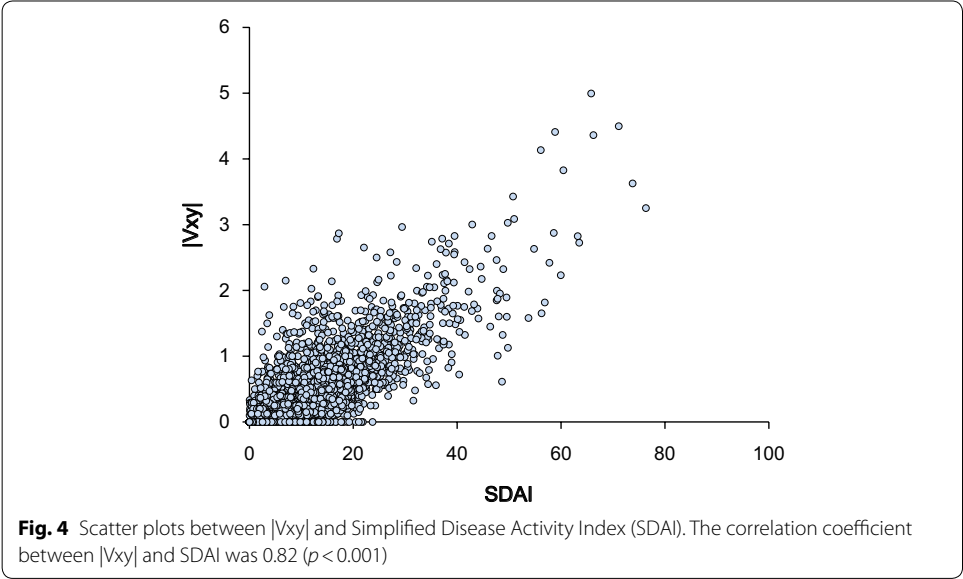
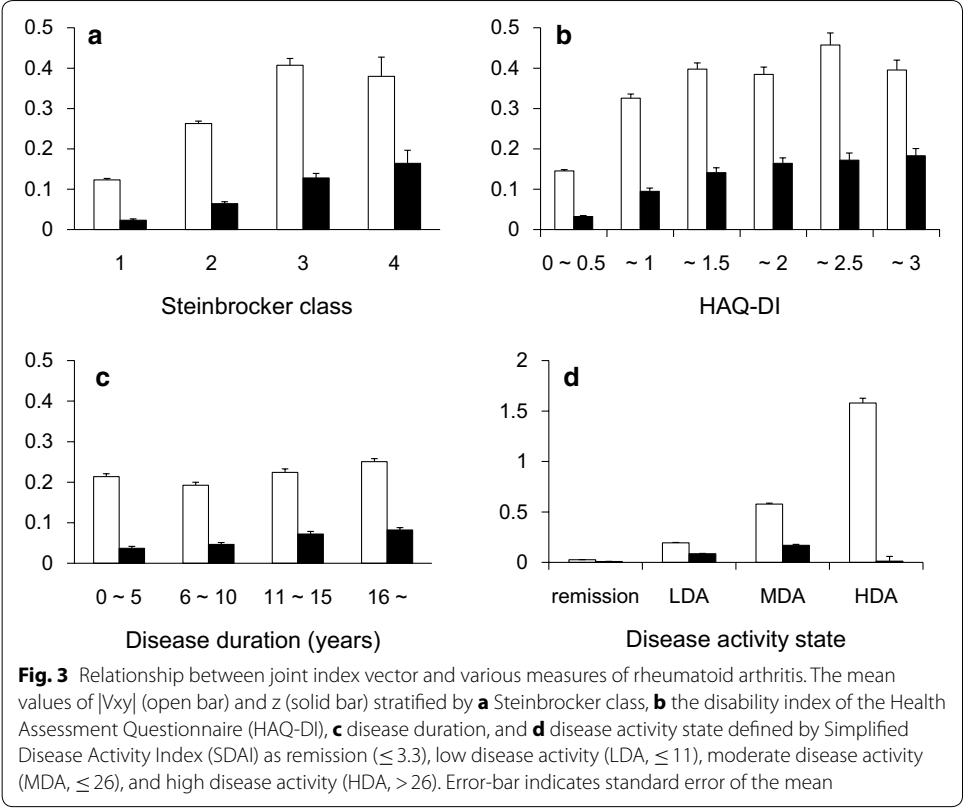


Table 1 shows the clinical characteristics of classified patients according to  $|V_{xy}|$  and  $z$  value. The LDA group had the lowest SDAI and HAQ-DI and also had the smallest proportion of females. The EVN group had higher SDAI and HAQ-DI than the LDA group, though its SDAI was lower than those of the SML and LAR groups and its HAQ-DI was lower than that of the LAR group. Patients in the SML group were the youngest and had

**Table 1** Demographic data of classified patients according to the joint index vector

	N	Disease duration	Age	Female (%)	SDAI	HAQ-DI
LDA	5135	11.9 (10.4)	64.5 (13.3)	76.7	2.8 (2.7)	0.4 (0.7)
EVN	2718	13.7 (11.1)	65.0 (12.5)	82.9	8.7 (6.8)	0.7 (0.8)
SML	797	12.1 (10.4)	62.1 (12.9)	82.2	14.0 (10.4)	0.7 (0.8)
LAR	2363	15.2 (11.7)	66.7 (12.5)	83.9	11.4 (7.0)	1.0 (0.9)
Total	11,013	13.1 (10.9)	64.9 (13.0)	80.2	6.9 (7.1)	0.6 (0.8)

Data are shown as mean (SD) unless otherwise described

SDAI Simplified Disease Activity Index, HAQ-DI the disability index of the Health Assessment Questionnaire, LDA low disease activity ( $|V_{xy}| \leq 0.1$ ), EVN evenly affected ( $|V_{xy}| > 0.1$  and  $|z| \leq 0.2$ ), SML small joints dominant ( $|V_{xy}| > 0.1$  and  $z < -0.2$ ), LAR large joints dominant ( $|V_{xy}| > 0.1$  and  $z > 0.2$ )

the highest SDAI. Patients in the LAR group were the oldest and had the longest disease duration. The LAR group had the highest HAQ-DI and, after the SML group, the second highest SDAI.

The concordance rate between the real four groups as observed in 2016 and the predicted groups from the vectors estimated using the transformation matrix was 56.6%. Positive predictive values (PPV) of the LDA, EVN, SML, and LAR groups were 70.9, 36.2, 53.1, and 58.7%, respectively. Comparing the real groups of 2015 and those of 2016, 51.4% of the LAR patients stayed in the same group the next year; therefore, using the estimated vectors added 7.3% to the predictive value of the LAR group compared to the real group of 2015. The predicted LAR group had higher SDAI ( $12.3 \pm 9.1$  vs.  $5.9 \pm 6.1$ ,  $p < 0.001$ ) and higher HAQ-DI ( $0.96 \pm 0.88$  vs.  $0.46 \pm 0.68$ ,  $p < 0.001$ ) than the other groups, a trend comparable to that seen in the real LAR group and other groups (SDAI:  $10.7 \pm 6.4$  vs.  $5.3 \pm 6.2$ ,  $p < 0.001$  and HAQ-DI:  $0.85 \pm 0.82$  vs.  $0.42 \pm 0.65$ ,  $p < 0.001$ ).

Using the real LAR group of 2015 as a predictor for the LAR group of 2016, sensitivity, specificity, PPV, and negative predictive value (NPV) were 54.5%, 87.6%, 51.4%, and 88.9%, respectively. Positive likelihood ratio (LR) = sensitivity/(1 – specificity) was 4.4. On the other hand, sensitivity, specificity, PPV, NPV, and LR were 31.1%, 94.8%, 58.7%, 85.1%, and 5.9, respectively, when the predicted LAR group was used as a predictor.

Table 2 shows LR in predicting the LAR group with estimated vectors from various target patient groups using different transformation matrices. LR using target patients with less than 10 years disease duration was higher than that using patients with 10 years or more. Methotrexate (MTX) users had higher LR than non-users. When a condition of patients for computing the transformation matrix was matched with that of target patients, patients with less than 10 years disease duration had the highest LR at 7.5 (Table 2).

Table 3 shows the predictive accuracy of the LAR group according to actual next-year patient status. Serial registration data from 2015 to 2016 for patients without orthopedic surgery were extracted from the *NinJa* database for the validation. Targets were all patients ( $n = 10,118$ ), MTX users ( $n = 6254$ ), bDMARD users ( $n = 2399$ ), and users of both drugs ( $n = 1436$ ). The transformation matrix was computed using data registered in *NinJa* from 2013 to 2014 from patients who matched the conditions of the target patients. In MTX users, PPV, NPV, and LR in predicting the LAR group were 60.8%, 85.9%, and 7.2 respectively. These values were higher than those in all patients. In MTX



**Table 2 Positive likelihood ratio in predicting the large-joint-dominant patients**

Transformation matrix derived from patients	Target patients in 2015				
	All	disease duration		MTX	
		< 10 years	≥ 10 years	Users	Non-users
Of all	5.9	6.0	5.6	6.6	5.8
With disease duration (years)					
< 10	6.8	7.5	5.9	7.5	6.9
≥ 10	5.3	5.5	4.9	5.7	5.4
Treated					
With MTX	6.4	6.6	6.0	7.2	6.3
Without MTX	4.9	5.0	4.5	5.2	5.0

The large-joint-dominant patients of 2016 were predicted with estimated vectors from various target patients in 2015 using different transformation matrices derived from the joint index vectors of patient data serially registered in 2013 and 2014

MTX methotrexate

**Table 3 Prediction of the large-joint-dominant patients with estimated vectors from target patients stratified by medication**

Target patients in 2015 (number of patients)	PPV (%)	NPV (%)	LR
All patients (10,118)	58.7	85.1	5.9
MTX users (6254)	60.8	85.9	7.2
bDMARD users (2399)	67.9	82.3	7.4
MTX and bDMARD users (1436)	68.4	84.0	8.6

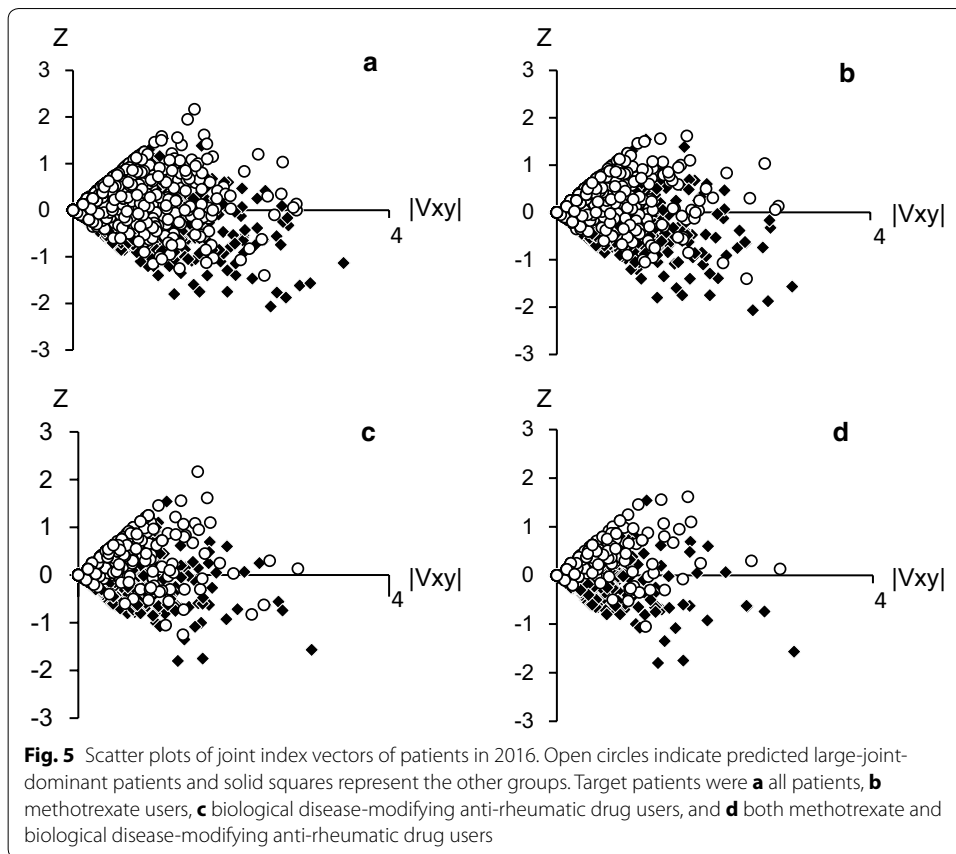
Transformation matrices were derived from the patients in 2013 and 2014 with concordant medication of target patients

PPV positive predictive value, NPV negative predictive value, LR positive likelihood ratio = sensitivity/(1 – specificity), MTX methotrexate

and bDMARD users, PPV, NPV, and LR were 68.4%, 84.0%, and 8.6, respectively. Figure 5 shows the scatter plot of joint index vectors in 2016. The predicted patients of the LAR group are shown by open circles. MTX and bDMARD users are distributed closer to the upper part compared to all patients (Fig. 5a, d).

## Discussion

As RA is an autoimmune disease that affects the synovial joints, clinical examination of inflamed joints is essential and important in order to assess the disease activity in patients with RA [8]. In this context, several articular indices have been developed, including the Ritchie articular index, which grades patients according to the severity of pain in tender joints [9], the Lansbury articular index, which assigns a weighted grade according to the joint surface area [10], and the 68/66 and 28 simple inflamed joint count methods [11, 12]. Thompson et al. [13] reported that the simultaneous presence of joint tenderness and swelling weighted for joint size yielded higher correlation with inflammation than did simple counts. On the other hand, Prevoo et al. [14] found that the validity and reliability of these articular indices did not differ substantially, and proposed that the 28-joint count index, not graded and not weighted, was preferable because of its simplicity. Currently, Disease Activity Score 28 [15] and SDAI [5], which incorporates



the 28-joint count index, are commonly used for assessing disease activity in patients with RA.

A weighted articular index requires complex calculations, but computing can reduce this burden. We proposed a novel joint index that consists of the sum of tender and swollen joint counts in a particular joint category (upper/large, upper/small, lower/large, and lower/small) divided by the number of evaluable joints in that category [3]. Then we found that each joint category affected activity-related and damage-related physical function (HAQ) differently: large joint involvement was associated with an increase in both activity-related and damage-related HAQ, whereas upper/small joint involvement was the most significant predictor of activity-related HAQ [4]. Joint index vector  $V(x, y, z)$  consists of the joint index of the upper limbs ( $x$ ), that of the lower limbs ( $y$ ), and large- or small-joint dominance ( $z$ ). A strong correlation was found between  $|V_{xy}|$  and SDAI, and  $z$  increased linearly as functional damage progressed. You can easily understand a patient's condition graphically separating activity and damage components. The 28 simple inflamed joint count method, which ignores joint size and its distributions, does not have this merit of the joint index vector. The mean value of  $|V_{xy}|$  in patients with low disease activity ( $SDAI \leq 11$ ) was around 0.1, and 5% of upper or lower limit of  $z$  value ( $-4$  to  $4$ ) is  $\pm 0.2$ ; therefore, we temporarily used  $|V_{xy}| = 0.1$  and  $|z| = 0.2$  as cut-off points. Then patients in the LAR group with ( $|V_{xy}| > 0.1$  and  $z > 0.2$ ) had relatively high disease activity and poor physical function. Since this is the first attempt in the world to

forecast joint involvement using the joint index vector, we do not have any established cut-off scores. We believe further prospective studies using big data will optimize and validate our data.

The likelihood ratio (LR) in predicting the LAR group using a transformation matrix was 5.9, indicating that, if prior probability of belonging to the LAR group was 20%, the odds ratio ( $= 0.2/0.8 = 0.25$ ) rises to approximately 1.5 ( $\approx 0.25 \times 5.9$ ), and the posterior probability of the LAR group increases to 60% ( $0.6/0.4 = 1.5$ ). LR in predicting the LAR using a transformation matrix was superior to that using the former-year group as a predictor (5.9 vs. 4.4). Patients in the predicted LAR group had higher SDAI and HAQ-DI than other patients, comparable to the trends seen in the real LAR group, which indicates that the prediction of the LAR group was relatively accurate.

The LR in predicting the LAR group from among target patients with less than 10 years disease duration was higher than that from among patients with 10 years or more. When we used a transformation matrix from patients with or without MTX for predicting the LAR group, the LR was higher among patients using MTX than among those not using it (Table 2). Predictive accuracy increased when we used a transformation matrix from patients corresponding to the target patients; LR was over seven both in patients with less than 10 years and in those using MTX. In addition, LR in predicting the LAR group from among patients using both MTX and bDMARDs reached the highest value, 8.6 (Table 3). These findings indicate that predictive accuracy is greatest in tightly controlled populations with short duration of RA; therefore, this predictive method will be beneficial for clinical practice.

Joint index vector has other advantages besides predicting high disease activity and poor function among patients in the LAR group. The fact that the joint index vector is easily understood is an advantage in dealing with big data. It aids visual comprehension as the graphical depiction of its results clearly shows the patients in the LAR group located in the upper part of the diagram (Fig. 1b). The predicted LAR group using target patients who use MTX and bDMARDs is distributed closer to the upper part of the diagram, as is the real LAR group (Fig. 5). Moreover, this prediction method can be used for patients treated with IL-6 inhibitor, which strongly suppresses C-reactive protein, as well as for patients taking other DMARDs, since joint index vector is derived from pure joint assessment without laboratory data. Furthermore, joint index vector permits physicians to estimate disease activity and functional damage even if a patient's assessment or laboratory data are missing, as  $|V_{xy}|$  and  $z$  are correlates with SDAI and HAQ-DI, respectively; joint index vector encourages utilization of missing data effectively.

RA is not a simple monogenic disease, but a polygenic disease whereby environmental and multiple genetic loci increase the individuals risk of developing disease. Successful application of common polygenic modeling approaches would require sample sizes greater than 1000 individuals for traits with less than 50% heritability [1]. The joint index vector enables us to handle 10,000 or more data and to discriminate patients with high disease activity and poor physical function from others. Pharmacogenetic and genomic studies have the potential to enable precision medicine by providing biomarkers to target the right drug to the right patients; however, a large number of studies results to date have been disappointing and have not yielded a change in clinical practice [1]. One of the reasons is a difficulty to capture response in RA. Many research

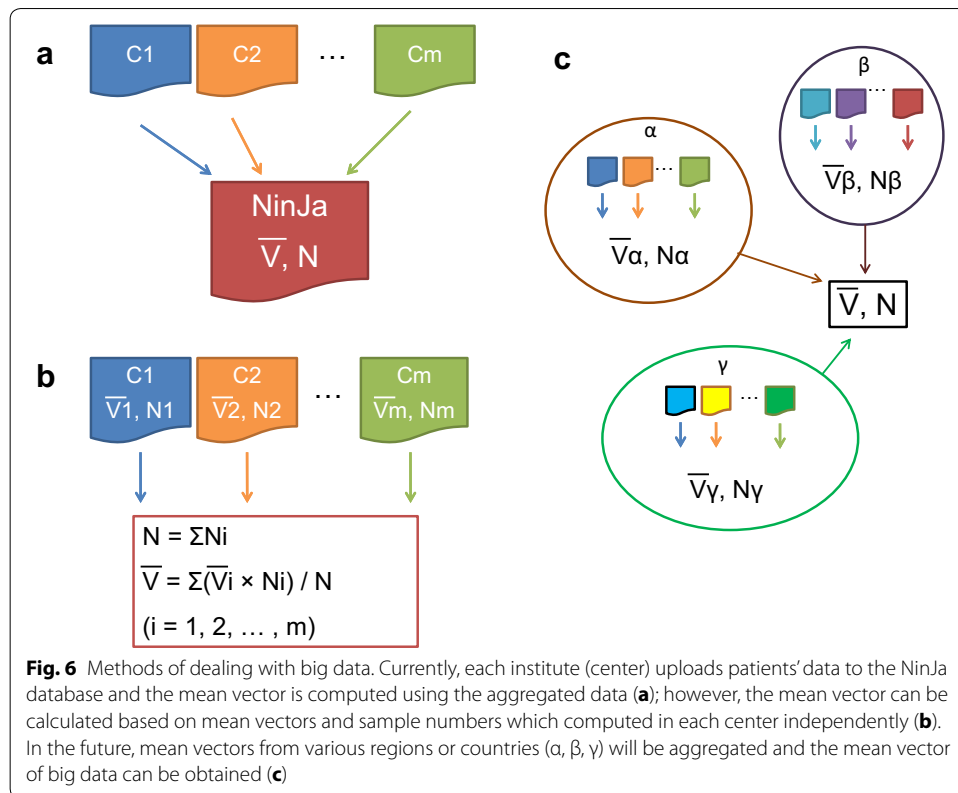
studies use composite measures such as DAS28, which include subjective measures of disease and are known to have a placebo effect [1]. Researchers are searching a more personalized approach toward treatment to achieve rapid remission in every patient to prevent disability in a cost-effective manner [2], and our method is likely to meet the requirements. The predictive accuracy using the joint index vector and its transformation matrix was greatest among target patients who used both MTX and bDMARDs. An appropriate transformation matrix may enable us to predict which agent will have an optimal effect on each patient before administering any drug. In combination with biological markers such as rheumatoid factors and anti-citrullinated protein antibodies, this method using the joint index vector may increase the accuracy with which we can predict which patients will have poor prognosis and can contribute to precision medicine.

Our method is not restricted to the *NinJa* database. A larger scale of database that fulfills the following requirements would be available.

1. Affected joints and other information of each patient are entered into the database annually. At least bilateral shoulder, sternoclavicular, elbow, wrist, proximal interphalangeal, metacarpophalangeal, hip, knee, ankle, tarsometatarsal, and metatarsophalangeal joints must be evaluated whether tenderness and/or swelling exist.
2. Patients' information should include age, disease duration, therapeutic agents, inflammatory markers (C-reactive protein, erythrocyte sedimentation rate), titers of rheumatoid factor and anti-citrullinated protein antibody, patients' global assessment, physicians' global assessment, and HAQ-DI.
3. Security to protect from leaking personal information must be established.

Currently, each institute (center) uploads patients' data to the *NinJa* database and the mean vector is computed using the aggregated data (Fig. 6a); however, the mean vector can be calculated based on mean vectors and sample numbers which computed in each center independently (Fig. 6b). If 100 centers compute 1000 patient data in parallel, then the mean vector of 100,000 patients can be calculated easily and speedy. In the future, mean vectors from various regions or countries will be aggregated and the mean vector of big data can be obtained (Fig. 6c).

Transformation matrices calculated from stratified patients' data would be prepared for users. For example, if there was a seropositive (rheumatoid factor and anti-citrullinated protein antibodies positive) RA patients with less than 5 years duration of the disease and treated with MTX but not with bDMARDs, you could search the database and get a transformation matrix 1 [T1] matched the condition of this patient. Then you could get another transformation matrix 2 [T2] from seropositive RA patients with less than 5 years duration of the disease and treated with MTX and bDMARDs. Comparing the estimated vectors applying [T1] and [T2] would help you make a decision to start bDMARDs or not. As mobile devices are ubiquitous now, big data in a dynamic and real-time environment are analyzed with wireless sensor networks [16]. In the future, a physician will get a transformation matrix immediately when entering a condition into a mobile device by wireless sensor networks, that aggregates computed mean vectors of



patients who matched the condition in each center. And a new dataset generated from feedback data will be used for improving predictive accuracy.

New methods of big data analysis have been emerging and predictive accuracy is improving; for example, Ke et al. [17] proposed a feature learning algorithm based on the adaptive independent subspace analysis and this method showed higher classification accuracy than the independent component analysis. This computer learning method needs relatively small sample data to predict larger data. Applying this method on change in pattern of involved joints may predict single joint involvement separately.

One limitation of this study is the relatively low diversity of patients' demographic features and therapeutic agents. Data were collected from patients at 51 participating institutes. Almost all of the institutes are teaching hospitals for rheumatology approved by the Japan College of Rheumatology, and the patients are treated under a tightly controlled (treat to target) strategy in general. Another limitation was the assessment manner was left to each physician; therefore, there may be differences in joint assessment among physicians, which affects predictive accuracy. Despite these limitations, this study indicates that a transformation matrix derived from joint index vectors can be used to predict the next-year status in each patient from multicenter non-interventional data. As computing can reduce the time required to calculate the joint index vector, the burden of applying the joint index vector in daily practice is low.

## Conclusions

Joint index vector, a novel joint assessment measure, consists of the proportion of affected joints in the upper and lower limbs, along with large- or small-joint dominance. The joint index vector shows big data visually and enables us to identify patients with high disease activity and poor physical function. A transformation matrix derived from the joint index vector can predict the next-year status of individual patients, which will help them receive precision medicine. Since forecasting joint involvement is not perfect now, further investigations are needed to improve predictive accuracy by combining other predictive factors with the joint index vector.

## Additional files

[Additional file 1.](#) DataSet of 11,013 patients without missing values for SDAI or HAQ-DI and serial registration data from patients without orthopedic surgery between 2013 and 2014 ( $n = 10,206$ ) and between 2015 and 2016 ( $n = 10,118$ ). Transformation matrices and estimated vectors are also included.

[Additional file 2.](#) Sample Data for testing EXCEL macro.

## Abbreviations

Jl: joint index; UL: upper/large; US: upper/small; LL: lower/large; LS: lower/small; LDA: low disease activity; EVN: evenly affected; SML: small joint dominant; LAR: large joint dominant; SDAI: Simplified Disease Activity Index; HAQ-DI: the disability index of the Health Assessment Questionnaire; RA: rheumatoid arthritis; bDMARDs: biological disease-modifying anti-rheumatic drugs; *NinJa*: The National Database of Rheumatic Diseases in Japan; PPV: positive predictive value; NPV: negative predictive value; LR: likelihood ratio; MTX: methotrexate.

## Authors' contributions

SN and JN conceived and designed this study. SN analyzed and interpreted the patient data. TS drafted the manuscript critically. ST made substantial contribution to acquisition of data. All authors read and approved the final manuscript.

## Author details

<sup>1</sup> Rheumatic Disease Center, Kurashiki Medical Center, 250 Bakuro-cho, Kurashiki-shi, Okayama 710-8522, Japan. <sup>2</sup> Department of Rheumatology, Tokyo Medical University Hospital, 6-7-1 Nishishinjuku, Shinjuku-ku, Tokyo 160-0023, Japan.

<sup>3</sup> Orthopedic Surgery, The University of Tokyo Hospital, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8655, Japan. <sup>4</sup> Department of Rheumatology, National Hospital Organization Tokyo National Hospital, 3-1-1 Takeoka, Kiyose-shi, Tokyo 204-8585, Japan.

## Acknowledgements

We thank Mayumi Yokoyama and Akiko Komiya for providing assistance with the cleaning and maintenance of the *Ninja* database.

## Competing interests

ST was supported by research grants from Abbott Japan Co., Ltd., Astellas Pharma Inc., Chugai Pharmaceutical Co., Ltd., Eisai Co., Ltd., Mitsubishi Tanabe Pharma Corporation, Pfizer Japan Inc., Takeda Pharmaceutical Company Limited, and Teijin Pharma Limited. ST received honoraria from Asahi Kasei Pharma Corporation, Astellas Pharma Inc., AbbVie GK., Chugai Pharmaceutical Co., Ltd., Ono Pharmaceutical Co., Ltd., Mitsubishi Tanabe Pharma Corporation, and Pfizer Japan Inc.

## Availability of data and materials

All data generated or analyzed during this study are included in this published article and its Additional files 1, 2.

## Consent for publication

The authors consent to the publication of this work.

## Funding

This work was supported by Grants-in-Aid for Japan Agency for Medical Research and Development (AMED).

## Publisher's Note

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

Received: 20 July 2018 Accepted: 4 October 2018

Published online: 13 October 2018

# References

1. Bluett J, Barton A. Precision medicine in rheumatoid arthritis. *Rheum Dis Clin North Am*. 2017;43:377–87. <https://doi.org/10.1016/j.rdc.2017.04.008>.
2. Wijnbrandts CA, Tak PP. Prediction of response to targeted treatment in rheumatoid arthritis. *Mayo Clin Proc*. 2017;92:1129–43. <https://doi.org/10.1016/j.mayocp.2017.05.009>.
3. Nishiyama S, Aita T, Yoshinaga Y, Kishimoto H, Toda M, Yoshihara Y, Miyoshi S, Manki A, Miyawaki S. Proposing a method of regional assessment and a novel outcome measure in rheumatoid arthritis. *Rheumatol Int*. 2012;32:2569–71. <https://doi.org/10.1007/s00296-011-2058-9> (Epub 2011 July 26).
4. Nishiyama S, Nishino J, Tohma S. Importance of large joint involvement in functional disability of patients with rheumatoid arthritis. *Intern Med Rev*. 2016. <https://doi.org/10.18103/imr.v2i8.155>.
5. Smolen JS, Breedveld FC, Schiff MH, Kalden JR, Emery P, Eberl G, van Riel PL, Tugwell P. A simplified disease activity index for rheumatoid arthritis for use in clinical practice. *Rheumatology (Oxford)*. 2003;42:244–57.
6. Fries JF, Spitz P, Kraines RG, Holman HR. Measurement of patient outcome in arthritis. *Arthritis Rheum*. 1980;23:137–45.
7. Bruce B, Fries JF. The Stanford health assessment questionnaire: dimensions and practical applications. *Health Qual Life Outcomes*. 2003;1:20. <https://doi.org/10.1186/1477-7525-1-20>.
8. Bombardier C, Tugwell P, Sinclair A, Dok C, Anderson G, Buchanan WW. Preference for endpoint measures in clinical trials: results of structured workshops. *J Rheumatol*. 1982;9:798–801.
9. Ritchie DM, Boyle JA, McInnes JM, Jasani MK, Dalakas TG, Grieveson P, Buchanan WW. Clinical studies with an articular index for the assessment of joint tenderness in patients with rheumatoid arthritis. *Q J Med*. 1968;37:393–406.
10. Lansbury J, Haut DD. Quantitation of the manifestations of rheumatoid arthritis. 4. Area of joint surfaces as an index to total joint inflammation and deformity. *Am J Med Sci*. 1956;232:150–5.
11. The cooperating clinics committee of the American Rheumatism Association. A seven-day variability study of 499 patients with peripheral rheumatoid arthritis. *Arthritis Rheum*. 1965;8:302–34.
12. Fuchs HA, Brooks RH, Callahan LF, Pincus T. A simplified twenty-eight-joint quantitative articular index in rheumatoid arthritis. *Arthritis Rheum*. 1989;32:531–7.
13. Thompson PW, Silman AJ, Kirwan JR, Currey HLF. Articular indices of joint inflammation in rheumatoid arthritis. Correlation with the acute-phase response. *Arthritis Rheum*. 1987;30:618–23.
14. Prevoo ML, van Riel PL, van't Hof MA, van Rijswijk MH, van Leeuwen MA, Kuper HH, van de Putte LB. Validity and reliability of joint indices. A longitudinal study in patients with recent onset rheumatoid arthritis. *Br J Rheumatol*. 1993;32:589–94.
15. Prevoo ML, van't Hof MA, Kuper HH, van Leeuwen MA, van de Putte LB, van Riel PL. Modified disease activity scores that include twenty-eight-joint counts. Development and validation in a perspective longitudinal study of patients with rheumatoid arthritis. *Arthritis Rheum*. 1995;38:44–8.
16. Wei W, Qi Y. Information potential fields navigation in wireless *Ad-Hoc* sensor networks. *Sensors*. 2011;11:4794–807. <https://doi.org/10.3390/s110504794>.
17. Ke Q, Zhang J, Song H, Wan Y. Big data analytics enabled by feature extraction based on partial independence. *Neurocomputing*. 2018;288:3–10.

**Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:**

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

---

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)