Scoring of the radiological picture of idiopathic interstitial pneumonia: a study to verify the reliability of the method

Eva Kocova¹, Jiri Vanasek¹, Vladimir Koblizek², Jakub Novosad⁴, Pavel Elias¹, Vladimir Bartos² and Martina Sterclova³

Abstract

Background: Idiopathic pulmonary fibrosis (IPF) is a clinical form of usual interstitial pneumonia (UIP). Computed chest tomography (CT) has a fundamental role in the multidisciplinary diagnostics. However, it has not been verified if and how a subjective opinion of a radiologists or pneumologists can influence the assessment and overall diagnostic summary.

Purpose: To verify the reliability of the scoring system.

Material and Methods: Assessment of conformity of the radiological score of high-resolution CT (HRCT) of lungs in patients with IPF was performed by a group of radiologists and pneumologists. Personal data were blinded and the assessment was performed independently using the Dutka/Vasakova scoring system (modification of the Gay system). The final score of the single assessors was then evaluated by means of the paired Spearman’s correlation and analysis of the principal components.

Results: Two principal components explaining cumulatively a 62% or 73% variability of the assessment of the single assessors were extracted during the analysis. The groups did not differ both in terms of specialty and experience with the assessment of the HRCT findings.

Conclusion: According to our study, scoring of a radiological image using the Dutka/Vasakova system is a reliable method in the hands of experienced radiologists. Significant differences occur during the assessment performed by pneumologists especially during the evaluation of the alveolar changes.

Keywords
Usual interstitial pneumonia (UIP), score system, HRCT, reliability, alveolar pattern, interstitial pattern

Date received: 9 February 2015; accepted: 22 August 2015

Introduction

Idiopathic pulmonary fibrosis (IPF) has been recently defined as a specific form of chronic progressive fibrosing interstitial pneumonia of unknown etiology (1). It is associated with a various degree of interstitial fibrosis and fibrous destruction of the pulmonary tissue. Determination of clinical diagnosis of IPF is based on an interdisciplinary opinion of pneumologists, radiologists, and pathologists regarding the condition. There has been a long endeavor to create an integrated radiological scoring system that would be easily interpretable and repeatable while providing objective and valid information about the severity of the disease. A number of scoring systems have been developed (2,3).

¹Department of Radiology, University Hospital and Faculty of Medicine, Hradec Kralove, Czech Republic
²Department of Pneumology, University Hospital and Faculty of Medicine, Hradec Kralove, Czech Republic
³Department of Pneumology First Faculty of Medicine, Thomayer Hospital, Prague, Czech Republic
⁴Institute of Clinical Immunology and Allergology, University Hospital and Faculty of Medicine, Hradec Kralove, Czech Republic

Corresponding author: Eva Kocova, Department of Radiology, University Hospital and Faculty of Medicine, Sokolska 581, Hradec Kralove, Czech Republic.
Email: eva.kocova@fnhk.cz
IPF is a disease of unknown etiology. A concurrence of the environmental factors in a genetically predisposed individual has been considered. The most common supposed factors include cigarette smoking, exposure to organic antigens, metal dusts like steel and lead. Furthermore, viral infections such as hepatitis C, cytomegalovirus, influenza virus, Ebstein-Barr virus, parainfluenza, measles, herpesvirus 6, mycoplasma, and human immunodeficiency virus (HIV) have also been considered. Last but not least, the genetic factors and gastroesophageal reflux are under discussion.

The reported incidence worldwide has been reported to be between 1/100,000 and 2–29/100,000, respectively. However epidemiological data are probably underevaluated as this condition is often underdiagnosed. Men are affected more often than women. Incidence increases also with age – most of the patients with IPF are aged over 50 years. Survival is reported to be between 2–3 years from the date of diagnosis and, despite aggressive therapy the mortality due to this disease, is very high.

The clinical condition of a patient is critical in correlation with the radiological picture of usual interstitial pneumonia (UIP). At the same time other forms of interstitial pneumonia must be completely excluded – interstitial pneumonia associated with exposure to exogenic agents, with systemic connective tissue disease, and other forms of idiopathic interstitial pneumonias. The clinical symptoms of IPF include prolonged cough, dyspnea, and fatigue. Crackles located basally bilaterally and finger clubbing are present.

The purpose of the study was to verify the reliability of a scoring system of high-resolution computed tomography (HRCT) changes of UIP using the Dutka/Vasakova scoring system.

**Material and Methods**

**Patients**

The first 20 consecutive patients (the first 10 women and first 10 men) from our IPF hospital database from the years 2009-2014 were included in the study. All patients had typical clinical symptoms of IPF and all of them had histological verification of UIP. The mean age of patients was 65 years (age range, 39–84 years). Patient data were blinded and the assessors were not informed about the clinical status of the patients.

**Assessors**

In our cross-over study we compared evaluation of the alveolar and interstitial score by a team of 13 radiologists and pneumologists (6 radiologists, 7 pneumologists). All radiologists were at least partially specialized in chest radiology, three were consultants in radiology (marked as 1, 2, and 6 in the analysis) and three were older residents (marked as 3, 4, and 5 in the analysis). Two out of the group of radiologists used scoring routinely (marked as 1 and 3 in the analysis). The pneumologists also included consultants in pneumology (marked as 7, 8, 10, and 12 in the analysis) and residents (marked as 9, 11, and 13 in the analysis). Two pneumologists focused on the issue of interstitial pulmonary processes (marked as 8 and 9 in the analysis). The assessment was performed individually and the assessors could not influence the assessment of their colleagues.

**Scoring system**

Evaluation in the Dutka/Vasakova scoring system is performed on four levels: aortic arch; carina; maximum diameter of the right ventricle; and top of the right diaphragmatic dome (Fig. 1). The right and left lungs should be evaluated separately. The HRCT picture is divided by sectors of 5 percent each (Fig. 2). The percentage representation in the single levels is evaluated separately for the alveolar and interstitial changes according to the scheme: no affection = 0; affection up to 5% = 1; 5–25% = 2; 25–50% = 3; 50–75% = 4; above 75% = 5. The extent of affection in terms of alveolar and interstitial changes is then counted up separately from each level for each lung, the values for both lung (right and left) are then added and the result is averaged and rounded to the whole number. The result includes...
two values representing alveolar and fibrous affection, respectively.

**Statistical analysis**

Spearman’s correlation analysis was used for the initial statistical evaluation of 13 variables (corresponding to 13 assessors of the alveolar or interstitial score). The correlation and covariance matrix were subsequently analyzed using the principal component analysis (PCA) with the method of oblique rotation (oblimin) and Keiser’s normalization. The Keiser-Meyer-Olkin (KMO) criterion was regarded for the assessment of suitability of the model and how it exceeded the value of 0.8 in the alveolar and interstitial score (the limit of acceptability is conventionally determined as 0.5). In both cases there was a highly significant Bartlett test spherically \( P < 0.001 \) indicating a significant correlation among the single assessors which enabled performance of reasonable PCA.

The purpose of the analysis of the principal components is a reduction of the number of variables on the basis of their variability and mutual relations. The principal components occur as a linear combination of the original variables (single assessors). They are new variables which measure a different feature (dimension) of the analyzed variables which are not in a mutual correlation. Using these newly proposed variables we can explain a non-negligible percentage of variability in decisions of the single assessors and reveal distinction or similarity in their assessment of the same clinical situation. The rate of correlation among the single variables and newly proposed components is marked as a factor load or component fulfilment.

The correlation matrixes are presented in the tables for the single evaluated scores (Tables 1 and 2), furthermore the component fulfilment for the single assessors with the values below 0.3 being suppressed (Tables 3

**Fig. 2.** Dividing of HRCT scan to areas for evaluating of severity of UIP – the picture is divided into sectors of 5% (the whole scan is 100 percent).

|   | A_1 | A_2 | A_3 | A_4 | A_5 | A_6 | A_7 | A_8 | A_9 | A_10 | A_11 | A_12 | A_13 |
|---|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|------|------|------|
| A_1 | 1.000 | 0.612* | 1.000* | 0.893* | 0.665* | 0.484* | 0.256 | 0.409 | 0.339 | 0.379 | 0.890* | 0.372 | 0.308 |
| A_2 | 0.612* | 1.000 | 0.612* | 0.598* | 0.406 | 0.407 | 0.288 | 0.421 | 0.261 | 0.409 | 0.657* | 0.216 | 0.660* |
| A_3 | 1.000* | 0.612* | 1.000 | 0.893* | 0.665* | 0.484* | 0.256 | 0.409 | 0.339 | 0.379 | 0.890* | 0.372 | 0.308 |
| A_4 | 0.893* | 0.598* | 0.893* | 1.000 | 0.669* | 0.475* | 0.221 | 0.564* | 0.551† | 0.304 | 0.756* | 0.342 | 0.492† |
| A_5 | 0.665* | 0.406 | 0.665* | 0.669* | 1.000 | 0.472‡ | 0.321 | 0.388 | 0.306 | 0.223 | 0.642* | 0.159 | 0.249 |
| A_6 | 0.484† | 0.407 | 0.484† | 0.475‡ | 0.472‡ | 1.000 | 0.360 | 0.289 | 0.239 | 0.312 | 0.666* | 0.546† | 0.259 |
| A_7 | 0.256 | 0.288 | 0.256 | 0.221 | 0.321 | 0.360 | 1.000 | 0.220 | 0.331 | 0.604* | 0.452† | 0.567* | 0.077 |
| A_8 | 0.409 | 0.421 | 0.409 | 0.564* | 0.388 | 0.289 | 0.220 | 1.000 | 0.847* | 0.569* | 0.318 | 0.097 | 0.366 |
| A_9 | 0.339 | 0.261 | 0.339 | 0.551† | 0.306 | 0.239 | 0.331 | 0.847* | 1.000 | 0.453‡ | 0.290 | 0.171 | 0.238 |
| A_10 | 0.379 | 0.409 | 0.379 | 0.304 | 0.223 | 0.312 | 0.604* | 0.569* | 0.453‡ | 1.000 | 0.406 | 0.528† | 0.210 |
| A_11 | 0.890* | 0.657* | 0.890* | 0.756* | 0.642* | 0.666* | 0.452‡ | 0.318 | 0.290 | 0.406 | 1.000 | 0.404 | 0.348 |
| A_12 | 0.372 | 0.216 | 0.372 | 0.342 | 0.159 | 0.546* | 0.567* | 0.097 | 0.171 | 0.528† | 0.404 | 1.000 | 0.160 |
| A_13 | 0.308 | 0.660* | 0.308 | 0.492† | 0.249 | 0.259 | 0.077 | 0.366 | 0.238 | 0.210 | 0.348 | 0.160 | 1.000 |

The correlation matrix expressing the Spearman’s correlation among the single assessors (in the cells always a value of the correlation coefficient \( \rho \) and its statistical significance), statistically significantly correlated relationships are highlighted in bold.

*Correlation coefficient significant on the level of 0.01 (bi-directional).
†Correlation coefficient significant on the level of 0.05 (bi-directional).
A1–6, radiologists; A7–13, pneumologists.
and 4), and finally the graphs of the component fulfillment for the assessors enabling to graphically differentiate at least two groups of assessors according to the mutual correlation in decision for scoring of the alveolar and interstitial score (Figs 3 and 4). The IBM software SPSS statistics ver. 22 (IBM Corporation, Somers, NY, USA) was used for the statistical analysis.

Table 2. Correlation matrix – interstitial score (Spearman’s correlation coefficient).

|     | I_1  | I_2  | I_3  | I_4  | I_5  | I_6  | I_7  | I_8  | I_9  | I_10 | I_11 | I_12 | I_13 |
|-----|------|------|------|------|------|------|------|------|------|------|------|------|------|
| I_1 | 1.000 | 0.504* | 0.902* | 0.791* | I_2  | I_3  | I_4  | I_5  | I_6  | I_7  | I_8  | I_9  | I_10 | I_11 | I_12 | I_13 |
| I_2 | 0.504* | 1.000 | 0.680* | 0.716* | 0.554* | I_3  | I_4  | I_5  | I_6  | I_7  | I_8  | I_9  | I_10 | I_11 | I_12 | I_13 |
| I_3 | 0.902* | 0.680* | 1.000 | 0.808* | 0.744* | 0.908* | 0.752* | 0.558* | 0.542* | 0.739* | 0.632* | 0.730* | 0.707* |
| I_4 | 0.791* | 0.716* | 0.808* | 1.000 | 0.847* | 0.758* | 0.648* | 0.639* | 0.701* | 0.716* | 0.510* | 0.709* | 0.499* |
| I_5 | 0.800* | 0.554* | 0.744* | 0.847* | 1.000 | 0.729* | 0.470* | 0.481* | 0.478* | 0.635* | 0.416 | 0.613* | 0.407 |
| I_6 | 0.842* | 0.715* | 0.908* | 0.758* | 0.729* | 1.000 | 0.711* | 0.632* | 0.627* | 0.787* | 0.663* | 0.841* | 0.614* |
| I_7 | 0.658* | 0.574* | 0.752* | 0.648* | 0.470* | 0.711* | 1.000 | 0.436 | 0.624* | 0.543* | 0.789* | 0.472* | 0.598* |
| I_8 | 0.453* | 0.671* | 0.558* | 0.639* | 0.481* | 0.632* | 0.436 | 1.000 | 0.784* | 0.528* | 0.520* | 0.665* | 0.550* |
| I_9 | 0.514* | 0.687* | 0.542* | 0.701* | 0.478* | 0.627* | 0.624* | 0.784* | 1.000 | 0.514* | 0.641* | 0.558* | 0.460* |
| I_10 | 0.652* | 0.632* | 0.739* | 0.716* | 0.655* | 0.787* | 0.543* | 0.528* | 0.514* | 1.000 | 0.450* | 0.705* | 0.667* |
| I_11 | 0.561* | 0.400 | 0.632* | 0.510* | 0.416 | 0.663* | 0.789* | 0.520* | 0.641* | 0.450* | 1.000 | 0.436 | 0.571* |
| I_12 | 0.656* | 0.698* | 0.730* | 0.709* | 0.613* | 0.841* | 0.472* | 0.665* | 0.558* | 0.705* | 0.436 | 1.000 | 0.439 |
| I_13 | 0.555* | 0.487* | 0.707* | 0.499* | 0.407 | 0.614* | 0.598* | 0.550* | 0.460* | 0.667* | 0.571* | 0.439 | 1.000 |

The correlation matrix expressing the Spearman’s correlation among the single assessors (in the cells always a value of the correlation coefficient rho and its statistical significance), statistically significantly correlated relationships are highlighted in bold.

*Correlation coefficient significant on the level of 0.05 (bi-directional).
*Correlation coefficient significant on the level of 0.01 (bi-directional).
11–6, radiologists; 17–13, pneumologists.

Table 3. Analysis of the main components – components fulfilment (alveolar score).

| Component | 1    | 2    |
|-----------|------|------|
| A_1       | 0.941|      |
| A_3       |      | 0.941|
| A_4       | 0.928|      |
| A_11      | 0.824|      |
| A_5       | 0.755|      |
| A_2       | 0.709|      |
| A_13      | 0.595|      |
| A_7       | 0.879|      |
| A_10      | 0.874|      |
| A_12      | 0.683|      |
| A_9       | 0.592|      |
| A_8       | 0.309| 0.471|
| A_6       | 0.328| 0.393|

Components fulfillment express a numerical correlation among the single variables and newly proposed variable – component. The level of correlation is critical for assignment of the single variables in the groups.

Table 4. Analysis of the main components – components fulfilment (interstitial score).

| Component | 1    | 2    |
|-----------|------|------|
| I_2       | 0.924|      |
| I_12      | 0.906|      |
| I_4       | 0.890|      |
| I_5       | 0.835|      |
| I_8       | 0.747|      |
| I_9       | 0.648|      |
| I_6       | 0.634| 0.395|
| I_1       | 0.577| 0.377|
| I_3       | 0.563| 0.460|
| I_10      | 0.536| 0.347|
| I_11      | 0.914|      |
| I_7       | 0.794|      |
| I_13      | 0.735|      |

Components fulfillment express a numerical correlation among the single variables and newly proposed variable – component. The level of correlation is critical for assignment of the single variables in the groups.
Results

Two main components (evaluation of the alveolar and interstitial score) explaining a 62% or 73% variability of all assessors were extracted for both situations, respectively. There was a statistically significant correlation on the 5% level of significance in evaluation of the interstitial score both among the radiologists as well as between the radiologists and pneumologists (Tables 1 and 2). The radiologists showed a statistically significant correlation mutually in the assessment of the alveolar score (only one radiologist did not correlate with the others; on the other hand, only one pneumologist correlated with the radiologists). On the other hand, there were also significant differences between the radiologists and pneumologists during evaluation of the given scores. In the group of radiologists the highest conformity in both types of score was among
those assessors who use the scoring system regularly (marked as 1 and 3). In the group of pneumologists there was a statistically significant conformity in the assessment of the alveolar score (only one pneumologist did not correlate with the others), but not in the evaluation of the interstitial score.

Discussion

The diagnosis of IPF is based on clinical symptoms, pulmonary functional tests and HRCT findings.

In the past, HRCT of the chest was considered as a method that can predict response to the administered therapy and mortality of patients based on the amount of the present ground glass (7,9,10). This opinion is considered as obsolete today. The ground glass picture is hypothesized as a possible pattern of initial fibrotic changes (11). Chest examination using CT with a HRCT reconstructions versus chest CT without these reconstructions has a higher ability to find small structures such as interlobular septums, cyst walls, interstitial nodules, and changes of the bronchial wall using collimation of 1–1.5 mm (Fig. 5) (12). Identification of these changes is absolutely necessary for diagnosis of UIP. A positive predictive value of HRCT is reported to be 90–100% in diagnosis of UIP (6). The most common differentially diagnostic problem in the HRCT are non-specific interstitial pneumonia (NSIP), chronic forms of interstitial hypersensitive pneumonia, or sarcoidosis (13). Some studies even reported that differentiation of these forms of pulmonary affection on the basis of the HRCT was solely possible only in 50% of cases (14). Identification of fibrosis and honeycombing is a principal radiological criterion.

The diagnosis of UIP includes also evaluation of the extent of affection by the scoring system. Some scoring systems have been developed (2,3). Some of them were too complicated and they appeared to be useless in common practice. Nevertheless, these more complicated and complex variants correlated well with the real clinical status of patients and functional tests.

The Dutka/Vasakova scoring system that is widely used in the Czech Republic was initially used by Gay SE (7) who evaluated ground glass opacities (CT - alv) and reticular opacities (CT - fib) using the scale from 0 to 5. Subsequently the values were counted and the final CT score (CT-tot) was obtained. The individual lobes were evaluated separately. This system was taken from the study in which the limited examination of the chest only three levels – aortic arch, main carina, and 1 cm above the diaphragm was performed – was compared to the volume evaluation of the whole lung. It was found that the evaluation in the limited examination corresponded very well to the evaluation in the volume examination of the lung. It means that for scoring system it possible to use a conventional HRCT with spacing of 10 mm which markedly reduces the radiation load for patients.

Evaluation of the alveolar and fibrous affection was carried out on the basis of the scheme provided in Table 5. The ground glass opacities are considered to represent alveolar changes, while reticulations and honeycombing represent fibrous changes.

The scoring system Dutka/Vasakova modified Gay's system – CT is evaluated on four levels and only alveolar (CT – alv) and reticular opacities (CT – fib) are counted. In our study we tried to verify the reliability of the scoring system.

According to our analysis there is a statistically significant conformity (on the 0.05 level of significance) among the radiologists in evaluation of the alveolar and interstitial score. However, the conformity between the radiologists and pneumologists is much lower in the evaluation of alveolar score compared to interstitial
score. The reason could be in mistakenly interpreted ground-glass changes among the pneumologists, in their insufficient experience in assessment of the CT findings. Non-conformity of three pneumologists could be explained by a short clinical practice experiences (two of them being young residents). Another limiting factor can be an insufficient number of the diagnostic monitors at the pulmonary department. The limitation of the study is relatively low number of patients in the study. Despite a lower statistical power a probability of randomly created statistically significant but not clinically relevant correlations among the single assessors can paradoxically be increased in this case. However, despite this limitation the presented mathematical model shows high signs of reliability (highly significant Bartlett test of spherical data and a high KMO criterion).

In conclusion, the Dutka/Vasakova scoring system is a reliable method for the assessment of the radiological picture of UIP in the hands of radiologists. However, validity of this method will need to be verified in a larger group of patients. Further standardization in scoring of HRCT scans is needed to reduce misinterpretations. The main advantage of the scoring system is possibility to express severity and progression of this disease by a clearly definable number.

Declaration of conflicting interests
The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding
The author(s) received no financial support for the research, authorship, and/or publication of this article.

References
1. Raghu G, Harold RC, Jim JE, et al. An Official ATS/ERS/JRS/ALAT Statement: Idiopathic pulmonary fibrosis: evidence-based guidelines for diagnosis and management. Am J Respir Crit Care Med 2011;183:788–824.
2. Šterclová M, Vašáková M. Kvantifikace rozsahu postižení u nemocných s fibrotizujícími intersticiálními plícními procesy. Česká radiologie 2013;67:204–208. (in Czech).
3. Sterclova M, Vasakova M, Dutka J, et al. Extrinsic allergic alveolitis: comparative study of the bronchoalveolar lavage profiles and radiological presentation. Postgrad Med J 2006;82:598–601.
4. King TE Jr, Schwarz MI, Brown K, et al. Idiopathic pulmonary fibrosis relationship between histopathologic features and mortality. Am J Respir Crit Care 2001;164:1025–1032.
5. Selman M, King TE, Parod A, et al. Idiopathic pulmonary fibrosis: prevailing and evolving hypotheses about its pathogeneses and implications for therapy. Ann Intern Med 2001;134:136–151.
6. Watadani T, Sakai F, Johkoh T, et al. Interobserver variability in the CT assessment of honeycombing in the lungs. Radiology 2013;266:936–944.
7. Gay SE, Kazerooni EA, Toews GB, et al. Idiopathic pulmonary fibrosis predicting response to therapy and survival. Am J Respir Crit Care Med 1998;157:1063–1072.
8. Kaiser HF. A Second. Generation Little Jiffy. Psychometrika 1970;35:401–415.
9. Akira M, Sakatani M, Ueda E. Idiopathic pulmonary fibrosis: progression of honeycombing at thin-section CT. Radiology 1993;189:687–691.
10. Best AC, Meng J, Lynch AM, et al. Idiopathic pulmonary fibrosis: physiology tests, quantitative CT index, and CT visual scores as predictors of mortality. Radiology 2008;246:935–940.
11. Fujimoto K, Taniguchi T, Johkoh T, et al. Acute exacerbation of idiopathic pulmonary fibrosis: high-resolution CT scores predict mortality. Eur Radiol 2012;22:83–92.
12. Kazerooni EA, Martinez FJ, Flint A, et al. Thin-section CT obtained at 10-mm increments versus limited three-level thin section CT for idiopathic pulmonary fibrosis: correlation with pathologic scoring. Am J Roentgenol 1997;169:977–983.
13. Sverzellati N, Wells AU, Tomassetti S, et al. Biopsy-proved idiopathic pulmonary fibrosis: spectrum of nondiagnostic thin-section CT diagnoses. Radiology 2010;254:957–964.
14. Silva CI, Muller NL, Lynch DA, et al. Chronic hypersensitivity pneumonitis: differentiation from idiopathic pulmonary fibrosis and nonspecific interstitial pneumonia by using thin-section CT. Radiology 2008;246:288–297.