

# The application of 24-hour pharyngeal pH-monitoring and Reflux Finding Score and Reflux Symptom Index questionnaires in the diagnostics of laryngopharyngeal reflux

Elżbieta Włodarczyk, Beata Miaśkiewicz, Danuta Raj-Koziak, Agata Szkietkowska, Piotr H. Skarżyński, Henryk Skarżyński

Institute of Physiology and Pathology of Hearing, Kajetany, Poland

Gastroenterology Rev 2019; 14 (4): 274–282  
DOI: <https://doi.org/10.5114/pg.2019.90253>

**Key words:** laryngopharyngeal reflux, 24-hour pharyngeal pH-monitoring, Reflux Symptom Index, Reflux Finding Score.

---

**Address for correspondence:** Elżbieta Włodarczyk PhD, Institute of Physiology and Pathology of Hearing, 17 Mokra St, 05-830 Kajetany, Poland, e-mail: [e.wlodarczyk@ifps.org.pl](mailto:e.wlodarczyk@ifps.org.pl)

## Abstract

**Introduction:** Twenty-four-hour oesophageal pH-monitoring is a gold standard in the diagnostics of gastroesophageal reflux (GERD); however, this examination does not always perform well in patients in whom laryngeal symptoms of reflux are observed.

**Aim:** To test the effectiveness of Reflux Symptom Index (RSI) and Reflux Finding Score (RFS) in confirming the occurrence of laryngopharyngeal reflux (LPR).

**Material and methods:** Eighty-two patients with symptoms suggesting the occurrence of LPR were studied. The mean age of the participants amounted to  $48.79 \pm 12.02$ . The patients were asked to fill in the RSI. The pharynx was assessed using videolaryngostroboscopy and then the RFS was filled in. Next, 24-hour pharyngeal pH-monitoring was performed with the use of a Dx pH – Restech System device. All the statistical analyses were conducted with the use of the R computational environment.

**Results:** The authors attempted to determine the reference values of the RSI and RFS questionnaires for the Polish population. For a very general approximation, normality of the distributions of the results in the RSI and RFS questionnaire can be done and an approximation mean + 2 SD can be used as a reference value. Then, for RSI the recommended cut-off limit of LPR would fluctuate, depending on the study, between 8 and 17, whereas for RFS it would be between 8 and 14.

**Conclusions:** When used alone, RSI/RFS questionnaires do not allow an unambiguous diagnosis of LPR assessed by 24-hour pharyngeal pH-monitoring.

## Introduction

Twenty-four-hour oesophageal pH-monitoring is a gold standard in the diagnostics of gastroesophageal reflux (GERD) because it enables us to observe all the reflux episodes in a patient. This examination does not always perform well in patients in whom laryngeal symptoms of reflux are observed [1]. Pharyngeal pH-monitoring was thought to be a standard management of laryngopharyngeal reflux (LPR). Shaker *et al.* demonstrated that pharyngeal reflux was much more common in people with laryngeal symptoms of reflux than in the group of patients with GERD but without any laryngeal symptoms or in the control group – without the reflux ailments [2].

In order to facilitate LPR diagnostics and increase its availability, Belafsky proposed the use of two questionnaires [3, 4] measuring the intensity of symptoms associated with LPR. The first one, the Reflux Syndrome Index (RSI), is a subjective assessment done by the patient, who assesses his/her daily ailments and their intensity, whereas the second questionnaire, the Reflux Finding Score (RFS), is filled in by a specialist physician based on a laryngeal image obtained in laryngoscopy.

In the initial studies by Belafsky, these questionnaires turned out to be very good tools for recording the effects of treatment of LPR and confirming the effectiveness of the selected therapy. Such use of these questionnaires was the main conclusion in the papers published so far, which was confirmed in many other

studies, e.g. by Vailati *et al.* [5], Reichel *et al.* [6], and Yadlapati *et al.* [7] and does not raise any doubts. However, regardless of the above, Belafsky reported also reference values that would indicate the occurrence of LPR on the basis of 95% confidence intervals of two control groups consisting of 40 (RFS scale) and 25 people (RSI scale), respectively. Despite the fact that even the author himself warned that “the observations indicating reflux occur among the people without the clinical diagnosis of LPR as well” [3], RSI scores higher than 13 and RFS scores above 7 were considered a universal indicator of LPR, commonly used in the diagnostics [5, 8].

There are many reasons for which the diagnostic use of RSI and RFS questionnaires may raise doubts [9]. Firstly, the RSI questionnaire measures the patient’s subjective perception of his or her ailments. Thus, a cultural specificity of the country and the manner of translation of the questionnaire [8, 10] or psychological aspects [6] may be a problem here. Secondly, the questionnaires are usually directed to a specific population of patients and can differ depending on the purpose of the study. It is difficult to estimate to what extent the reference values coincide in such cases, for example when different countries are considered. Problems with repeatability of the RFS questionnaire between different diagnosticians were also reported in the literature [11]. Moreover, the control groups chosen in the study by Belafsky are relatively small, and what is more important, they were selected in such a way as to correspond to the treatment group only in terms of age and sex distribution. From the statistical point of view, a recommendation based on such control groups is burdened with a large error.

All the above-mentioned aspects may be a reason for which the cases of inconsistency between RSI and RFS scores and the observed LPR are present in the literature. In case of RFS, a lack of connections is usually indicated [12, 13], whereas reports concerning the RSI scale are contradictory [14].

Taking into account all the above circumstances, doubts concerning the application of RSI and RFS scales in diagnostics, as well as the lack of a comprehensive study on this issue in relation to the Polish population, the authors considered it important to thoroughly describe the associations between the objective medical observations of reflux and the results of RSI and RFS questionnaires. Twenty-four-hour pharyngeal pH-monitoring was conducted on a group of 82 patients, together with making attempts to perform LPR diagnostics based on the questionnaires.

## Aim

The aim of the study was to test the effectiveness of RSI and RFS in confirming the occurrence of LPR.

## Material and methods

Eighty-two patients with symptoms suggesting the occurrence of LPR were studied. The mean age of the participants amounted to  $48.79 \pm 12.02$  years, with a predominance of women (79%). The study was conducted in accordance with the Declaration of Helsinki, and every patient gave their consent to participate in the study.

A medical interview concerning the symptoms of laryngological, phoniatic, and general diseases was conducted in all the patients. The patients were asked to fill in the RSI questionnaire. Pharynx was assessed using videolaryngostroboscopy, and then the RFS questionnaire was filled in. Next, 24-hour pharyngeal pH-monitoring was performed with the use of a Dx pH – Restech System device. All the procedures were conducted in accordance with the Declaration of Helsinki.

The Restech® pH sensor was calibrated in solutions of pH 7 and pH 4 prior to use. The sensor was inserted until the flashing LED was seen in the back of the subject’s throat and then positioned so that the flashing light was 5–10 mm below the uvula. The length of the LED light was 5 mm, and it served as a useful guide for placement. The catheter was secured to the patient’s face, as close to the nares as possible using a Tegaderm™ and then passed over the ear and secured to the neck with a second Tegaderm™. The transmitter at the end of the catheter was either taped to the skin or attached to the subjects’ clothing using a clip-on case. A data recorder was attached to the patients’ belt. Patients were asked not to shower during the recording period and to keep a diary indicating the time of the meal periods and the time spent in the supine and upright positions. The meal periods were excluded in the analyses of pharyngeal pH recordings. The Restech® data recorder was downloaded to a proprietary software program and correlated with the patient’s diary.

## Statistical analysis

All the statistical analyses were conducted with the use of the R [15] computational environment. Conventionally, it was accepted that a *p*-value below 0.05 indicates rejection of the zero hypothesis. The Cramer-von-Mises test was applied to verify the normality of the data. In the case of comparisons with other studies that reported only the mean and standard deviation of the measurements, two-sample Student’s *t*-test for different variances was used for testing. In the case of comparisons between the groups in the present study, Student’s *t*-test was applied as well, for the coherence, but the conclusion was confirmed every time with the use of non-parametric Mann-Whitney-Wilcoxon test.

**Table I.** Results of the elements of the RSI questionnaire

Variable	Mean	SD	Median
RSIscore_sum	23.26	7.20	23.5
RSIscore_1	3.22	1.27	3.0
RSIscore_2	3.59	1.37	4.0
RSIscore_3	3.27	1.41	3.0
RSIscore_4	1.83	1.26	2.0
RSIscore_5	2.18	1.51	2.0
RSIscore_6	1.88	1.21	2.0
RSIscore_7	2.22	1.46	2.0
RSIscore_8	2.87	1.48	3.0
RSIscore_9	2.23	1.38	2.0

**Table II.** Results of the elements of the RFS questionnaire

Variable	Mean	SD	Median
RFSscore_sum	8.12	3.41	8
RFSscore_1	0.29	0.71	0
RFSscore_2	1.41	0.97	2
RFSscore_3	2.34	0.93	2
RFSscore_4	1.01	0.56	1
RFSscore_5	1.06	0.62	1
RFSscore_6	1.40	0.70	1
RFSscore_7	0.15	0.52	0
RFSscore_8	0.44	0.83	0

**Table III.** Results of pH-monitoring examination

Parameter	Mean	SD	Median
Number of vertical episodes	20.87	34.66	7.50
% of vertical time	4.43	8.78	0.76
Number of horizontal episodes	6.00	15.42	0.00
% of horizontal time	1.84	6.85	0.00
Ryan Score vertical	154.06	285.37	49.21
Ryan Score horizontal	7.96	15.54	2.17

Correlations between the variables were calculated as Pearson's correlations. Significance of correlation was assessed using a test based on Student's *t* distribution, with the use of the `cor.test` function of the R package. In the case of multiple testing, the Bonferroni correction was applied [16]. In regression models, standard assumptions on the linear models were accepted, whereas the estimation was conducted using the method of least squares implemented in the `lm` function of the R package.

Values of 13 and 7 were accepted as standard reference values for the RSI and RFS questionnaires, respectively, according to Belafsky's recommendation. In the case of pH monitoring, a Ryan Score greater than 9.4 in the vertical position or greater than 6.8 in the horizontal position was recognised as an indication of a health problem [14, 17].

## Results

The mean value obtained in the RSI questionnaire is  $23.26 \pm 7.2$ , whereas in the RFS questionnaire it amounts to  $8.12 \pm 3.41$ . The tables present the detailed statistics of responses for each element of the RSI (Table I) and RFS (Table II) questionnaires. In the case of pH-monitoring, six different measures were chosen to describe the reflux character during a 24-hour observation – the number of episodes, their percentage duration, and the Ryan Score – each one for the horizontal and the vertical position. The mean results obtained for each of the measures, together with standard deviation, are presented in Table III.

A summary result on the RSI scale may be considered as close to the normal distribution (Cramer-von Mises test,  $p = 0.36$ ), unlike the summary result on the RFS scale, which does not correspond to this model (Cramer-von Mises test,  $p = 0.01$ ). Also, none the pH-monitoring measurements meet the requirements of the normal distribution (Cramer-von Mises test, for each:  $p < 0.001$ ), and what is more they are characterised by high asymmetry (skewness for each pH-monitoring measurement higher than 2.5) and a relatively large number of high measurements (kurtosis for each pH-monitoring measurement higher than 6).

In the majority of patients, the reference values are higher than those set by Belafsky, for the RSI scale it is 95% of the studied patients, whereas for the RFS scale it is 52%. In the case of medical observations, based on Ryan Score, LPR can be observed in 73% of patients.

When analysing an interdependence between the results of pH-monitoring and the RSI and RFS questionnaires, simple correlations were used; multidimensional relations and potential interactions were verified using the regression and non-regression models.

Tables IV and V show the correlations between summary results of the RSI and RFS questionnaires and the pH-monitoring measurements. Statistical significance (*p*-values) for the correlation between summary results of the RSI and RFS questionnaires and the pH-monitoring measurements is presented in Table VI.

Each of the studied relations was described by a very low correlation. The correlation for the questionnaires with none of the pH-monitoring measurements

was higher than 0.2. The correlations were low – they did not exceed the value of 0.22 also in the case of analysing the individual elements of the scale. Such results are contradictory to those that can be found in the literature [17], in which the RSI scale was significantly correlated with the results of pH-monitoring, although the conclusions concerning the RFS scale are similar. The application of Spearman’s correlation did not have any qualitative effect on the result.

**Table IV.** Correlations between the individual elements of the RSI questionnaire and pharyngeal pH-monitoring measurements

	RFSscore_ sum	RFSscore_1	RFSscore_2	RFSscore_3	RFSscore_4	RFSscore_5	RFSscore_6	RFSscore_7	RFSscore_8	RFSscore_9
Ryan Score vertical	-0.01	0.15	-0.01	0.01	-0.12	-0.01	0.01	0.01	-0.07	-0.1
Ryan Score horizontal	0.12	0.17	0.16	0.03	-0.1	-0.02	0.05	0.09	0.18	0.01
Number of vertical episodes	0.17	0.22	0.18	0.043	-0.09	0.02	0.17	0.17	0.14	0.06
Number of horizontal episodes	0.15	0.2	0.15	-0.01	-0.07	0.03	0.11	0.09	0.19	0.07
% of vertical time	0.04	0.16	0.01	0.03	-0.08	0.03	0.02	0.06	-0.03	-0.07
% of horizontal time	-0.02	0.02	0.06	-0.12	-0.16	-0.07	0.08	0.02	0.07	-0.03

**Table V.** Correlations between the individual elements of the RFS questionnaire and pharyngeal pH-monitoring measurements

	RFSscore_ sum	RFSscore_1	RFSscore_2	RFSscore_3	RFSscore_4	RFSscore_5	RFSscore_6	RFSscore_7	RFSscore_8
Ryan Score vertical	-0.01	0.05	0	-0.1	-0.1	0.13	0.01	0.16	-0.03
Ryan Score horizontal	0.06	0.05	0.04	-0.03	0.01	0.21	0.16	-0.08	0.05
Number of vertical episodes	0.06	-0.07	0.17	0	-0.02	0.18	0.06	-0.08	-0.01
Number of horizontal episodes	0.09	-0.03	0.06	-0.01	0.02	0.19	0.15	-0.09	0.01
% of vertical time	0.03	0.08	0.02	-0.05	-0.08	0.17	0.05	0.11	-0.02
% of horizontal time	-0.03	-0.04	-0.12	-0.06	0.01	0.04	0.12	-0.05	0.04

**Table VI.** Statistical significance (*p*-values) for the correlation between summary results of the RSI and RFS questionnaires and the pH-monitoring measurements. The obtained *p*-values were subjected to the Bonferroni correction

Variable	RSI sum	RFS sum	Age
Ryan Score vertical	-0.01	-0.01	0.2
Ryan Score horizontal	0.12	0.06	0.23
Number of vertical episodes	0.17	0.06	-0.01
Number of horizontal episodes	0.15	0.09	0.24
% of vertical time	0.04	0.03	0.17
% of horizontal time	-0.02	-0.03	0.28

**Table VII.** Results of linear regression models with different explained variables

Variable	Intercept	Age	Sex	RSI	RFS	RSI : RFS	R <sup>2</sup> correlated
Ryan Score vertical	577.03	4.65	22.97	-25.48	-83.21	3.20	0.06
<i>P</i> -values	0.06	0.08	0.77	<b>0.02</b>	<b>0.01</b>	<b>0.01</b>	0.08
Ryan Score horizontal	-18.55	0.32	-4.16	0.42	0.57	-0.01	0.02
<i>P</i> -values	0.28	<b>0.03</b>	0.34	0.49	0.75	0.83	0.24
Number of vertical episodes	-29.42	0.07	-13.76	1.83	3.88	-0.13	0.00
<i>P</i> -values	0.45	0.83	0.16	0.19	0.34	0.42	0.38
Number of horizontal episodes	-24.93	0.34	-3.16	0.55	0.81	-0.02	0.04
<i>P</i> -values	0.14	<b>0.02</b>	0.46	0.36	0.65	0.75	0.15
% of vertical time	14.69	0.13	-0.34	-0.67	-2.24	0.09	0.04
<i>P</i> -values	0.13	0.12	0.89	0.05	<b>0.03</b>	<b>0.02</b>	0.16
% of horizontal time	-6.50	0.17	-2.29	0.03	0.00	0.00	0.04
<i>P</i> -values	0.39	<b>0.01</b>	0.23	0.90	1.00	0.94	0.14

In the next stage it was ascertained whether a combination of the two scales might explain the diversity of the pH-monitoring measurements. In order to do this, a model of linear regression in which the individual pH-monitoring measurements were described by the result of the RSI and RFS questionnaires and their interactions and, to control the other effects, by age and sex, was adjusted. Table VII shows the results of estimations of these models together with the *R*<sup>2</sup> coefficient that demonstrates the level of adjustment of the model to the data and *p*-values indicating statistical significance. Lines show the estimation of the power of the effect of the said variable on the explained variable. The “Msex” variable corresponds to a zero-one variable indicating male sex (the basic level assumes that the patient’s sex is female). The “RSI : RFS” variable corresponds to interactions between the results of RSI and RFS questionnaires. *P*-values under the parameter

estimations correspond to their statistical significance (whether they are significantly different from 0, *t*-test), whereas *p*-values under *R*<sup>2</sup> indicate statistical significance of the whole model (whether all the parameters are significantly different from 0, *F*-test).

The adjustment of none of these models is higher than 6% (corrected *R*<sup>2</sup>; *R*<sup>2</sup> does not exceed 12%), which is very low and indicates that even a combination of RSI and RFS questionnaires enables us to describe reflux behaviour measured by pH-monitoring only to a very low extent (or not at all). It is also confirmed by the *F*-test of overall model significance that never exceeds the threshold of statistical significance (*p* > 0.08). When analysing the individual explaining variables it is worth noting that in the case of reflux measurement in a vertical position, the only significant variable was age (the higher the age, the higher the reflux intensity). However, the variables concerning the RSI and RFS

questionnaires and their interaction turned out to be significant for the Ryan Score and the overall reflux duration in a vertical position. Unfortunately, facing the overall non-significance of the entire model, such a result is not reliable and does not enable us to state that there is a significant relation between the RFS and RSI and pH-monitoring, so it does not explain their use in diagnostics.

For integrity and completeness of the analysis, it was also verified to what extent the other type of the model, e.g. a non-linear relation between the RSI and RFS and the pH-monitoring measurements, is able to explain the obtained data. After systematic testing of the most popular transformations of the individual variables (polynomial, logarithmic, exponential) the adjustment of the model to the data was achieved at the maximum level of 30% ( $R^2$  correlated, statistically significant, unpublished). The RSI and RFS questionnaires had a statistically significant influence on the measures determining the vertical episodes. However, similarly as for simple linear regression, the degree of adjustment was so small that it did not enable us to confirm the predictive properties of the RSI and RFS questionnaires. What is more, using non-linear models in the case of a sample size of approx. 80 gives a possibility of formation of the so-called overfitting (modelling of randomness of the specific sample, not the general connections). Regardless of these serious restrictions, even if the results of our non-linear analysis were considered reliable, they would suggest an improper scoring of the RSI and RFS scales, whereas the attempts for improving it would create an impractical tool that would require using a mathematically complex formula. Finally, at the statistical level, it is not possible to justify a significant coexistence of the RSI and RFS results with the pH-monitoring measurements. It is also impossible to indicate a reliable, direct, linear, multidimensional relation, taking into account both scales and their interactions.

## Discussion

### Reference values for the assessment of LPR

Based on control groups of size of 40 and 25 subjects, respectively, Belafsky determined the reference values that would confirm the presence of LPR in a patient. Despite the warnings of the author himself, these values were commonly accepted in practice. The authors made an attempt to investigate to what extent the cut-off points of 13 for the RSI and 7 for the RFS would be an indicator of reflux disease in the studied group of patients.

Studies in which the authors analysed the results of the RSI and RFS questionnaires also for the asymptomatic control groups and then compared them to

Belafsky's control groups were selected from the study literature. Tables VIII and IX present the results of this comparison for the RSI and RFS questionnaires, respectively. All the reported control populations were significantly different ( $p < 0.001$ ) from the control population of Belafsky, based on which the reference values were established. The authors did not have a possibility to assess which reference values result from the other studies, because access to the microdata of these studies would be necessary to do this. However, it can be certainly stated that: a) they would be significantly different from those given by Belafsky, and b) they would differ significantly between themselves as well. For a very general approximation, normality of the distributions of the results in the RSI and RFS scales can be done and an approximation mean + 2 SD can be used as a reference value. Then, for RSI the recommended cut-off limit of LPR would fluctuate, depending on the study, between 8 and 17, whereas for RFS it would be between 8 and 14. This observation confirms to a great extent the necessity for verification and establishment of reference values for the populations of the specific countries or specific applications because they are not comparable. Even for this reason the original recommendations by Belafsky should be used with great caution.

**Table VIII.** Comparison of the results of the RSI questionnaire in literature data

Study	RSI total	N	P-value
Belfasky	11.6 (2)	25	NA
Printza	2.41 (3)	172	< 0.01
Farahat	3.59 (3.93)	100	< 0.01
Schindler	6.3 (5.6)	193	< 0.01
Musser	3.8 (2.25)	10	< 0.01

Groups of asymptomatic patients were selected for the statement. The presented results are mean (SD). The "p-value" column presents the p-value of the t-test in a comparison of the specific study with the study by Belafsky. Statistically significant differences are marked in **bold**.

**Table IX.** Comparison of the results of the RFS questionnaire in literature data

Study	RFS total	N	P-value
Belfasky	5.2 (1.6)	40	NA
Printza	2.41 (3)	172	< 0.01
Musser	8.07 (2.96)	10	< 0.01

Groups of asymptomatic patients were selected for the statement. The presented results are mean (SD). The "p-value" column presents the p-value of the t-test in a comparison of the specific study with the study by Belafsky. Statistically significant differences are marked in **bold**.

**Table X.** pH-monitoring measurements divided into groups according to the standard criteria of diagnosing LPR

pH-monitoring measure	RSI criterion	RFS criterion
	Mean (RSI ≤ 13 vs. RSI > 13)	Mean (RFS ≤ 7 vs. RFS > 7)
N	4 vs. 78	39 vs. 43
Ryan Score vertical	459.42 vs. 138.4, <i>p</i> = <b>0.027</b>	163.73 vs. 145.29, <i>p</i> = 0.772
Ryan Score horizontal	2.17 vs. 8.26, <i>p</i> = 0.448	6.4 vs. 9.38, <i>p</i> = 0.39
Number of vertical episodes	18.5 vs. 20.99, <i>p</i> = 0.89	18.54 vs. 22.98, <i>p</i> = 0.566
Number of horizontal episodes	0 vs. 6.31, <i>p</i> = 0.428	4.21 vs. 7.63, <i>p</i> = 0.319
% of vertical time	11.01 vs. 4.09, <i>p</i> = 0.125	4.54 vs. 4.33, <i>p</i> = 0.914
% of horizontal time	0 vs. 1.93, <i>p</i> = 0.585	2.27 vs. 1.44, <i>p</i> = 0.588

A comparison between the groups using Student's *t*-test. Statistically significant differences are marked in **bold**.

In order to finally assess the reliability of the reference values recommended in the literature, i.e. 13 for RSI and 7 for RFS, the differences in pH-monitoring measurements between the groups were examined divided according to these criteria. The results are presented in Table X.

A standard criterion of LPR diagnostics does not distinguish in a statistically significant manner between the intensity of any of the medically objective measures of reflux intensity. On the contrary, for the vertical Ryan Score, the group with the lower RSI presented the higher indicator (statistically significant).

### Characteristic of patients with abnormal pH-monitoring result

The previous section presents the arguments that raise serious doubts concerning the use of the RSI and RFS questionnaires for diagnostic purposes. The reference values of these scales reported by Belafsky cannot be directly and uncritically applied to the other populations of patients; hence, the authors focused on the analysis of the subgroup of patients, which, according to the readings of pH-monitoring, have reflux episodes.

**Table XI.** Differences in pH-monitoring divided into groups according to the occurrence of LPR

Measure of pH-monitoring	LPR criterion Mean (LPR– vs. LPR+)
N	22 vs. 60
Number of vertical episodes	0.23 vs. 28.43, <i>p</i> = <b>0.001</b>
Number of horizontal episodes	0.59 vs. 7.98, <i>p</i> = 0.054
% of vertical time	0.01 vs. 6.05, <i>p</i> = <b>0.005</b>
% of horizontal time	0.44 vs. 2.35, <i>p</i> = 0.267

A comparison between the groups using Student's *t*-test. Statistically significant differences are marked in **bold**.

A group with (LPR+), according to the literature [14, 17], was identified base on the Ryan Score value – when it was higher than 9.4 for the vertical position or 6.8 for the horizontal position.

In order to establish how such a division of patients is correlated with the other measures of pH-monitoring, these groups were compared according to the observed number of episodes and percentage of reflux duration. The results are presented in Table XI. According to the assumptions, differences in the measures of episodes in the vertical position, which usually characterise LPR [14, 17], are statistically significant. Thus, a coherence of the division to LPR+ and LPR– can be accepted in our data.

A distribution of the results of the RSI and RFS questionnaires divided into LPR+ and LPR– groups was examined, as shown below. The LPR+ group was not statistically significantly different from the LPR– groups in any of the cases, in the results of the RSI and RFS questionnaires (Table XII).

In order to finally verify the usefulness of the RSI and RFS questionnaires for prognosis of the disease, a relation between the LPR+ and LPR– groups and the division into subgroups according to the recommended reference values was examined. It is presented as shown in Tables XIII and XIV. Parameters of this prognosis were similar as in Friedman – they were not useful for the classification of LPR. However, based on RSI, all

**Table XII.** Differences in summary results of the RSI and RFS questionnaires divided into groups according to the occurrence of LPR

Questionnaire	LPR criterion Mean (LPR– vs. LPR+)
RSI sum.	23.23 vs. 23.27, <i>p</i> = 0.983
RFS sum.	7.64 vs. 8.3, <i>p</i> = 0.439

A comparison between the groups using Student's *t*-test.

**Table XIII.** Diagnostic parameters of the standard recommendations concerning the observation of LPR based on the RSI questionnaire

Variable	LPR–	LPR+
RSI ≤ 13	0	4
RSI > 13	22	56
Effectiveness of prediction of the RSI:		
Accuracy	0.6829268	
P-value	0.8681608	
Sensitivity	0.9333333	
Specificity	0.0000000	

The first statement is a distribution of all the studied subjects depending on belonging to the specific groups. The next one sums up the effectiveness of prediction based on the RSI. The p-value corresponds to the statistical significance towards a non-informative value.

**Table XIV.** Diagnostic parameters of the standard recommendations concerning the observation of LPR based on the RFS questionnaire

Variable	LPR–	LPR+
RFS ≤ 7	10	29
RFS > 7	12	31
Effectiveness of prediction of the RFS:		
Accuracy	0.5000000	
P-value	0.9999977	
Sensitivity	0.5166667	
Specificity	0.4545455	

The first statement is a distribution of all the studied subjects depending on belonging to the specific groups. The next one sums up the effectiveness of prediction based on the RFS. The p-value corresponds to statistical significance toward a non-informative value.

the patients with LPR can be detected (high sensitivity), but at the same time, all the people without problems with reflux are wrongly classified (zero specificity). In the case of the criterion based on RFS, in half of the cases the diagnosis is appropriate for the LPR+ and LPR– groups (sensitivity and specificity of approx. 50%). What is more, high p-value of the diagnostic procedures based on RSI and RFS indicates that trivial procedures, such as considering all the subjects as ill, would perform at the same level or even better.

All the other simple schemes of LPR identification on the basis of a summary result of the RFS and RSI questionnaires were analysed, assuming all the possible cut-off points. None of them enabled high accuracy of the prognosis and even high sensitivity and specificity (they were also not statistically significant).

A precise comparison of results of the RSI and RFS questionnaires with the measurements of pH-monitoring, suggesting a problem with reflux, show a low usefulness of these questionnaires for prognosis of the occurrence of LPR. Not only for the reference values recommended in the literature, but also for other procedures that would be applied in practice, the use of these questionnaires does not enable an effective diagnosis for the population studied by the authors.

## Conclusions

As shown in studies by other authors, the RFS and RSI questionnaires can be used for assessing the effectiveness of treating LPR. Their application in diagnostics, especially as a basic tool, raises doubts. More systematic studies in the Polish population are needed to establish the reference values for both the applied questionnaires and the 24-hour pharyngeal pH-monitoring. In case of further studies, performing validation of the RSI and RFS questionnaires is indicated.

## Conflict of interest

The authors declare no conflict of interest.

## References

1. Vaezi MF, Hicks DM, Abelson TI, et al. Laryngeal signs and symptoms and gastroesophageal reflux disease (GERD): a critical assessment of cause and effect association. *Clin Gastroenterol Hepatol* 2003; 1: 333-44.
2. Shaker R, Milbrath M, Ren J, et al. Esophagopharyngeal distribution of refluxed gastric acid in patients with reflux laryngitis. *Gastroenterology* 1995; 109: 1575-82.
3. Belafsky PC, Postma GN, Koufman JA. The validity and reliability of the Reflux Finding Score (RFS). *The Laryngoscope* 2001; 111: 1313-7.
4. Belafsky PC, Postma GN, Koufman JA. Validity and reliability of the Reflux Symptom Index (RSI). *J Voice* 2002; 16: 274-7.
5. Vailati C, Mazzoleni G, Bondi S, et al. Oropharyngeal pH monitoring for laryngopharyngeal reflux: is it a reliable test before therapy? *J Voice* 2013; 27: 84-9.
6. Reichel O, Dressel H, Wiederänders K, et al. Double-blind, placebo-controlled trial with esomeprazole for symptoms and signs associated with laryngopharyngeal reflux. *Otolaryngol Head Neck Surg* 2008; 139: 414-20.
7. Yadlapati R, Pandolfino JE, Lidder AK, et al. Oropharyngeal pH testing does not predict response to proton pump inhibitor therapy in patients with laryngeal symptoms. *Am J Gastroenterol* 2016; 111: 1517-24.
8. Schindler A, Mozzanica F, Ginocchio D, et al. Reliability and clinical validity of the Italian Reflux Symptom Index. *J Voice* 2010; 24: 354-8.
9. Barry DW, Vaezi MF. Laryngopharyngeal reflux: more questions than answers. *Cleve Clin J Med* 2010; 77: 327-34.
10. Maldhure S, Chandrasekharan R, Dutta AK, et al. Role of PH monitoring in laryngopharyngeal reflux patients with voice disorders. *Iran J Otorhinolaryngol* 2016; 28: 377-83.



11. Kelchner LN, Horne J, Lee L, et al. Reliability of speech-language pathologist and otolaryngologist ratings of laryngeal signs of reflux in an asymptomatic population using the reflux finding score. *J Voice* 2007; 21: 92-100.
12. Musser J, Kelchner L, Neils-Strunjas J, et al. A comparison of rating scales used in the diagnosis of extraesophageal reflux. *J Voice* 2011; 25: 293-300.
13. Park KH, Choi SM, Kwon SUK, et al. Diagnosis of laryngopharyngeal reflux among globus patients. *Otolaryngol Head Neck Surg* 2006; 134: 81-5.
14. Friedman M, Hamilton C, Samuelson CG, et al. The value of routine pH monitoring in the diagnosis and treatment of laryngopharyngeal reflux. *Otolaryngol Head Neck Surg* 2012; 146: 952-8.
15. R Core Team. R. A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing 2015.
16. Dunn O. Multiple comparisons among means. *J Am Stat Assoc* 1961; 56: 52-64.
17. Mesallam TA, Malki KH, Farahat M, et al. Voice problems among laryngopharyngeal reflux patients diagnosed with oropharyngeal pH monitoring. *Folia Phoniatr Logop* 2013; 65: 280-7.

**Received:** 26.09.2018

**Accepted:** 14.03.2019