SCIENTIFIC Publishing Limited

ENT Updates

https://ojs.ukscip.com/index.php/ENTU

Article

Improved Scoring of Lund Mackay Score (LMS)

Satyendra Nath Chakrabartty [1,2]

[1] Indian Ports Association, New Delhi 110027, India; chakrabarttysatyendra3139@gmail.com; Tel.: +91-9831597909

[2] Indian Maritime University, Kolkata 700017, India

Received: 18 June 2024; **Revised:** 8 September 2024; **Accepted:** 29 November 2024; **Published:** 6 December 2024

Abstract: Lund Mackay Score (LMS), modified Lund Kennedy (MLK) scores quantify degree of opacification for each sinus by numerical scoring systems. Correlations between LMS on paranasal sinus CT scans, MLK nasoendoscopic scores and associated tools like SNOT-22, RSDI, EPOS, PROMIS-29, etc. showed contrasting empirical results. The paper discusses methodological limitations of numerical scoring systems and provides a method of transforming ordinal discrete scores of a multi-point item to normally distributed proposed scores (P_i). Scale scores $\mathcal{S}$-scores) as sum of P_is also follow normal distribution and can include all indicators irrespective of scale formats. Normality of monotonically increasing P-scores and $\mathcal{S}$-scores of LMS/MLK satisfy desired properties, provide unique ranks to the individuals, facilitate parametric analysis for diagnosis (ROC analysis), classification and comparisons of different aspects of chronic rhinosinusitis measured by LMS/MLK and subjective measures of symptom scores reflecting disease severity. The method also facilitates statistical tests of equality of mean and variance of LMS/MLK for two groups or a single group at different time periods, significance of disease progression and better computation of reliability, factorial validity. Proposed method can better assess severity/disability of CRS and include all tools (pathological, clinical, patient-reported- outcomes and HRQoL instruments) irrespective of scale formats without any bias for advantaged or disadvantaged groups.

Keywords: Chronic Rhinosinusitis; Lund–Mackay CT Score; Sinusitis; SNOT-22; Normal Distribution; Disease Progression

1. Introduction

Computed tomography (CT) scan of the head and neck involves paranasal sinuses in chronic rhinosinusitis (CRS) patients who are asymptomatic. The degree of such opacification is quantified by Lund Mackay Score (LMS), which is a numerical scoring system. For radiographic disease severity, CT scans are scored to get LMS which scores each sinus (anterior ethmoid, posterior ethmoid, maxillary, frontal, and sphenoid sinuses) as: 0 (no opacification), 1 (partial opacification), or 2 (complete opacification). The ostiomeatal complex is scored as 0 (not occluded) or 2 (occluded). Left and right sides are staged separately. The sinuses are grouped into: frontal sinus, anterior ethmoidal cells, posterior ethmoidal cells, maxillary sinus, sphenoid sinus and ostiomeatal complex. The scores are summed and total LMS for a patient ranges between 0 (complete lucency of all sinuses) to 24 (complete opacity of all sinuses), where higher score implies increasing grade of polyposis. The LMS **attempts** to reflect radiographic disease severity for CRS patients **and** has shown good inter-observer reliability and is popular in research and clinical practices [1]. However, **the absence of developed sphenoid** and frontal sinuses **can** reduce the score range to 16 from 24 points, **leading to an** underestimation of the disease and introducing bias, particularly in pediatric applications [2].

To overcome the problem of insufficiency of gradation of LMS, modified Lund Kennedy (MLK) score was proposed where each sinus is scored considering the percentage of opacification from mucosal thickening as follows: 0 = 0%, 1 = 1% to 25%, 2 = 26% to 50%, 3 = 51% to 75%, 4 = 76% to 99%, and 5 = 100% or completely occluded [3]. The ostiomeatal complex is given a score of 0 to 2, depending on whether it is completely patent, partially obstructed, or completely obstructed. Each side is graded, and their sum is the total score out of maximum of 54 [4]. The modified staging systems showed more efficiency in evaluating rhinosinusitis inflammation compared to LMS [5]. While no correlation between changes in LMS on paranasal sinus CT scans and MLK nasoendoscopic scores in CRS patients was found [6], strong positive rank correlation (Spearman ρ) among subjects with polyps and weak rank correlation in subjects without polyps were observed [7]. MLK endoscopic scoring system does not include the items of scarring and crusting. Nasal cavity opacification is not considered as part of total score by LMS or MLK, despite the fact that nasal polyposis causes nasal cavity opacification and characterizes the phenotype of CRS with nasal polyps (CRSwNP).

Empirical investigations of CT scans measured by LMS often involve consideration of other clinical parameters and patient-reported outcome measures (PROMs) to find associations or to validate or both. Popular tools of CRS-specific PROMs are Sino-Nasal Outcome Test (SNOT-22), Rhinosinusitis disability index (RSDI), European Position Statement on Rhinosinusitis (EPOS) [8], Quality of Life Questionnaires, Patient-Reported Outcomes Measurement Information System (PROMIS-29) covering general quality of life across domains like anxiety, depression, fatigue, sleep disturbance, satisfaction with social roles, pain interference, and physical function domains, etc. Such tools generally use ordinal K-point items (K= 2, 3, 4, 5, ...) and differ in scopes, dimensions covered, scale formats, scoring methods, etc. and are often blurred what is being measured and do not always match with clinical and research goals. Similarly, generic and disease-specific Health related quality of life (HRQoL) instruments result in confusion about the best use of an instrument and even popular instruments show different correlations with the dimensions.

Frequently used statistical techniques presume meaningful aggregation (either arithmetic or multiplicative) and set of assumptions. Non-satisfaction of the assumptions and failure to ensure meaningful addition can distort the results. For example, reference [9] found that good number of scientific articles published in the European Annals of Head & Neck Diseases during 2018 and 2019, normality was checked only in 14.2% of articles using , p-value significance threshold were not always defined, confidence intervals and power analyses were documented only in 10.7% and 5.3% respectively. LMS did not follow the normal distribution [10].

The paper describes major methodological weaknesses of numerical scoring scales including LMS and other assessment tools used in otorhinolaryngology and proposes a method to transform the ordinal raw scores of an item to equidistant score (E-scores) followed by further linear transformations to normally distributed continuous scores ranging from 1 to 100, making easier the utilization and the interpretation of the instruments.

2. Literature Survey

LMS correlates well with Percentage Opacification scores, other markers of disease severity, nature of surgery offered, and its outcome. While references [1, 10] found no correlation between LMS and SNOT-22 scores, positive association with pre-operative and post-operative SNOT-22 scores with pre-operative Lund-Mackay CT scores (LMCTS) was observed [11]. LMS was found to be associated with symptom reduction, complication rates and revision rates but, poorly correlated with subjective patient-reported symptom scores. Thus, LMS measures different aspects of chronic rhinosinusitis than the subjective measures of symptom scores reflecting disease severity, often used for comparison [1]. The LMS scoring is not a good measure of chronic rhinosinusitis severity or the prognosis of patients after surgery [10] who preferred SNOT-22 as a predictive tool in deciding to operate and the possibility of obtaining a relative recovery.

2.1. Major Drawbacks of LMS

Non-meaningful Addition in numerical scoring system: Levels of a K-point item are ordered but not equidistant. Trait-distance from "none" to "sometimes" ≠ distance between "always" and "sometimes". Similarly, for a sinus, trait difference between "no opacification" and "partial opacification" could be different from the difference between "partial opacification" and "complete opacification". Subjectivity of the 0-versus-2 score has been questioned

since Ostiomeatal complex (OMC) occlusion is not associated with draining sinuses for patients with eosinophilic rhinosinusitis (ECRS) and simple surgical interventions directed at the OMC **are** not beneficial to this CRS subgroup [12]. Non-satisfaction of equidistant property implies addition is not meaningful. Meaningful addition of $X+Y=Z$ requires similar distribution of X and Y and also knowledge of distribution of Z. Thus, meaningful addition is an important pre-requisite statistical perspective in using such scales. However, PROMs, LMS consider summative scores without checking distribution of scores. Reference [13] emphasized that ordinality, discreteness, nonlinearity, skew, ceiling and floor effects in rating data create problems for undertaking parametric statistical analysis. Normality checks of rating data are necessitated for inference procedures [14].

Possible solution: Transform each item score to normally distributed scores (P_i) and take dimension scores($\mathcal{D}_i$). Scale score (S_i) of i-th subject is sum of P_is so that $\mathcal{D}_i$ and S_i follows normal distribution and parameters of their distribution can be obtained from the data. It is well known that **for two normally distributed variables X and Y, (X + Y) also follows normal distribution.**

Use of zero as an anchor value: Zero attached to a level does not help to find expected values (value of the variable × probability of that value) of level-wise scores; reduces mean and variance of the scale, item-total correlations and regression **could be** be inappropriate due to presence of many zeroes [15]. If each respondent of a sub-group selects the level marked as "0" to an item then it is difficult to find between group variance since mean = variance = 0 for the sub-group and correlation with that item is undefined.

Possible solution: Mark the anchor values as 1, 2, 3..... and so on, keeping the convention of higher score ⇔ higher value of the variable being measured.

Non-satisfaction of normality assumptions: Usual procedure to test equality of mean score of two groups is through t-test or paired t-test, and techniques like ANOVA, F-test, Principal component analysis (PCA), Factor analysis (FA), etc. assume normally distributed score. **Thus**, empirical checking of normality before adopting such techniques **is needed**. Problems arise if test of normality fails.

Possible solution: Transformation of item scores to normally distributed scores confirms normality and no need to undertake test of normality.

Factoring cavity size: LMS does not factor cavity size into its scoring system [16]. Accordingly, the authors computed percentage sinus opacification (%SO) considering averaged sinus volume and not an overall total opacification percentage. However, computerized volumetric analysis depending on manual image segmentation is difficult, time-consuming and has limitations in practical applicability [17] who suggested 0has limitations in practical applicability [17] who suggested use of convolutional neural network (CNN) for objective, volumetric segmentation of the paranasal sinuses on CT.

Multiple linear regressions: Used to find empirical relationship of a dependent variable (Y) and associated factors of ENT disorders as independent variables X_1, X_2, ……, X_m. Major assumptions of multiple linear regressions are: linearity; normal distribution with zero mean and constant variance (homoscedasticity) for errors in prediction (residuals), and no pair of independent variables are highly correlated (absence of multicollinearity). However, high value of correlation coefficient (r) or coefficient of determination (R^2) **does** not **always** justify linearity. If the variable X takes integers 1, 2, 3……30, $r_{X,f(X)} \geq 0.92$ for $f(X) = X^2$, X^3, log_{10}^{X}, and $Sin\ X$ and $f(X) = \alpha + \beta X$ was not justified since error scores did not follow normal distribution [18]. Moreover, observed value of correlation depends heavily on group heterogeneity. Let $X \sim N(0, 1)$ and $Y = \frac{1}{\sqrt{2\pi}} e^{\frac{-1}{2}X^2}$. Here, $r_{XY} = -0.93302$ for $0 \leq X \leq 3.9$ and $r_{XY} = 0.0004$ for $-3.9 \leq X \leq 3.9$ [18], indicating homogeneity of data can underestimate or overestimate the correlation. Thus, regression analysis based on value of (r) or R^2 only, are not **always** justified.

Diagnosis: It distinguishes Normals with others. In addition to positive objective endoscopic or CT findings, diagnosis of CRS requires that the patients have at least two of the following four symptoms: nasal congestion, nasal drainage, facial pain/pressure, and/or diminished smell. For diagnosis of CRS using CT scan, reference [19] found high area (0.802) under the receiver operating characteristic (ROC) curve and high sensitivity (about 85%) and specificity scores (59%). High sensitivity of the CT scan for mucosal inflammation in paranasal sinuses indicates possibility of "incidental" mucosal findings that do not represent true "sinus" disease [20] and may lead to over diagnosis. Surprisingly, children without CRS exhibited a rather low incidental Lund score of approximately 2.8 [21] which is less common than adult patients without CRS [22]. However, diagnostic efficiency depends heavily on the base rate prevalence of CRS in the population being evaluated.

Possible solution: Combine normally distributed scale scores of LMS or MLK and symptoms scores satisfying basic assumption of parametric ROC curve. Use such transformed scores to undertake ROC analysis.

Classification: Researchers differed in deciding boundary points for classification of subjects based on LMS. As per reference [22], LMS ≤ 3 are normal, 4–5 are indeterminate and ≥6 are pathological. Typically, LMS ≤ 3 require additional clinical judgment and/or further data to establish diagnoses [6]. However, classifications suggested by reference [23, 24] were different primarily due to different sample sizes and heterogeneity of the samples with respect to chronic sinusitis [25]. However, reference [26] suggested LMS < 8 as low and LMS ≥ 8 as high which reduced differences in scoring and increased inter-observer reliability and grouped the sample with LMS up to 8, 8 to 16, and greater than 16.

Possible solution: Find optimal cut-off values and measure of efficiency of classification.

Reliability: Reliability of LMS is often computed by inter-rater reliability (IRR) or test-retest reliability ($r_{test-retest}$). IRR reflecting agreement between experts can be obtained by methods like percentages of agreement (PA), kappa (κ), weighted kappa, resulting in different values and conflicting results of IRR [27]. Kappa treats all the disagreements equally and ignores marginal distributions. Weighted kappa depends heavily on the chosen of weights which are subjective. Value of weighted kappa usually exceeds value of unweighted kappa. LMS performed by independent scorers vary. Differences exceeding 1 are resolved by consensus, failing which a third party decides the score [16]. However, agreement among raters and scale reliability are different concepts [28]. Considering different inter-rater agreement (IRA) across different levels, reference [29] proposed Coefficient of variation (CV) as a measure of agreement among raters.

High value of $r_{test-retest}$ **does** not imply robustness of LMS. $r_{test-retest}$ may be high if there is no effect of treatments or scores of each subject improved or deteriorated uniformly due to treatments. Such reliability **does** not reflect true stability of the construct(s), which **is** also influenced by time gap. Reference [30] preferred correlation (not agreement) to compute $r_{test-retest}$ of Internet Addiction Test by reference [31]. Clearly, $r_{test-retest}$ is not be a sufficient condition to demonstrate agreements.

It is desirable to find reliability reflecting both correlation and agreement between measurements. Intraclass correlation coefficient (ICC) satisfies such desirable properties. However, various forms of ICCs are there which differ in assumptions leading to different values and interpretations. The "absolute agreement" approach usually results in smaller ICC estimate than the "consistency" approach. The MLK score showed ICC coefficient =0.68 implying good inter-rater reliability [32].

Validity: Concurrent validity of LMS or MLK was evaluated by comparing the scores with retest scores at 6 months after surgery [33]. However, this could be a measure of $r_{test-retest}$ or responsiveness. The authors found discriminating validity by comparing the symptomless and symptomatic sides by Wilcoxon signed-rank test.

Others: Use of different analysis gave different results, For example, reference [1] used usual Pearson's correlation coefficient to fit regression equation of LMS on SNOT-22 scores, despite weak correlation ($r_{Pre-operative\ SNOT-22,\ LMS}$) at the level of 0.058; one-way ANOVA to see association between LMS and ordinal variables like extent of polyposis and logistic regression to see power of LMS to predict perioperative complications or revision surgery.

For a given sinus cavity, LMS showed poor ability to discriminate among patients with mild-to-moderate levels of disease [16]. LMS should not be used as the sole basis to proceed with surgery, especially for those with higher likelihood of undergoing endoscopic sinus surgery (ESS) [34].

3. Method

Transform ordinal raw scores of say 5-point item (i-th item) to continuous, monotonic equidistant scores (E_i-scores) by taking weights W_{i1}, W_{i2}, W_{i3}, W_{i4}, W_{i5} considering frequency of the levels of the item such that $5W_{i5} - 4W_{i4} = 4W_{i4} - 3W_{i3} = 3W_{i3} - 2W_{i2} = 2W_{i2} - W_{i1}$ = Constant. Calculations of weights are based on following stages:

I. Find maximum ($f_{i\,max}$) and minimum frequency ($f_{i\,min}$) of the response-categories of the i-th item. Take initial weights $\omega_{ij} = \frac{f_{ij}}{n}$. Arrange the $\omega_{ij}^{'}s$ so that

$$\omega_{i1} = \frac{f_{imin}}{n} < \omega_{i2} < \omega_{i3} < \omega_{i4} < \omega_{i5} = \frac{f_{imax}}{n}$$

II. Let intermediate weight $W_{i1} = \omega_{i1}$
Take common difference $\alpha = \frac{5f_{imax} - f_{i\,min}}{4n}$ since $W_{i1} + 4\alpha = 5W_{i5}$
Define intermediate weights of the *j*-th level of the item as $W_{ij} = \frac{1}{j}[\omega_{i1} + (j-1)\alpha]$
for *j*= 2, 3, 4, 5
III. Consider the final weights $W_{ij(Final)} = \frac{W_{ij}}{\sum_{j=1}^{5} W_j}$ so that $\sum W_{ij(Final)} = 1$ and
$kW_{ik(Final)} - (k-1)W_{i(k-1)(Final)}$ = Constant, value of which differs for different items.
E_i-scores are standardized as $Z_i = \frac{E_i - \overline{E_i}}{SD(E_i)} \sim N(0,1)$ and $Z_i s$ are transformed to proposed score

$$P_i = (100-1)\left[\frac{Z_i - MinZ_i}{MaxZ_i - MinZ_i}\right] + 1 \tag{1}$$

where $1 \leq P_i \leq 100$ ensures uniformity in item score–range.
Dimension scores ($\mathcal{D}_i$) is the sum of normally distributed scores of items (P_i) belonging to a dimension. Scale score ($\mathcal{S}_i$) of *i*-th subject is the sum of his/her dimension scores = sum of all item-wise P_i-scores. $\mathcal{S}_i$ and $\mathcal{D}_i$ will follow normal. For example, if scores of the *i*-th item $\sim N(\mu_i, \sigma_i)$, $\mathcal{S}$-scores (transformed LMS or transformed MLK) follows normal with mean $\sum_i \mu_i$ and variance $[\sum \sigma_i^2 + 2\sum_{i \neq j} Cov(P_i, P_j]$. Thus, probability density function of $\mathcal{S}$ is convolution of normally distributed P_i-scores. Parameters of normal distribution of $\mathcal{S}$ can be estimated from the data.

4. Results

Continuous and monotonically increasing $\mathcal{S}$-scores following normal are based on pattern of responses, enabling unique ranks to the individuals. In addition, $\mathcal{S}$-scores satisfy following desired properties:

- Same range of scores for each P_i
- Avoid bias for advantaged or disadvantaged groups
- For *i*-th dimension, contribution to $\mathcal{S}$ and elasticity are quantified respectively by $\frac{\mathcal{D}_i}{\mathcal{S}} \times 100$ and $\frac{\frac{\Delta \mathcal{S}}{\mathcal{S}}}{\frac{\Delta \mathcal{D}_i}{\mathcal{D}_i}}$ to show relative importance of the dimensions from two different angles.

Benefits

- Provides total score of an individual for LMS or MLK or any associated scale like SNOT-22, RSDI, EPOS, PROMIS-29, Quality of Life Questionnaires, etc., irrespective of scale formats and factor structures.
- Progress of the *i*-th patient in *t*-th time-period over the previous time-period by $\frac{\mathcal{S}_{i_t} - \mathcal{S}_{i_{(t-1)}}}{\mathcal{S}_{i_{(t-1)}}} * 100$ which reflects responsiveness of $\mathcal{S}$-scores and effectiveness of treatment plan adopted. $\mathcal{S}_{i_t} < \mathcal{S}_{i_{(t-1)}} \Longrightarrow$ Decline in *t*-th period over $(t-1)$-th period. Such deterioration needs to be probed for identification of the dimension(s) where deteriorations occurred and initiation of possible corrective actions. For a sample of patients, progress is indicated by $\overline{\mathcal{S}_{i_t}} > \overline{\mathcal{S}_{i_{(t-1)}}}$.

Disease progression across time (progress path) CONFIRMED. PLEASE DELETE of one patient or a sample of patients can be plotted over time to facilitate comparisons of response to the treatments from the beginning. A decreasing (negatively slopped) graph of $\mathcal{S}_{i_t}$ and time *(t)* imply improvement across time and an increasing graph indicates the opposite. Such plot is akin to hazard function of survival.

- Responsiveness of $\mathcal{S}$-scores enables practitioners or researcher to know time-to-event outcomes from the time of diagnosis to the occurrence of the relevant events (disease recurrence or progress/deterioration) by a continuous variable.
- Linear association between $\mathcal{S}$-scores and HRQoL-scores can be evaluated by correlation or by multiple correlations between $\mathcal{S}$-scores on HRQoL dimension scores as independent variables.
  Univariate regression equation of the form $\mathcal{S} = \alpha_1 + \beta_1(HRQoL\text{-}scores)$ or multiple linear regression $\mathcal{S} = \alpha + \sum_{i=1}^{k} \beta_i HRQoL_{Dimenision_i}$ can be fitted using dimensions scores of HRQoL as predictors of LMS/MLK. Regres-

sion equation of the form $HRQoL = \alpha + \beta.\mathcal{S}$ helps to know effect of LMS/MLK on HRQoL. However, checking of assumptions of the regression equations are needed while going for such analysis.

- Facilitates statistical tests of equality of mean and variance of LMS/MLK for two groups or a single group at different time periods like $H_0 : \mu_1 = \mu_2$ or $H_0 : \sigma_1^2 = \sigma_2^2$. Statistical tests of significance of progress of LMS/MLK can be tested by $H_0 : \frac{\mathcal{S}_{i_t} - \mathcal{S}_{i_{(t-1)}}}{\mathcal{S}_{i_{(t-1)}}} = 0$ since ratio of two normally distributed variables follows χ^2-distribution.
- Normally distributed $\mathcal{S}$-scores of LMS or MLK and symptoms scores satisfy the basic assumption of parametric ROC curve. Such transformed scores facilitate undertaking of ROC analysis where for a threshold parameter T, the score is "positive" if X > T, and "negative" if X < T where X follows known probability distribution with density $f_1(x)$ if X > T and density $f_0(x)$ if X < T. The true positive rate (Sensitivity i.e., probability of detection) is given by TPR(T) = $\int_T^\infty f_1(x)dx$ and the false positive rate (1-Specificity i.e., probability of false alarm) FPR(T) = $\int_T^\infty f_0(x)dx$. The ROC curve plots TPR (T) versus FPR(T) with T as the varying parameter. In other words, if distribution of diseased X_1 and non-diseased X_2 are $N(\mu_1, \sigma_1)$ and $N(\mu_2, \sigma_2)$ respectively, then the ROC curve is $ROC(t) = \Phi(a + b\Phi^{-1}(t))$ where Φ denotes the cumulative normal distribution function, $a = \frac{\mu_1 - \mu_2}{\sigma_1}, b = \frac{\sigma_2}{\sigma_1}$, and $AUC = \emptyset(\frac{a}{\sqrt{1+b^2}})$ where $\emptyset$ is the normal probability density function [35] and AUC = $\int_0^1 ROC(u)du$. For a single diagnostic test, the hypothesis H_0: $AUC = AUC_0$ can be tested by $Z = \frac{\hat{AUC} - AUC_0}{SE(\hat{AUC})}$ where $\hat{AUC}$ and its standard error (SE) can be estimated by parametric approaches.
- Helps to assess efficiency of classification by Davies-Bouldin Index (DBI) considering ratio of within-cluster and between-cluster distances [36]. For K-number of classes DBI is computed by $DBI_K = \frac{1}{K}\sum_{i=1}^{K} \sum_{j=1(i \neq j)}^{K} Max[\frac{DiamC_i - DiamC_j}{||C_i - C_j||}]$ where diameter of *i*-th class $DiamC_i = \sqrt{\frac{\sum_{x_i \in C_i} ||x_i - C_i||^2}{n_i}}$; C_i: Centroid or mean of the *i*-th class; n_i: Number of members in the *i*-th class. Upper limit of DBI is 1 and lower value implies better efficiency. The optimal number of clusters obtained from the graph of DBI and number of clusters has the smallest DBI value [37].
- Normality of *P*-scores and $\mathcal{S}$-scores of LMS/MLK enable undertaking PCA and finding factorial validity as the ratio of the highest eigenvalue (λ_1) and sum of the eigenvalues i.e.,

$$\text{FV} = \frac{\lambda_1}{\sum \lambda_i} \tag{2}$$

FV indicates validity of the main factor for which the scale was developed [38] and avoids selection of criterion scale with different score distributions, different factor structures and different domains of one or more constructs etc. For a unidimensional tests, FV is high.

-Maximum value of test reliability (α_{PCA}) of a scale with *m*-number of items was derived by reference [39] and is given by

$$\alpha_{PCA} = \left(\frac{m}{m-1}\right)\left(1 - \frac{1}{\lambda_1}\right) \tag{3}$$

Non-linear Relationship between FV and α_{PCA} [40] is:

$$\alpha_{PCA} = \left(\frac{m}{m-1}\right)\left(1 - \frac{1}{\lambda_1}\right) = \left(\frac{m}{m-1}\right)\left(1 - \frac{1}{FV.\sum \lambda_i}\right) = \left(\frac{m}{m-1}\right)\left(1 - \frac{1}{m.FV_{Z-scores}}\right) \tag{4}$$

Clearly higher value of $FV_{Z-scores}$ increases α_{PCA}.

5. Discussion

The paper addresses methodological issues of tools measuring LMS, MLK and associated tools like SNOT-22, RSDI, EPOS, PROMIS-29, Quality of Life Questionnaires, etc. and proposes conversion of ordinal scores of each scale to normally distributed scores. **Normality helps** meaningful evaluation of measurement properties and better utilization of such tests. Normality of *P*-scores and $\mathcal{S}$-scores of LMS/MLK satisfy desired properties, facilitate meaningful aggregation, better comparisons and rankings, offer platform for parametric analysis. **Such scoring method facilitates** statistical testing, fitting regression equations of LMS on HRQoL or HRQoL on LMS. $\mathcal{S}$-scores also helps to undertake parametric ROC analysis for diagnosis and better measure of reliability, factorial validity

avoiding criterion variable, assessment of progress or deterioration, identification of critical dimension(s), etc. **Relationship of maximum value of test reliability (α_{PCA}) and factorial validity can help to decide desirable value of α_{PCA} which maximizes factorial validity. The** proposed method can better assess **the severity and disability** of CRS and **incorporate** all **relevant** tools (pathological, clinical, patient-reported- outcomes and HRQoL instruments) irrespective of scale formats without any bias for advantaged or disadvantaged groups.

6. Conclusions

Normally distributed scores for LMS, MLK and HRQoL-score avoiding limitations of the existing methods can compare different aspects of chronic rhinosinusitis measured by LMS and subjective measures of symptom scores reflecting disease severity. **Major advantages of the proposed method are: parametric ROC analysis for diagnosis, classification, and** plotting the **disease progression** path of CRS, which in turn provides another criterion for comparisons. **In addition, normality helps to improve psychometric measures and their relationship.**

The proposed method satisfying desired properties advances scholarly. Practitioners and researchers can take advantages of the proposed **scoring system of Rhinologic Investigations** for meaningful analysis, **better assessment of severity/disability of CRS without any bias for advantaged or disadvantaged groups.** Empirical verifications of robustness of the proposed method, estimation of hazard function and clinical validations are proposed as future studies, along with comparisons with AI technology based sinus CT evaluation for quantification of radiologic disease burden in both clinical and research applications.

Funding

This work received no external funding.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Data Availability Statement

No datasets were generated or analyzed in the study.

Conflicts of Interest

The authors declare no conflict of interest CONFIRMED.

References

1. Hopkins C.; Browne J.P.; Slack R.; et al. The Lund-Mackay staging system for chronic rhinosinusitis: How is it used and what does it predict? *Otolaryngol. Head Neck Surg.* **2007**, *137*, 555–561.
2. de Araújo Neto, S.A.; Baracat, E.C.E.; Felipe, L.F. A new score for tomographic opacificacation of paranasal sinuses in children. *Braz. J. Otorhinolaryngol.* **2010**, *76*, 491–498. [CrossRef]
3. Okushi, T.; Nakayama, T.; Morimoto, S.; et al. A modified Lund–Mackay system for radiological evaluation of chronic rhinosinusitis. *Auris Nasus Larynx* **2013**, *40*, 548–553. [CrossRef]
4. Likness, M.; Pallanch, J.; Sherris, D.; et al. Computed Tomography Scans as an Objective Measure of Disease Severity in Chronic Rhinosinusitis. *Otolaryngol. Head Neck Surg.* **2014**, *150*, 305–11. [CrossRef]
5. Zinreich, S.J. Imaging for staging of rhinosinusitis. *Ann. Otol. Rhinol. Laryngol. Suppl.* **2004**, *113*, 19–23. [CrossRef]
6. Frączek, M.; Masalski, M.; Guziński, M. Reliability of computed tomography scans in the diagnosis of chronic rhinosinusitis. *Adv. Clin. Exp. Med.* **2018**, *27*, 541–545. [CrossRef]
7. Sudiro, M.; Kuntara, A.; Waldi.; D. Correlation of Lund-Mackay Score on Computed Tomography Scan and Nasoendoscopic Score in Chronic Rhinosinusitis. *Acta Inform. Med.* **2023**, *31*, 53–56. [CrossRef]
8. Fokkens, W.J.; Lund, V.J.; Mullol, J.; et al. European position paper on rhinosinusitis and nasal polyps. A sum-

mary for otorhinolaryngologists. *Rhinology* **2012**, *50*, 1–12. [CrossRef]
9. Laccourreye, O.; Fakhry, N.; Franco-Vidal, V.; et al. Statistics in scientific articles published in the European Annals of Otorhinolaryngology Head & Neck Diseases. *Eur. Ann. Otorhinolaryngol. Head Neck Dis.* **2021**, *138*, 89–92. [CrossRef]
10. Taheri, A.; Hasani, S.; Esfeedvajani, M.S.; et al. Sino-nasal outcome Test-22 and Lund–Mackay CT score to select endoscopic sinus surgery in chronic rhinosinusitis. *World J. Otorhinolaryngol. Head Neck Surg.* **2023**, *10*, 200–205. [CrossRef]
11. Brooks, S.G.; Trope, M.; Blasetti, M.; et al. Preoperative Lund-Mackay computed tomography score is associated with preoperative symptom severity and predicts quality-of-life outcome trajectories after sinus surgery. *Int. Forum Allergy Rhinol.* **2018**, *8*, 668–675. [CrossRef]
12. Snidvongs, K.; Chin, D.; Sacks, R.; et al. Eosinophilic rhinosinusitis is not a disease of ostiomeatal occlusion. *Laryngoscope* **2013**, *123*, 1070–1074. [CrossRef]
13. Šimkovic, M.; Träuble, B. Robustness of statistical methods when measure is affected by ceiling and/or floor effect. *PLoS ONE* **2019**, *14*, e0220889.
14. Clason, D.L.; Dormody, T.J. Analyzing Data Measured By Individual Likert-Type Items. *J. Agric. Educ.* **1994**, *35*, 31–35. [CrossRef]
15. Goulet, J.; Buta, E.; Carroll, C.; et al. Statistical Methods for the Analysis of NRS Pain Data. *J. Pain* **2015**, *16*, S7. [CrossRef]
16. Massey, C.J.; Ramos, L.; Beswick, D.M.; et al. Clinical Validation and Extension of an Automated, Deep Learning-Based Algorithm for Quantitative Sinus CT Analysis. *Am. J. Neuroradiol.* **2022**, *43*, 1318–1324. [CrossRef]
17. Humphries, S.M.; Centeno, J.P.; Notary, A.M.; et al. Volumetric assessment of paranasal sinus opacification on computed tomography can be automated using a convolutional neural network. *Int. Forum Allergy Rhinol.* **2020**, *10*, 1218–1225. [CrossRef]
18. Chakrabartty, S.N. Improving linearity in health science Investigations. *Health Sci. J.* **2023**, *17*, 1–6.
19. Bhattacharyya, N.; Fried, M.P. The accuracy of computed tomography in the diagnosis of chronic rhinosinusitis. *Laryngoscope* **2003**, *113*, 125–129. [CrossRef]
20. Bhattacharyya, N. Do maxillary sinus retention cysts reflect obstructive sinus phenomena? *Arch Otolaryngol. Head Neck Surg.* **2000**, *126*, 1369–1371. [CrossRef]
21. Hill, M.; Bhattacharyya, N.; Hall, T.R.; et al. Incidental paranasal sinus imaging abnormalities and the normal Lund score in children. *Otolaryngol. Head Neck Surg.* **2004**, *130*, 171–175. [CrossRef]
22. Ashraf, N.; Bhattacharyya, N. Determination of the incidental Lund score for the staging of chronic rhinosinusitis. *Otolaryngol. Head Neck Surg.* **2001**, *125*, 483–486. [CrossRef]
23. Do, B.A.; Lands, L.C.; Mascarella, M.A.; et al. Lund–Mackay and modified Lund–Mackay Score for sinus surgery in children with cystic fibrosis. *Int. J. Pediatr. Otolaryngol.* **2015**, *79*, 1341–1345.
24. Rathor, A.; Bhattacharjee, A. Clinical-radiological correlation and role of computed tomography staging in chronic rhinosinusitis. *World J. Otolaryngol. Head Neck Surg.* **2017**, *3*, 169–175.
25. Emmanuel, I.O.; Festus, E.O. Lund-Mackay Scoring of Incidental Paranasal Sinus Collection on Computed Tomography Scan of Head and Neck in the University of Benin Teaching Hospital, Nigeria. *Borno Med. J.* **2018**, *15*, 59–64.
26. John Staniorski, C.; Price, C.P.; Weibman, A.R.; et al. Asthma onset pattern and patient outcomes in a chronic rhinosinusitis population. *Int. Forum Allergy Rhinol.* **2018**, *8*, 495–503. [CrossRef]
27. Quinn, T.J.; Dawson, J.; Walters, M.R.; et al. Reliability of the modified Rankin Scale: a systematic review. *Stroke* **2009**, *40*, 3393–3395.
28. Wilson, J.T.; Hareendran, A.; Hendry, A.; et al. Reliability of the modified Rankin Scale across multiple raters: benefits of a structured interview. *Stroke* **2005**, *36*, 777–781.
29. Kottner, J.; Streiner, D.L. The difference between reliability and agreement. *J. Clin. Epidemiol.* **2011**, *64*, 701–702.
30. Jelenchick, L.A.; Becker, T.; Moreno, M.A. Assessing the psychometric properties of the Internet Addiction Test (IAT) in US college students. *Psychiatry Res.* **2012**, *196*, 296–301.
31. Young, K.S. Caught in the Net: How to Recognize the Signs of Internet Addiction—And a Winning Strategy for Recovery. John Wiley & Sons: New York, NY, USA, 1998.
32. Koo, T.K.; Li, M.Y. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J. Chiropr. Med.* **2016**, *15*, 155–163. [CrossRef]
33. Zhang, L.; Zhang, L.H. Comparison of different endoscopic scoring systems in patients with chronic rhinosinusitis: reliability, validity, responsiveness and correlation. *Rhinology* **2017**, *55*, 363–368.

34. Mehta, M.P.; Hur, K.; Price, C.P.; et al. Radiographic disease severity in chronic rhinosinusitis patients and health care utilization. *Laryngoscope Investig. Otolaryngol.* **2021**, *6*, 924–931. [CrossRef]
35. McClish, D.K. Analyzing a portion of the ROC curve. *Med. Decis. Mak.* **1989**, *9*, 190–195. [CrossRef]
36. Davies, D.L.; Bouldin, D.W. A cluster separation measure. *IEEE Trans. Pattern Anal. Mach. Intell.* **1979**, *1*, 224–227.
37. Kim, M.; Ramakrishna, R. New indices for cluster validity assessment. *Pattern Recognit. Lett.* **2005**, *26*, 2353–2363.
38. Parkerson, H.A.; Noel, M.; Pagé, M.G.; et al. Factorial Validity of the English-Language Version of the Pain Catastrophizing Scale–Child Version. *J. Pain* **2013**, *14*, 1383–1389. [CrossRef]
39. ten Berge, J.M.F.; Hofstee, W.K. Coefficients alpha and reliabilities of unrotated and rotated components. *Psychometrika* **1999**, *64*, 83–90. [CrossRef]
40. Chakrabartty, S.N. Sound Measurement of Patients Reported Outcomes. *New Healthcare Advancements and Explorations (NHAE): Cheyenne, WY, USA.* 2024; Volume 1. [CrossRef]