Article published online: 2025-01-09

Statistics Primer for Radiologists: Part 2—Advanced statistics for Enhancing Diagnostic Precision and Research Validity

Adarsh Anil Kumar[1] Santhosh Kannath[1] Jineesh Valakkada[1]

[1] Department of Imaging Sciences and Interventional Radiology, Sree Chitra Institute of Medical Sciences, Trivandrum, Kerala, India

Indian J Radiol Imaging 2025;35(Suppl S1):S74–S92.

Address for correspondence Santhosh Kannath, PDCC, Department of Imaging Sciences and Interventional Radiology, Sree Chitra Institute of Medical Sciences, Trivandrum 695011, Kerala, India (e-mail: santhoshkannath@sctimst.ac.in).

Abstract

Second part of this statistics primer focuses on advanced statistical concepts continuing on the foundation of basic statistics built from the first part of this primer. This advanced primer aims to delve deeper into essential statistical concepts beyond the basics, equipping the reader with the knowledge to effectively analyze complex data sets, explore correlations and causality, employ regression analysis techniques, interpret survival curves, and evaluate diagnostic tests rigorously. It primarily focuses on the statistical tests used to analyze the relationship between groups of variables (the statistical tests to analyze the difference between groups of variables was discussed in the part 1 of this series). Toward the end of the article concepts of survival curves and methods for assessing the diagnostic accuracy of tests are stressed upon.

Keywords
- primer
- radiologists
- statistics

Introduction

In the realm of radiology, where precision and accuracy are paramount, statistics serves as a crucial tool for interpreting data, conducting research, and making informed clinical decisions. This advanced primer aims to delve deeper into essential statistical concepts beyond the basics, equipping the reader with the knowledge to effectively analyze complex data sets, explore correlations and causality, employ regression analysis techniques, interpret survival curves, and evaluate diagnostic accuracy tests rigorously.

Statistical tests used to test the relationship between variables are broadly divided into three categories:

1. Correlation analysis.
2. Regression analysis (prediction).
3. Time-dependent statistical analysis (Cox regression analysis).

The first section of this article focuses on the concept of correlation between two data sets and the degree of correlation (i.e., regression). Toward the next half of this article, the focus will primarily be on the diagnostic accuracy of tests as well as interobserver agreement. With this background, let us now look into the concepts of correlation and regression.

Correlation and Causality

Correlation

Correlation is a statistical technique that measures the strength and direction of the linear relationship between two variables. While correlation analysis is useful in identifying associations, it does not equate to causation.[1]

Over the last few decades, there has been a notable increase in the number of thyroid cancer cases diagnosed worldwide. During nearly the same period, there has also been a significant increase in the use of computed tomography (CT) scans. We might infer a causal relationship from this assuming that the radiation from these CT scans cause thyroid cancer. However, a more plausible explanation is that the increased use of CT has led to the detection of more

DOI https://doi.org/10.1055/s-0044-1800971.
ISSN 0971-3026.

Thieme Medical and Scientific Publishers Pvt. Ltd., A-12, 2nd Floor, Sector 2, Noida-201301 UP, India

incidental thyroid nodules. Many of these nodules are small, indolent tumors that might never have been discovered otherwise. This correlation does not imply that increased use of CT is causing the thyroid cancer; rather, CT is now detecting more cases that were previously undetected.

Correlation analysis can be used to assess the relationship between variables. The strength of correlation is determined by the correlation coefficient (ranges from −1 to +1). From correlation analysis, inferences can be made about the strength and direction of correlation. Direction indicates whether the correlation is positive or negative, and strength indicates whether the correlation is strong or weak.[1–3]

Positive correlation means that larger values of variable *x* are accompanied by larger values of variable *y*. Basically, positive correlation means relationship between two variables that tend to move in the same direction (either increasing or decreasing). In certain types of tumors, such as those with high vascularity, as the tumor size increases, the magnetic resonance imaging (MRI) signal intensity may also increase due to greater blood flow. *Negative correlation* means that larger values of variable *x* are accompanied by smaller values of variable *y*, and vice versa. In oncology imaging, a negative correlation can be seen between tumor response to effective treatment and tumor size. As the treatment progresses and is effective, the size of the tumor typically decreases, which can be monitored through serial imaging studies such as MRI or CT. Graphical representation of correlation between two variables is through a *scatter plot* (►**Fig. 1**). Strength of correlation coefficient (r) is as follows:

- $r = 0.0$–No correlation.
- $0 < r < 0.3$ or $0 < r < -0.3$–Weak correlation.
- $0.3 < r < 0.7$ or $-0.3 < r < -0.7$–Moderate correlation.
- $0.7 < r < 1.0$ or $-0.7 < r < -1.0$–Strong correlation.
- $r = 1.0$–Perfect correlation (+1–perfect positive correlation; −1–perfect negative correlation).

As the value of correlation coefficient is near to 1 on either side, it shows that there is a strong relationship between the variables. If there is a correlation in the sample, then it is necessary to test whether there is sufficient evidence to suggest a correlation in the population. Significance of correlation coefficients can be tested using a *t*-test. In this case, the null hypothesis is that the two variables under consideration have no correlation. If the calculated *p*-value is < 0.05, the null hypothesis is rejected and it is assumed that there is a correlation between the variables.

Correlation hypothesis can be directional or nondirectional. In nondirectional correlation hypothesis the researcher is interested only in identifying whether there is a correlation between the variables and not interested in the direction of the correlation. However, in directional correlation hypothesis the researcher is interested in the direction of the correlation as well (i.e., whether there is a positive or negative correlation between the variables).

Pearson's Correlation

Pearson's correlation assesses the degree of linear relationship between metric scaled variables. It is basically a parametric measure (on variables having a Gaussian distribution). It is the covariance that is used for calculation. If the covariance has a positive value it indicates a positive correlation and if the covariance has a negative value then it indicates a negative correlation. Covariance can assume values in between plus and minus infinity. This makes it difficult to calculate the strength of relationship between different variables. It is for this reason that correlation coefficient (also called product-moment correlation coefficient) is calculated. This correlation coefficient is obtained by normalizing the covariance. For this normalization, variances of both variables are used and correlation coefficient is calculated.[4–6]

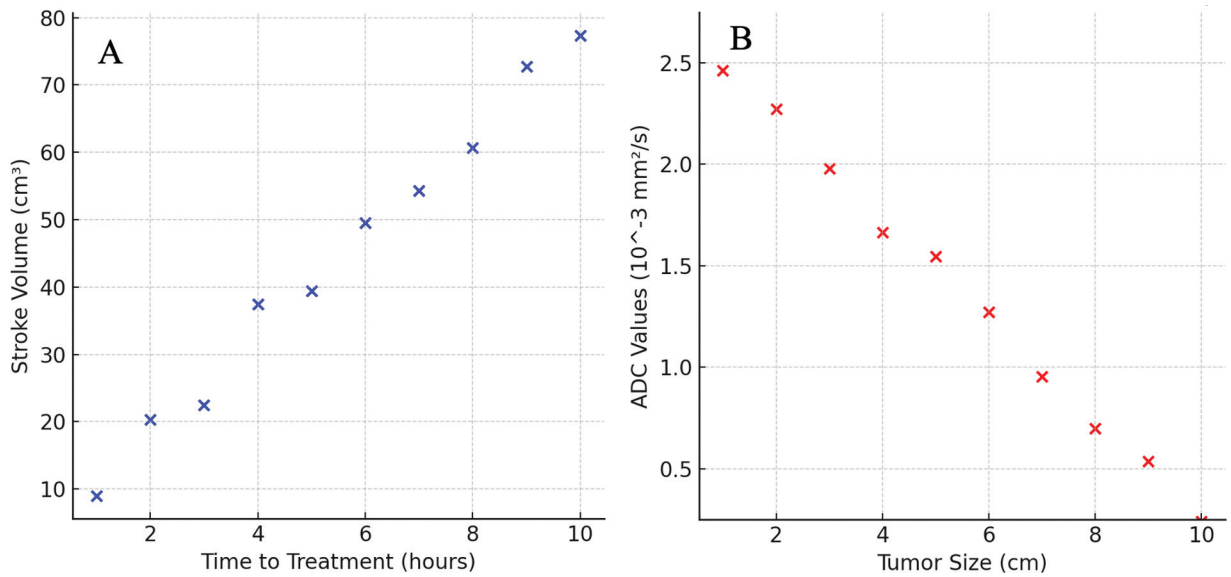

Fig. 1 Scatter plot (**A**) demonstrating positive correlation between the final infarct volume in acute stroke and the delay in time to revascularization. Scatter plot (**B**) demonstrating negative correlation between tumor size in case of brain tumors and the lower apparent diffusion coefficient (ADC) values on diffusion-weighted imaging.

Pearson's correlation coefficient (r) can take values anywhere between –1 and +1 and is interpreted as follows:

- Value of +1 indicates that there is a positive linear relationship.
- Value of –1 indicates that there is a negative linear relationship.
- Value of 0 indicates there is no linear relationship (i.e., the variables do not correlate with each other).
- Scatter plot is used to assess whether a linear relationship exists. It can be used to visually represent the relationship between variables. Pearson's correlation is only useful if linear relationships are present.

Variables must be normally distributed and must have a linear relationship between them for Pearson's correlation to be used. Normal distribution can be tested either analytically or graphically with the Q-Q plot. Whether the variables have a linear correlation is assessed by a scatter plot. If these conditions are not satisfied, then Spearman's correlation is used.

Example: Pearson's correlation coefficient can be used to assess the linear relationship between pulmonary nodule size on CT thorax scans and the likelihood of malignancy. A positive correlation (e.g., $r = 0.8$) would indicate that larger nodules tend to have a higher likelihood of being malignant, while a negative correlation would imply the opposite (though this is less expected in this context). A near-zero value would suggest no linear association between nodule size and malignancy likelihood. This analysis assumes both variables are continuous, normally distributed, and have a linear relationship, though other correlation methods might be more appropriate if these assumptions are not met.

Spearman's Rank Correlation

Spearman's rank correlation assesses the relationship between two variables that have ordinal level of measurement. It is the nonparametric equivalent of Pearson's correlation analysis. This correlation is used when the prerequisites for a parametric correlation analysis are not met, that is, when there is no metric data and no normal distribution. It is also known as Spearman's rho.[4–6]

Rank correlation calculation is based on the ranking system of the data series. Measured values are not used for the calculation, but instead they are transformed into ranks. Spearman's rank correlation is then performed using these ranks. For the rank correlation coefficient ρ, values can range between –1 and 1. If there is a value less than zero ($\rho < 0$), it indicates a negative linear correlation. If the value is greater than zero ($\rho > 0$), there is a positive linear relationship. If the value is zero ($\rho = 0$), it means that there is no relationship between the variables.

Example: Spearman's correlation coefficient is useful for ranking radiologists based on their diagnostic accuracy across different imaging modalities, as it measures the strength and direction of the monotonic relationship between two ranked variables. In this context, if each radiologist is assigned a rank based on their accuracy within each imaging modality (e.g., MRI, CT, ultrasound), the Spearman's correlation can help assess if there is a consistent pattern in diagnostic performance across modalities. A high positive Spearman's correlation (close to +1) would suggest that radiologists who rank highly in one modality tend to rank highly in others as well, indicating consistent diagnostic accuracy. Conversely, a low or negative Spearman's correlation would suggest variability in accuracy across modalities, with some radiologists excelling in certain types of imaging while underperforming in others. This analysis does not assume a linear relationship, making it suitable for ranked data.

Kendall's Tau

Kendall's rank correlation is a nonparametric test procedure. For this the data must not be normally distributed and both the variables must have an ordinal scale level. It is very similar to Spearman's rank correlation. Spearman's rank correlation calculates the correlation based on the difference in ranks of paired values, placing more emphasis on the magnitude of these differences. On the other hand, Kendall's rank correlation (often referred to as Kendall's tau) is based on the concept of concordant and discordant pairs, assessing the order or "direction" consistency between paired values rather than rank difference size. Kendall's tau ignores the degree of rank differences and focuses solely on whether pairs agree or disagree in rank order. Kendall's tau is preferred over Spearman's rank correlation and is used when there are small or highly variable data sets with potential outliers, or a more conservative estimate of correlation based on directionality rather than the magnitude of differences is to be assessed.[7]

Example: In a study linking radiologists' experience levels to diagnostic confidence in detecting early findings of Alzheimer's disease on MRI, Kendall's tau would offer a stable correlation estimate, focusing on the consistency in rank order rather than the size of rank differences. This is particularly useful when there are potential outliers, such as experienced radiologists with unexpectedly low confidence, as Kendall's tau is less affected by such variations. Thus, it provides a more conservative and robust measure of association in studies with limited data, ensuring reliable insights for preliminary analyses or future research planning.

Point-Biserial Correlation

Point-biserial correlation is used when one of the variables is dichotomous and the other has a metric scale. To calculate point-biserial correlation, one of the two expressions of the dichotomous variable is coded as 0 and the other as 1. In point-biserial correlation, the binary variable (0 or 1) acts like a two-group categorical variable, with the mean difference between the continuous variable in each group playing a role similar to covariance in the Pearson formula. Like Pearson's r, r_{pb} standardizes this relationship by dividing the mean difference by the standard deviation of the continuous variable, thereby producing a coefficient that ranges from –1 to +1. Essentially, the point-biserial correlation can be seen as a special case of Pearson's correlation where one variable is dichotomous, allowing it to quantify the linear association between a continuous variable and a binary variable.[8]

A $$r = \frac{\sum(X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum(X_i - \overline{X})^2 \cdot \sum(Y_i - \overline{Y})^2}}$$

B $$r_s = 1 - \frac{6\sum d_i^2}{n(n^2 - 1)}$$

C $$r_{pb} = \frac{\overline{X_1} - \overline{X_0}}{s_X}\sqrt{\frac{n_1 \cdot n_0}{n^2}}$$

Fig. 2 Formula for Pearson's correlation coefficient is shown in (**A**) [X_i and Y_i are individual sample points, $\bar{X}$ and $\bar{Y}$ are the mean values of X and Y, the summation (Σ) runs over all sample points]. Spearman's correlation coefficient is calculated as per the formula in (**B**) [rank the values of X and Y separately, d_i is the difference between the ranks of each pair of observations (X_i and Y_i), n is the number of observations]. Point-biserial correlation is calculated using the formula in (**C**) [X is a continuous variable and Y is a binary variable coded as 0 and 1, $\bar{X}_1$ is the mean of X for the group where $Y = 1$, $\bar{X}_0$ is the mean of X for the group where $Y = 0$, s_x is the standard deviation of X, n_1 and n_0 are the number of observations in the groups where $Y = 1$ and $Y = 0$, respectively, n is the total number of observations ($n = n_1 + n_0$)].

Formulae for Pearson's correlation coefficient, Spearman's correlation coefficient, and point-biserial correlation are shown in ►**Fig. 2**.

Correlation analysis is not possible with nominal data. Equivalent tests are: chi-square test, Cramér's V, contingency coefficient, and phi coefficient.

Causality

To establish causality one needs to demonstrate that changes in one variable directly results in changes in another. It requires rigorous experimentation and control of confounding factors.[9,10]

Basic Criteria for Causality

1. *Temporal precedence*: Cause must precede the effect in time.
2. *Association*: Statistical association must be there between the cause and effect.
3. *Nonspuriousness*: Relationship must not be due to a third, confounding variable.

Case Study: Correlation versus Causation in Radiology

A study finds a positive correlation between the use of a new gadolinium-based contrast agent in MRI and the incidence of nephrogenic systemic fibrosis (NSF) in patients with renal impairment. While correlated, additional research is needed to establish whether gadolinium exposure directly causes NSF or if other factors contribute to the association.

Regression Analysis

Regression analysis models the relationship between a dependent variable and one or more independent variables (thereby analyzing the degree of relationship between the variables). It thus makes it possible to infer or predict another variable based on one or more variables. The variable to be inferred is called the dependent variable and that used for prediction is the independent variable.[11] A regression analysis can be used to predict one variable from another only when a statistically significant correlation between the variables is obtained.

For example, regression analysis can be used to determine the relationship between the grade of dural arteriovenous fistula and the patient's cognitive function score (measured using a standardized cognitive assessment tool). Here, the independent variable (predictor) is the grade of dural arteriovenous fistula and the dependent variable (criterion) is the cognitive function score.

Regression analysis can be used basically for two purposes:

1. Measurement of the influence of one or more variables on another variable.
2. Prediction of a variable by one or more variables.

Types of regression analysis (►**Table 1**, ►**Fig. 3**):

1. Simple linear regression (univariate regression).
2. Multiple linear regression and multivariate regression.
3. Logistic regression.
4. Cox proportional regression.

For performing a linear regression, a linear relationship is needed between the independent variables and the dependent variables.

Control Variable (Covariate)

A control variable (covariate) is an extra independent variable, which is included in the regression analysis to account for possible confounding factors. This is done to isolate the

Table 1 Types of regression analysis and the type of dependent and independent variables in each of these regression models

	Number of independent variables	Scale of measurement dependent variable	Scale of measurement independent variable
Simple linear regression	One	Metric	Metric, ordinal, nominal
Multiple linear regression	Multiple	Metric	Metric, ordinal, nominal
Logistic regression	Multiple	Ordinal, nominal	Metric, ordinal, nominal

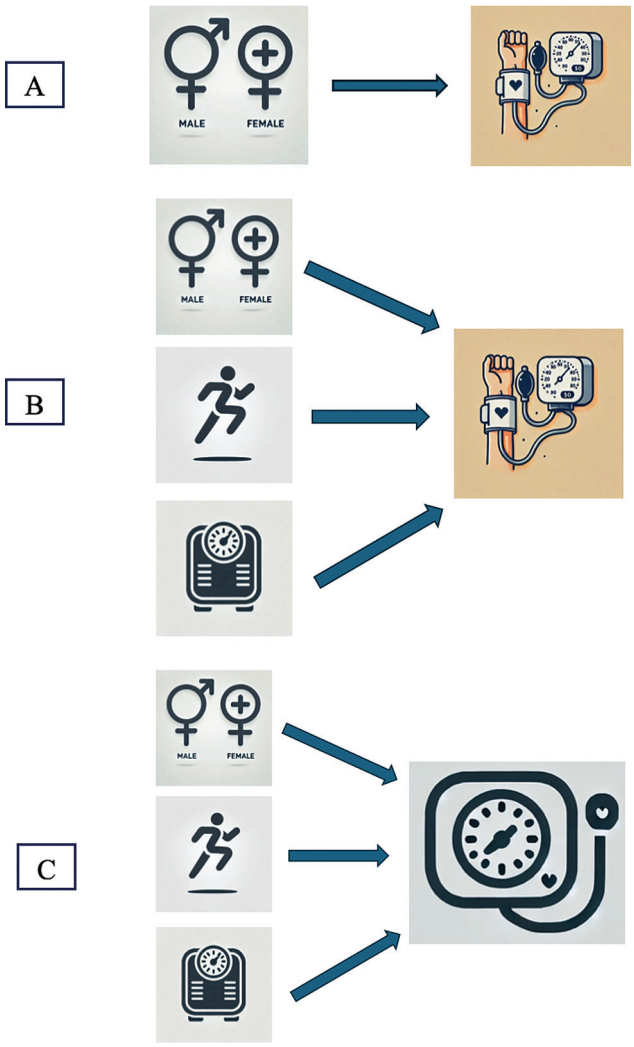

Fig. 3 (**A**) Demonstrates how simple linear regression can be used to analyze the relationship between gender and systolic blood pressure (in mm Hg). (**B**) Demonstrates how multiple linear regression can be used to analyze the relationship between gender, exercise, and weight with systolic blood pressure (in mm Hg). (**C**) Demonstrates how logistic regression can be used to analyze the previous three variables with presence or absence of hypertension (as a dichotomous variable).

relationship of interest between the independent variable and the dependent variable (ensuring that there are no unobserved factors affecting the relationship).[12]

Inclusion of control variables has the following advantages:

1. Reducing omitted variable bias.
2. Increasing precision (reduces residual variance).
3. Accounting for confounding.

Two basic things to keep in mind while performing a regression analysis are:

- Inclusion of irrelevant control variables can complicate the model resulting in reduced power of analysis.
- Omitting important control variables can lead to biased estimates.

To select relevant control variables (covariates) for a regression model, it is essential to consider factors that may confound the relationship between the independent and dependent variables, meaning they are associated with both and could bias the results if unaccounted for.[13] To select relevant control variables these steps need to be meticulously followed:

1. *Literature review*: Reviewing previous research helps identify commonly used control variables in similar studies and understand potential confounders based on existing evidence.
2. *Theoretical justification*: Only include variables that have a plausible theoretical relationship with both the independent and dependent variables. Irrelevant variables can add noise rather than clarity to the model.
3. *Statistical testing*: Use correlation matrices or partial regression plots to assess associations between potential covariates and the main predictor and outcome. Variables that are highly correlated with both may be relevant confounders.
4. *Domain expertise*: Collaborate with experts in the field (e.g., clinicians, radiologists) who can provide insights into factors likely to influence both the predictor and outcome, which may not always be obvious through statistical analysis alone.
5. *Avoid overfitting*: Avoid including too many control variables, as this can overfit the model, especially in smaller data sets. Focus on the most influential confounders to keep the model interpretable.

Including only well-chosen control variables strengthens the model by adjusting for confounding effects without introducing unnecessary complexity.

Simple Linear Regression

Simple linear regression examines the relationship between a dependent variable and a single independent variable. It is a type of univariate regression. Goal of simple linear regression is to predict the value of a dependent variable based on an independent variable. Greater the linear relationship between the two variables, more accurate will be the prediction.

Task of simple linear regression is to exactly determine the straight line that describes the linear relationship between the dependent and independent variable on a scatter plot. To achieve this, the method of least squares is used.[14]

Attempt is made to keep the error in estimation as small as possible, so that the distance between the estimated value and the true value should be as small as possible. The distance is called the "residual" and is abbreviated as $\in$ (epsilon).

When the regression line is calculated, it is attempted to determine the regression coefficients (a and b), so that the sum of the squared residuals is minimal. Regression coefficient (b) can be interpreted as follows:

- $b > 0$: Positive correlation between x and y.
- $b < 0$: Negative correlation between x and y.
- $b = 0$: No correlation between x and y.

There are multiple assumptions in linear regression analysis that should be kept in mind, which are: linearity,

homoscedasticity, normality, no multicollinearity, and no autocorrelation.[15]

1. *Linearity*: The relationship between predictor and outcome variables should be linear. For example, if predicting tumor size on follow-up MRI based on initial size, we assume that changes in tumor size over time are proportional. If growth is exponential or follows a complex pattern, linear regression might not be appropriate, and a different model might be needed.
2. *Homoscedasticity*: The variance of residuals (errors) should be constant across all levels of the predictor variable. For instance, if we are predicting lesion density on CT scans from patient age, homoscedasticity assumes that the spread of density values is similar across all ages. If older patients show a wider spread of densities around the predicted line, it indicates heteroscedasticity, which can make regression estimates less reliable.
3. *Normality*: The residuals (differences between observed and predicted values) should be normally distributed. For example, if we predict radiographic severity scores of a disease based on patient history and laboratory values, the residuals should ideally form a normal distribution. If the residuals are heavily skewed, it could suggest that outliers or nonnormal variables are influencing the predictions.
4. *No multicollinearity*: Predictor variables should not be highly correlated. For example, when predicting the likelihood of malignancy based on imaging features, using both "nodule diameter" and "volume" as predictors could introduce multicollinearity, as these two measures are highly correlated. This makes it difficult to interpret the separate effect of each variable on malignancy risk.
5. *No autocorrelation*: The residuals should be independent. In longitudinal studies, such as tracking tumor response on serial CT scans, autocorrelation might occur if measurements are taken close together in time, making one residual similar to the previous. Autocorrelation can bias the model, as it violates the assumption that each measurement is independent.

Meeting these assumptions in radiology studies is crucial to ensure the validity of the regression model's predictions and interpretability of its results.

Recommendation for minimum sample size in regression analysis is 10 observations per predictor variable. For example, if you have 3 predictor variables, a sample size of at least 30 is typically recommended. However, this is a rule of thumb, and larger samples are often preferred, especially if the data are complex or if you are using more sophisticated regression models, to ensure the results are robust and generalizable.[16]

Example:

Let us consider a small sample of data (on intimo-medial thickness [IMT] from carotid Doppler) collected from 10 patients:

Patient ID	IMT (mm)	Stenosis (%)
1	0.6	20
2	0.7	30
3	0.8	25
4	0.9	35
5	1.0	40
6	1.1	45
7	1.2	50
8	1.3	55
9	1.4	60
10	1.5	65

Performing the regression analysis:

Using simple linear regression, we aim to fit a line to the data that best describes the relationship between IMT and stenosis percentage.

The regression equation is:

$$\text{Stenosis}\ (\%) = b \times \text{IMT}\ (\text{mm}) + a$$

Where:

- a is the intercept (the expected stenosis percentage when IMT is 0).
- b is the slope (the change in stenosis percentage for each one-unit change in IMT).

Calculations:

Performing the calculations (which can be done using statistical software or manually), let us assume we find the following results:

- $a = 10$
- $b = 40$

So, the regression equation becomes:

$$\text{Stenosis}\ (\%) = 40 \times \text{IMT}\ (\text{mm}). + 10$$

Interpretation:

- Intercept (a): When IMT is 0 mm, the predicted stenosis percentage is 10%. This might not be clinically relevant but is part of the mathematical model.
- Slope (b): For each additional millimeter of IMT, the stenosis percentage increases by 40%.

Usage:

- If a patient has an IMT of 1.2 mm, the predicted stenosis percentage would be: $\text{Stenosis}\ (\%) = 10 + 40 \times 1.2 = 10 + 48 = 58\%$.

Conclusion

This simple linear regression model indicates a strong linear relationship between IMT and the percentage of carotid artery stenosis. This can be used by radiologists to estimate stenosis severity based on ultrasound measurements of IMT, thereby aiding in the diagnosis and management of patients with carotid artery disease.

Multiple Linear Regression

Multiple linear regression extends simple linear regression to include multiple independent variables.

Multiple linear regression should not be confused with multivariate regression. In multivariate regression, several regression models are calculated to allow conclusions to be drawn about several dependent variables.[17]

Example: Modeling the relationship between age, gender, and bone mineral density in osteoporosis studies.

Considerations for multivariable models are:

Independence of covariates: The relationship between variables included in the model should be examined using a separate regression or correlation analysis. If a significant interaction between variables is found, an interaction term should be added to the model.

Univariate model results: Before presenting multivariable model results, it is important to report the results of univariate models either in the same table or in a separate table for comparison.

Reporting of Regression Results

Units and reference categories: The units of measurement for continuous variables and the reference categories for categorical variables should be clearly specified in the regression table.

Independent table readability: Each regression table should be fully understandable on its own, and the title should clearly indicate the outcome variable being modeled.

Multivariate regression is used in radiology when there is a need to predict multiple continuous outcomes simultaneously based on several predictors. For example, in assessing patient outcomes after liver ablation procedures, a radiologist might want to predict both postprocedure liver function (measured by liver enzyme levels) and tumor size reduction on follow-up imaging, based on predictors such as patient age, baseline liver function, and ablation technique used. Multivariate regression allows the simultaneous evaluation of how these predictors influence both outcomes, providing a more comprehensive analysis of treatment effects and patient factors on multiple clinical metrics.[18]

To perform a multivariate analysis after a univariate analysis, include all statistically significant variables from the univariate analysis into a multivariable model, adjusting for potential confounders, and assess the combined effect of these variables on the outcome while controlling for their interrelationships.

Logistic Regression

Logistic regression is a type of regression analysis where the dependent variable is nominally scaled. Logistic regression makes it possible to explain the dependent variable or estimate the probability of occurrence of the categories of the variable.[19–21]

In analysis of data sets, objective is to predict outcomes in many cases (that too particularly binary outcome prediction). This is where the importance of logistic regression lies. In logistic regression such binary outcomes can be predicted where the input includes categorical or continuous variables. In the simplest form of logistic regression, dichotomous variables (0 or 1) can be predicted. The probability of the occurrence of value 1 (with a particular characteristic present) is estimated. In other words, the dependent variable is made dichotomous (0 or 1) and the probability that the expression 1 occurs is estimated.

In logistic regression, the odds ratio (OR) is a measure of association between a predictor variable and the outcome. It quantifies how the odds of the outcome change with a one-unit increase in the predictor, holding other variables constant. For example, if we are using logistic regression to predict the likelihood of lung cancer based on smoking status (smoker or nonsmoker), an OR of 3 for smoking would mean that smokers have three times the odds of having lung cancer compared with nonsmokers, assuming all other factors are constant. An OR greater than 1 indicates a positive association (higher odds), an OR less than 1 indicates a negative association (lower odds), and an OR of 1 means no association. ORs are commonly used in logistic regression to interpret the impact of individual predictors on a binary outcome.[19]

Key assumptions in logistic regression are the following[20]:

1. *Linearity of the logit*: Logistic regression assumes a linear relationship between continuous predictors and the log odds of the outcome. For instance, if predicting the likelihood of a positive biopsy based on tumor size, the log odds of malignancy should increase linearly with tumor size on imaging. If the relationship is nonlinear, it may distort predictions.
2. *Independent observations*: Each observation should be independent of others. For example, when predicting pneumonia presence based on chest radiograph features, each patient's chest radiograph should be analyzed independently. Correlated observations, such as repeated scans for the same patient, could violate this assumption and require adjustments.
3. *No multicollinearity*: Predictor variables should not be highly correlated. For instance, if using both "lesion density" and "lesion enhancement" as predictors for malignancy, multicollinearity could arise if these variables are strongly related, leading to unreliable estimates.
4. *Absence of outliers with large influence*: Outliers should not disproportionately influence the results. For example, if one patient has an unusually high lesion size compared with others in a study on brain tumor detection, this outlier might skew the model's predictions. Identifying and handling such outliers is essential.
5. *Adequate sample size*: Logistic regression requires a sufficient number of events per predictor to provide stable estimates. For instance, predicting rare conditions like a specific subtype of bone tumor on MRI would require enough cases to avoid unstable and unreliable results.

Meeting these assumptions ensures that the logistic regression model provides reliable and interpretable predictions for clinical applications in radiology.

$$\boxed{\text{A}} \quad \hat{y} = b \cdot x + a$$

$$\boxed{\text{B}} \quad \hat{y} = b_1 \cdot x_1 + b_2 \cdot x_2 + \cdots + b_k \cdot x_k + a$$

$$\boxed{\text{C}} \quad f(z) = \frac{1}{1 + e^{-z}} = \frac{1}{1 + e^{-(b_1 \cdot x_1 + \cdots + b_k \cdot x_k + a)}}$$

Fig. 4 Simple linear regression is analyzed on the basis of equation in (**A**) [$\hat{y}$ – estimated dependent variable (y value estimated for each x value), b – slope (gradient of the straight line), x – independent variable, a – y intercept]. Multiple linear regression is analyzed on the basis of equation in (**B**) [b_1, b_2, etc. representing the slopes of each set of variables and x_1, x_2, etc. representing the independent variables). Logistic regression is analyzed on the basis of equation in (**C**).

A frequent application of logistic regression in radiology is to find out which of the variables have an influence on a particular pathology. In this example, 0 could stand for those not having the pathology and 1 for those having it. Similarly, the effect of age and gender on this particular pathology could be examined (pathology is the dependent variable, and age as well as gender are the independent variables).

Logistic model is basically based on the logical function. Unique thing about the logistic function is that for values between minus and plus infinity, it assumes only values between 0 and 1. Equations for simple linear regression, multiple linear regression, and logistic regression are shown in ►**Fig. 4**.

Example:

Objective:

- To predict the probability that a pulmonary nodule detected on a chest CT scan is malignant based on various clinical and imaging features.

Data:

- Age: Age of the patient in years.
- Sex: Gender of the patient (0 for female, 1 for male).
- Nodule size: Size of the nodule in millimeters.
- Nodule location: Location of the nodule in the lung (0 for upper lobe, 1 for middle lobe, 2 for lower lobe).
- Nodule shape: Shape of the nodule (0 for round, 1 for irregular).
- Smoking history: Smoking history of the patient (0 for nonsmoker, 1 for current/former smoker).

Logistic regression model:

- Dependent variable (outcome) is malignancy (0 for benign, 1 for malignant).

Interpretation:

- The analysis will provide coefficients for each predictor variable along with their statistical significance. The coefficients represent log odds of the outcome occurring for a one-unit increase in the predictor variable. Statistically significant variables (usually $p < 0.05$) are considered to be strong predictors of the outcome.

Example output:

- Age: A 1-year increase in age increases the odds of malignancy.
- Sex: Being male (compared with female) affects the odds of malignancy.
- Nodule size: A 1-mm increase in nodule size increases the odds of malignancy.
- Nodule location: Nodules in different lung lobes have varying odds of being malignant.
- Nodule shape: Irregularly shaped nodules have higher odds of being malignant compared with round nodules.
- Smoking history: Having a history of smoking increases the odds of malignancy.

When the dependent variable has two characteristics (male, female), that is, it is dichotomous, then binary logistic regression is used. However, if the dependent variable has more than two characteristics, for example, using Borden grade (1, 2, or 3) of dural arteriovenous fistulae as the dependent variable and performing a logistic region with gender as the independent variable. Analyzing this logistic regression could help us understand whether being male (compared with female) affects the odds of having a higher Borden classification.

Case Study: Regression Analysis in Radiology Research

A research study aims to predict patient survival time following pancreatic adenocarcinoma diagnosis using demographic and clinical variables. Multiple linear regression identifies significant predictors such as tumor size, patient age, and treatment modality, providing insights into factors influencing patient outcomes and guiding personalized treatment strategies.

In summary, a regression analysis "model" refers to a mathematical framework or equation used to describe the relationship between one or more independent variables (predictors) and a dependent variable (outcome). The goal of a regression model is to predict the value of the dependent variable based on the values of the independent variables.

Table 2 Various output variables generated on processing univariate and multivariate regression in SPSS software

Output	Univariate regression	Multivariate regression	Description
R-squared (R^2)	Proportion of variance explained by one predictor	Proportion of variance explained by all predictors combined	Indicates how much of the variance in the dependent variable is explained by the predictor(s). Higher values suggest a stronger model fit
Adjusted *R*-squared	Not much different from R^2 in univariate	Adjusted for the number of predictors to avoid overestimation	Adjusts *R*-squared to account for the number of predictors, providing a more accurate measure of model performance in multivariate regression
Standard error	Measures the average distance from the regression line	Measures the average distance from the regression line	Shows the average error in predictions made by the model. Lower values indicate a better fit
F-statistic	Tests overall significance of the single predictor	Tests overall significance of all predictors together	Assesses whether the model significantly predicts the dependent variable. A significant *F* indicates the model is a good fit
p-Value for *F*-statistic	Tests whether the predictor significantly predicts the outcome	Tests whether all predictors combined significantly predict the outcome	If the *p*-value is below a threshold (e.g., 0.05), it suggests the overall model is statistically significant
Unstandardized coefficient	Coefficient of the single predictor (*B*)	Coefficients of all predictors (*B*)	Indicates how much the dependent variable changes with a one-unit change in the independent variable(s)
Standardized coefficient	Shows the strength of the relationship between predictor and outcome	Standardized coefficients allow for comparison between variables	Expresses the relationship in terms of standard deviations, useful for comparing the relative effect of predictors in multivariate models
***t*-Statistic**	Tests significance of the single predictor	Tests significance of each predictor in the model	Shows whether the individual predictor (s) have a significant effect on the outcome variable
p-Value for coefficient	Significance of the predictor	Significance of each predictor in the model	A *p*-value less than 0.05 typically indicates that the predictor is a statistically significant contributor to the model
Collinearity statistics	Not applicable.	Variance inflation factor (VIF) and tolerance to check multicollinearity	Indicates whether predictors are highly correlated, which can distort the model. High VIF (> 10) indicates multicollinearity issues
Residual statistics	Standardized residuals for model diagnostics	Standardized residuals for model diagnostics	Assesses how well the model fits by checking if the residuals (errors) are normally distributed and homoscedastic (constant variance)
Cook's distance	Not applicable	Cook's distance identifies influential data points	Helps detect outliers or influential points that might distort the regression results. Values > 1 indicate influential points to investigate further

Various output variables generated on processing univariate and multivariate regression in SPSS software have been demonstrated in ►**Table 2**.

In regression analysis, the unstandardized coefficient for a nominal or categorical independent variable represents the change in the dependent variable's predicted value when that categorical variable is present, relative to a reference category, while holding other variables constant. For instance, suppose we are analyzing the effect of MRI machine type (a categorical variable with three categories: "Type A," "Type B," and "Type C") on scan quality scores. If "Type A" is the reference category, the unstandardized coefficient for "Type B" indicates the average difference in scan quality between "Type B" and "Type A." A positive coefficient suggests that "Type B" produces a higher score than "Type A," while a negative coefficient indicates a lower score. The interpretation of these coefficients shows how each category compares to the reference group in terms of its effect on the dependent variable, allowing us to understand the impact of each categorical level. Unstandardized coefficient for continuous independent variable has been explained in ►**Table 2**.[22]

The validity of a regression model depends on how well it satisfies several key assumptions and its ability to generalize to new data. Assumptions such as linearity, independence of errors, homoscedasticity (constant error variance), normality of residuals, and no multicollinearity are essential to ensure accurate and unbiased estimates. If these assumptions are violated, the model's predictions and inferences may be unreliable. Additionally, for a model to be valid, it must generalize well to new or unseen data, indicating that it captures the underlying patterns rather than just fitting the specific data set used to train it. This generalization is often tested through techniques like cross-validation, and the model's performance on hold-out or test data can indicate its robustness and predictive power in real-world applications.[23]

Let us consider another example of univariate and multivariate regression analysis:

Univariate regression analysis (example: age and blood pressure): In univariate regression, only one predictor variable—age—is used to predict blood pressure. Suppose the *R*-squared value is 0.57, this means that age alone explains 57% of the variability in blood pressure. This is a relatively strong *R*-squared value, suggesting that age is a major factor in determining blood pressure. If the unstandardized coefficient for age is 1.23, it means that for every additional year in age, blood pressure increases by 1.23 units. The *p*-value we obtain for age is 0.000, indicating that this relationship is statistically significant (typically, a *p*-value below 0.05 is considered significant). This means we can be highly confident that age is a meaningful predictor of blood pressure.

Multivariate regression analysis (example: age, weight, and blood pressure): In multivariate regression, two predictors—age and weight—are included to predict blood pressure. If the model's *R*-squared value is 0.65, meaning that together, age and weight explain 65% of the variability in blood pressure. This is an improvement over the univariate model, indicating that adding weight as a predictor increases the explanatory power of the model. The coefficients reveal that both age and weight contribute to predicting blood pressure. Specifically, if the coefficient for age is now 0.95, it means that each additional year in age increases blood pressure by 0.95 units, holding weight constant. For weight, if the coefficient is 0.45, meaning that each additional kilogram increases blood pressure by 0.45 units, holding age constant. If both *p*-values are less than 0.05, it indicates that age and weight are statistically significant predictors of blood pressure in this multivariate model.

In multivariate regression, multicollinearity occurs when predictor variables are highly correlated, which can distort the estimated coefficients and make it difficult to assess the individual effect of each predictor. In this example, multicollinearity is assessed using the variance inflation factor (VIF), which measures how much the variance of a regression coefficient is inflated due to multicollinearity. Suppose, VIF values are within acceptable limits (typically, VIF values below 5 or 10 are considered acceptable), it means that multicollinearity is not an issue. This assures us that the estimated effects of age and weight on blood pressure are reliable and that both predictors are contributing independently to the model.[24]

In both analyses, the models explain a substantial amount of the variability, and the predictors significantly influence the outcome.

Cox's Regression Analysis (Cox Proportional Hazards Model)

Cox regression analysis assesses the effect of multiple independent variables on a time-to-event outcome. In other words, it assesses whether there are other parameters having an influence on the survival time. The main goal is to test a hypothesis about independent variables or to build a predictive model. For example, it can be used to identify prognostic factors affecting the survival of patients with glioblastoma using Cox proportional hazards model. If you want to assess age of the subjects as a predictor of survival, proportional hazards model is used. It evaluates the effect of each predictor (e.g., age) on the shape of the survival curve.[25,26]

Cox regression takes into account six assumptions[27]:

1. *Proportional hazards assumption*: The effect of each predictor on survival is constant over time, meaning the hazard ratios (HRs) are proportional throughout the study period.
2. *Independence of survival times*: Each individual's survival time is independent of others, without interference or dependency between subjects.
3. *Linearity assumption*: The relationship between continuous predictors and the log hazard is linear, ensuring that the predictors' effects are accurately modeled.
4. *No multicollinearity*: Predictors should not be highly correlated, as multicollinearity can distort the estimated effects and lead to unreliable results.
5. *No outliers*: Extreme values in the data should be avoided, as they can disproportionately influence the results and skew the model.
6. *No effect modification*: There should be no interactions between predictors affecting the hazard rate, as these would require a more complex model to account for varying effects.

Alternatives if the assumptions are violated[28]:

- *Time-dependent Cox regression*: This model allows HRs to change over time, relaxing the proportional hazards assumption by including interaction terms between predictors and time.
- *Stratified Cox model*: If certain variables violate proportionality, you can stratify by those variables, which allows for different baseline hazard functions in each stratum without estimating separate coefficients.
- *Accelerated failure time (AFT) model*: AFT is a parametric alternative that does not rely on the proportional hazards assumption. It models the effect of covariates on the survival time itself rather than on the hazard rate, offering flexibility with different distributions (e.g., Weibull, exponential).
- *Flexible parametric models*: Models like restricted cubic splines allow for a more flexible hazard function, accommodating nonlinear effects and time-dependent HRs.

In the context of Cox proportional hazards regression, there are important relationships between the β coefficient (β), standard error(SE), *p*-value, HR, and the 95% confidence interval (CI) for the HR.[29] Here is how these terms are connected:

- *Beta coefficient (β)*: The β coefficient (also known as the log hazard or log of the HR) represents the effect of a covariate (independent variable) on the hazard or risk of the event happening. A positive β indicates that the variable increases the hazard (i.e., higher risk), while a negative β suggests a protective effect (i.e., lower risk).
- *HR*: The HR represents the relative risk of the event occurring (e.g., death, disease) associated with one unit increase in the covariate. It is a multiplicative measure:
  - HR = 1 means no effect of the covariate on the hazard.
  - HR > 1 means increased hazard (higher risk).
  - HR < 1 means reduced hazard (lower risk).
- *SE*: The SE of the β coefficient measures the variability or uncertainty of the estimated coefficient. A large SE relative to the β coefficient suggests that the estimate is not very precise. The SE is used to compute the 95% CI and the *p*-value for the HR.
- *p-Value*: The *p*-value tests the null hypothesis that the β coefficient (and thus the HR) is equal to zero (no effect). A low *p*-value (typically < 0.05) suggests that the effect of the covariate on the hazard is statistically significant.
- *95% CI for the HR*: The 95% CI for the HR provides a range of values within which the true HR is likely to fall, with 95% confidence. If the 95% CI for the HR includes 1, the effect is not statistically significant at the 5% level, as a HR of 1 indicates no effect of the variable on the outcome.

These components together allow you to assess both the magnitude and the reliability of the effect of covariates on the hazard in survival analysis.

Cox proportional hazards model is widely used in survival analysis to evaluate the effect of several variables on the time until an event occurs, such as death or disease recurrence, without needing to specify the underlying survival distribution. This model estimates HRs for each predictor variable, indicating how the risk of the event changes with each unit increase in the variable, while holding other factors constant. For example, in a study examining the survival of patients with lung cancer based on imaging findings and clinical factors, the Cox model can assess how variables like tumor size, stage, or treatment type influence the hazard (risk) of mortality. A key assumption of the Cox model is proportional hazards, meaning that the relative effect of each predictor on the hazard rate is consistent over time. This model is valuable for understanding prognostic factors and guiding treatment decisions in clinical settings.

Survival Analysis

Survival analysis is a group of statistical methods that deals with the concept of time-to-event data (with the variable under study being the time until an event occurs). It is often encountered in studies that involve patient survival, disease progression, or treatment efficacy.[30–33]

One important concept to understand in survival analysis is that the event need not be a negative one. It could also be a positive one such as complete remission after start of a particular chemotherapeutic drug. So, basically it means that "survival analysis" in some scenarios may not have anything to do with actual survival time.

Survival analysis is essential in radiology because it helps evaluate patient prognosis and the effectiveness of imaging-based interventions over time. For example, in assessing the impact of radiofrequency ablation for liver tumors, survival analysis can determine the time to recurrence or overall survival following the procedure. By analyzing this data, radiologists and oncologists can better understand which patients are likely to benefit most from the treatment, guide follow-up imaging schedules, and adjust treatment plans for improved patient outcomes. Survival analysis in such cases provides critical insights into the efficacy of radiology-guided treatments and supports evidence-based decision-making in patient care.

Censored Data

In survival analysis, we need to understand that there may be scenarios where the event does not occur at all, occurs much later than our final follow-up time, or follow-up could not be obtained. All these are summarized under the concept of censoring. Primary reason for this is that the study cannot go on indefinitely and must stop at some point.[34] "At risk" refers to a subject who has not yet experienced the event and is still being monitored for the event.

Methods of Survival Time Analysis

1. Kaplan–Meier survival time curve.
2. Log-rank test.
3. Cox regression analysis.

Kaplan–Meier Survival Time Curve

The Kaplan–Meier survival time curve graphically represents the survival rate or survival function. Here, survival rate is plotted on the *y*-axis and time on the *x*-axis. Censored data points are also considered in this curve where the event of interest has not occurred.[35,36]

Median survival is the time half of the subjects survive. It is a useful summary measure and is often reported with Kaplan–Meier survival curves. This can be reported from the curve as long the curve dips below the 0.50 survival point on the *y*-axis (►**Fig. 5**).

Basically, the Kaplan–Meier survival curve shows the cumulative survival probability. A steeper slope indicates a higher event rate (death rate) and therefore a worse survival prognosis. A flatter slope on the other hand indicates a lower event rate and therefore a better prognosis. By using multiple curves, different groups can be compared (wherein diverging curves represent differences in survival between the groups). Survival probability at a particular point can also be located by drawing a vertical line at that time point (on the *x*-axis) to meet the curve, and then reading the corresponding survival probability from the *y*-axis.

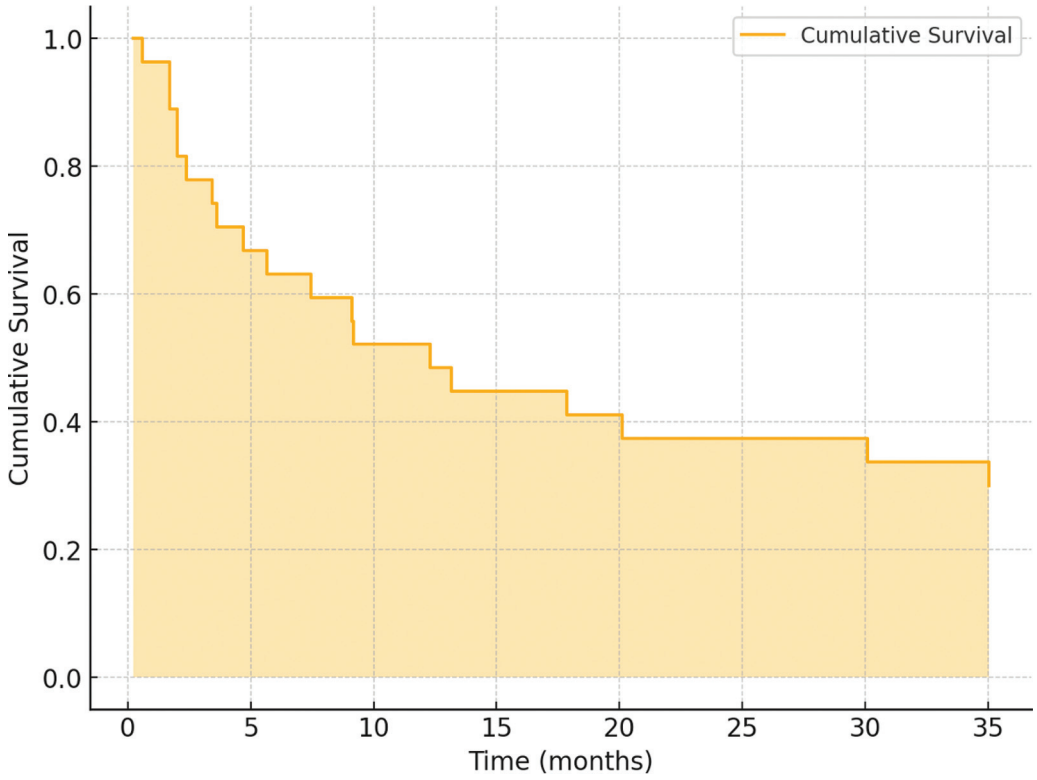

Fig. 5 Kaplan–Meier survival curve for hepatocellular carcinoma patients posttransarterial chemoembolization. It shows the survival probability decreasing over time, which is typical in survival analysis as more patients experience the event (death) or are censored.

Kaplan–Meier survival curve takes into account these assumptions[36]:

1. *Random or noninformative censoring*: Censoring (where follow-up is incomplete for some patients) should occur randomly and not be related to the likelihood of the event. For example, if patients with aggressive tumors on MRI are more likely to drop out of a survival study, this could bias the results, as their missing data might not be random.
2. *Independence of censoring*: The survival times of censored patients should be similar to those of patients who remain in the study. For instance, if some patients undergoing a radiological procedure are lost to follow-up due to moving away, this should not be related to their survival probability to avoid skewing survival estimates.
3. *Survival probabilities do not change over time*: The population's risk of survival should be consistent over the study period. In a study evaluating the survival of patients with metastatic cancer based on positron emission tomography-CT findings, this assumption would mean that advancements in treatment during the study period should not impact survival probabilities, which might otherwise make early and late survival probabilities incomparable.
4. *No competing risks*: There should be no alternative risks that could prevent the occurrence of the event of interest. For example, if studying survival after a radiologically guided tumor ablation, the analysis assumes that no other major health events (e.g., fatal cardiovascular events) interfere, as these would alter the survival probabilities independently of the tumor's progression.
5. *Homogeneity of groups*: This assumption implies that within each group, patients are similar in terms of underlying survival-related factors. For example, if comparing survival times after two different radiological interventions for liver cancer (e.g., radiofrequency ablation vs. microwave ablation), it is assumed that patients in both groups are similar in terms of tumor size, liver function, and overall health status. If the groups differ significantly in these factors, observed survival differences may reflect underlying patient characteristics rather than the true effect of the interventions. Without homogeneity, Kaplan–Meier curves might misrepresent the actual effect of each intervention on survival.
6. *Absence of time-dependent variables*: Kaplan–Meier assumes that factors influencing survival remain constant over time, which means it cannot accommodate time-dependent variables—factors that change as the study progresses. For instance, in a study on survival after CT-guided lung biopsy, patients may start on additional treatments (like chemotherapy) after the biopsy. This additional treatment, which changes over time, could influence survival but cannot be accounted for in a standard Kaplan–Meier analysis. Similarly, if a patient's disease stage worsens over time, affecting survival likelihood, Kaplan–Meier would not reflect this progression, potentially oversimplifying the survival outcome.

Log-Rank Test

Log-rank test compares distribution of time until an event occurs of two or more independent samples. For example,

long-rank test can be used to compare the survival times of two different groups of glioblastoma patients (one group treated with standard radiotherapy vs. another treated with radiotherapy plus temozolamide). It basically tells you whether there is a significant difference between the two curves. Null hypothesis in log-rank test is that both groups have similar survival rates (null hypothesis is then rejected if the p-value is < 0.05).[37,38]

Log-rank test takes into account three assumptions:

1. Random or noninformative censoring.
2. Independence of survival times.
3. Proportional hazards assumption (hazard rates [rate of an event occurring] should be consistent over time).

The log-rank test can be used to compare survival times between two groups, such as patients undergoing two different radiological treatments for liver cancer. For example, group A receives radiofrequency ablation, while group B receives microwave ablation. The log-rank test would assess if there is a statistically significant difference in survival times between the two groups, assuming that censoring is random (e.g., patients lost to follow-up are unrelated to treatment outcome), that each patient's survival time is independent of others, and that the hazard rates for both treatments remain proportional over time. If these assumptions hold, the log-rank test can reliably indicate whether one treatment leads to longer survival.

Cox proportional hazards model can be likened to the multivariable analysis of log-rank test.

Example: Survival analysis of patients with hepatocellular carcinoma (HCC) treated with transarterial chemoembolization (TACE).

Objective:

To evaluate the overall survival of patients with HCC who undergo TACE.

Study design:

- Population: Patients diagnosed with HCC and treated with TACE.
- Time frame: Follow-up period of 5 years.
- Data collection: Patient demographics, tumor characteristics, treatment details, and follow-up data including survival status and time of death or last follow-up.

Methodology:

1. Kaplan–Meier survival curves:
   - ➢ Survival time data are arranged in tabular format from shortest to longest survival time (assuming that none of the data are censored). Then, time 0 is added to this table.
   - ➢ Now, we calculate number of deaths at each time point (m). Then, we look at the number of patients who survived to that time point plus the number of deaths at that exact time point (n).
   - ➢ Survival times are calculated by dividing n by the total number.

Patient ID	Time (y)
1	3
2	5
3	5
4	8
5	8
6	8
7	9
8	10
9	11
10	11

Time (y)	m	n	$S(t)$
0	0	10	10
3	1	9	0.9
5	2	7	0.7
8	3	4	0.4
9	1	3	0.3
10	1	2	0.2
11	2	0	0

   - ➢ Construct Kaplan–Meier survival curves to estimate the overall survival probability over time.
   - ➢ Plot survival curves to visualize the median survival time and the probability of survival at various time points.
   - ➢ If we have to include censored data, we add a column (q) and then the denominator for total number at each time point will be $(n - q)/n$.
2. Log-rank test:
   - ➢ Use the log-rank test to compare survival distributions between different subgroups (e.g., patients with different tumor stages or different treatment responses).
   - ➢ To assess the log rank test we calculate the long-rank statistic. Log-rank statistic is equivalent to the chi-square value. The critical chi-square value can be determined using the chi-square distribution (with the degrees of freedom being the number of groups minus 1, and choosing an appropriate α [usually α of 0.05]).
3. Cox proportional hazards model:
   - ➢ Perform a Cox proportional hazards regression analysis to identify factors that significantly affect survival.
   - ➢ Include covariates such as age, sex, tumor size, liver function, and response to TACE.

Results:

- Kaplan–Meier analysis:
  - ➢ The median overall survival time for the cohort is found to be 8 years.
  - ➢ The 3-, 5-, and 10-year survival rates are 90, 70, and 20%, respectively.

- Log-rank test:
  - ➢ Significant differences in survival are observed between patients with early-stage HCC (e.g., median survival of 36 months) and those with advanced-stage HCC (e.g., median survival of 12 months) ($p < 0.05$).
- Cox proportional hazards model:
  - ➢ Significant predictors of poorer survival include larger tumor size (e.g., HR = 1.5, $p < 0.01$), poor liver function (e.g., HR = 2.0, $p < 0.01$), and lack of response to TACE (e.g., HR = 1.8, $p < 0.05$).

Conclusion

Survival analysis reveals that TACE provides a significant survival benefit for patients with HCC, particularly those with early-stage disease. Key factors affecting survival include tumor size, liver function, and response to treatment. These findings can guide clinical decision-making and patient counseling.

The typical order for performing these survival analyses is as follows:

1. Kaplan–Meier survival analysis: Begin with Kaplan–Meier analysis to estimate and visualize the survival curves for different groups without adjusting for covariates. This step provides a basic understanding of survival probabilities over time and allows comparison of survival distributions across groups.
2. Log-rank test: Use the log-rank test to statistically compare the survival curves from the Kaplan–Meier analysis. This test evaluates whether there is a significant difference in survival times between groups (e.g., treatment vs. control) without accounting for additional covariates.
3. Cox regression modeling: Finally, perform Cox regression modeling to analyze the impact of multiple covariates on survival while adjusting for potential confounders. Cox regression provides HRs, allowing you to quantify the effect of each predictor on the risk of the event occurring.

Other survival analysis techniques exist, such as time parametric survival models, dependent covariate models, and Nelson–Aalen estimator, which have not been covered.

Case Study: Survival Analysis in Radiology

A clinical trial evaluates the effectiveness of a novel radiotherapy technique in prolonging survival in patients with glioblastoma. Kaplan–Meier survival curves illustrate differences in survival probabilities between treatment groups, while Cox regression identifies treatment efficacy after adjusting for relevant covariates such as tumor size and patient age.

Kaplan–Meier survival analysis is a nonparametric and univariate method, whereas the Cox proportional hazards model is a semiparametric and multivariate method.

There are curves other than the Kaplan–Meier curve, such as the receiver operating characteristic (ROC) curve, cumulative incidence function curve, Nelson–Aalen curve, calibration curves, and hazard function curves. Let us now understand about the importance of a ROC curve in classification and diagnostic modeling.

ROC Curve

ROC curve is a graphical representation of a binary classification model performance across all classification thresholds. It visualizes the tradeoff between sensitivity and specificity across different diagnostic test thresholds.[39–41]

- It is basically a plot of true positive rate (sensitivity) against false positive rate (1–specificity) across various thresholds (►**Fig. 6**).
- Area under the curve (AUC): Measures overall test performance. AUC = 1 indicates perfect discrimination, while AUC = 0.5 implies no better than chance. Larger the AUC, better is the classifier.
- A new classifier may be created using logistic regression, and a ROC curve can be created for the different threshold values in logistic regression.

Example:

Objective:

- To assess the diagnostic performance of a new biomarker in predicting successful reperfusion (measured by the modified Thrombolysis in Cerebral Infarction or mTICI score) after endovascular treatment in patients with acute ischemic stroke.

Data:

- Biomarker level: Level of the new serum biomarker (continuous variable).
- Successful reperfusion: Outcome of the treatment (0 for unsuccessful reperfusion, 1 for successful reperfusion, defined as mTICI 2b/3).

Steps to generate a ROC curve:

- Fit a logistic regression model: We first fit a logistic regression model to predict the probability of successful reperfusion based on the serum biomarker level.
- Calculate predicted probabilities: Use the fitted model to calculate the predicted probabilities of successful reperfusion.
- Generate the ROC curve: Plot the ROC curve by varying the threshold for predicting successful reperfusion and calculate the corresponding true positive rate (sensitivity) and false positive rate (1–specificity) at each threshold.
- Calculate the AUC: Compute the AUC as a measure of the biomarker's diagnostic performance.

Interpretation:

- The ROC curve plots the true positive rate (sensitivity) against the false positive rate (1–specificity) at various threshold settings. The AUC provides a single measure of the biomarker's diagnostic performance:
  - ➢ AUC = 1.0: Perfect test.
  - ➢ AUC = 0.5: No discriminative power, equivalent to random guessing.
  - ➢ 0.5 < AUC < 1.0: The test has some discriminative power, with higher values indicating better performance.

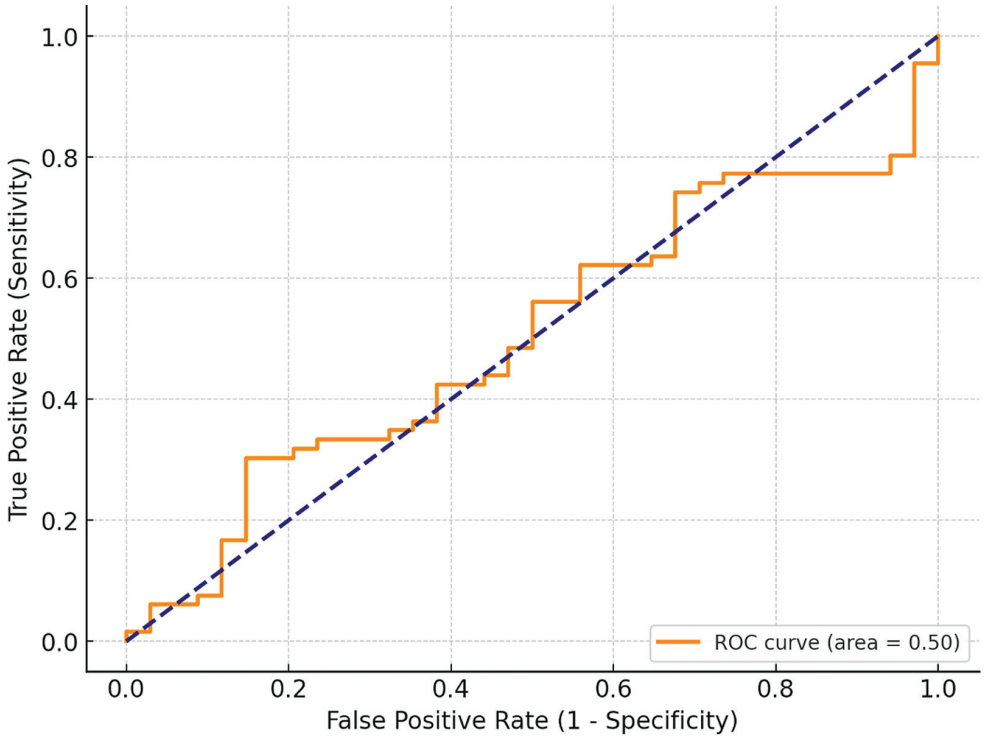

Fig. 6 Receiver operating characteristic (ROC) curve representing the diagnostic performance of a biomarker in predicting successful reperfusion (modified Thrombolysis in Cerebral Infarction [mTICI] score) after endovascular treatment in acute ischemic stroke patients. The area under the curve (AUC) provides a measure of the test's diagnostic performance.

Example output:

- ROC curve plot shows how well the biomarker discriminates between patients with and without successful reperfusion.
- The AUC value quantifies the overall ability of the biomarker to correctly classify patients regarding their reperfusion outcome.

Analysis of Diagnostic Tests

Bayesian theory offers a powerful framework for analyzing diagnostic tests by incorporating prior knowledge (pretest probabilities) along with new data (test results) to calculate posttest probabilities (►**Figs. 7** and **8**).[42]

Pretest probability = Likelihood ratio × Posttest probability

This basically means that if the pretest probability of a particular condition is 0, even if the test is 100% sensitive, posttest probability will be 0.

Contingency Tables

Contingency tables summarize the relationship between two categorical variables. It is essential for assessing diagnostic test performance. The frequency in the table is given in absolute or relative frequency.[43,44]

Example: Diagnostic test evaluation

Key metrics:

- *Sensitivity (Sn)*: TP/(TP + FN); proportion of true positives correctly identified.
- *Specificity (Sp)*: TN/(TN + FP); proportion of true negatives correctly identified.
- *Positive predictive value*: TP/(TP + FP); probability that a positive test result is correct.
- *Negative predictive value*: TN/ (TN + FN); probability that a negative test result is correct.
- *Positive likelihood ratio (LR+)*: Sn/(1–Sp); ratio of the probability of a positive test result in people with the disease to the probability of a positive test result in people without the disease.
- *Negative likelihood ratio (LR–)*: 1–Sn/Sp; ratio of the probability of a negative test result in people with the disease to the probability of a negative test result in people without the disease.
- *Accuracy*: (TP + TN)/(TP + TN + FP + FN); proportion of all test results that are correct (both true positives and true negatives).
- *AUC*: Measure of the overall performance of a diagnostic test. An AUC of 1 represents a perfect test, while an AUC of 0.5 represents a worthless test.

	Disease present	Disease absent	Total
Test positive	True positive (TP)	False positive (FP)	TP + FP
Test negative	False negative (FN)	True negative (TN)	FN + TN
Total	TP + FN	FP + TN	N

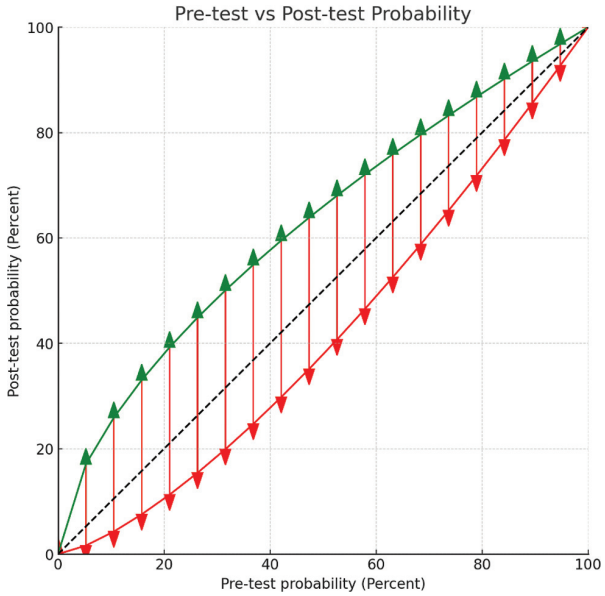

Fig. 7 Graph illustrating the relationship between pretest and posttest probabilities. Green curve (upper left) represents a positive test result, while the red curve (lower right) represents a negative test result. The length of the green arrows indicates the change in absolute probability after a positive test, and the red arrows represent the change in absolute probability after a negative test. The diagram shows that at low pretest probabilities, a positive test results in a greater absolute change in probability than a negative test, which generally holds true as long as the specificity is not significantly higher than the sensitivity. Conversely, at high pretest probabilities, a negative test causes a greater absolute change in probability than a positive test, a property that also holds when the sensitivity is not much higher than the specificity.

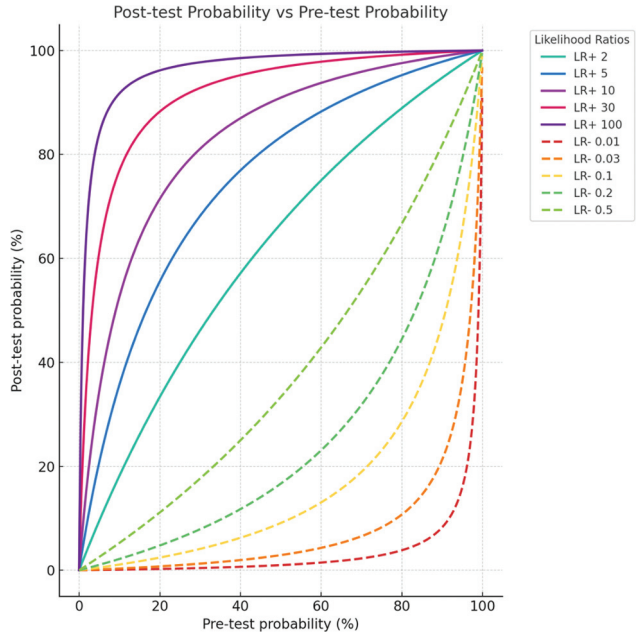

Fig. 8 Graph illustrating the relationship between pretest probability (*x*-axis) and posttest probability (*y*-axis) for various positive and negative likelihood ratios (LR). It is used to determine how diagnostic test results (either positive or negative) affect the probability of a condition being present. The solid lines represent positive likelihood ratios (LR+), which indicate how much a positive test result increases the probability of a condition being present. Higher positive likelihood ratios lead to greater increases in posttest probability. The dashed lines represent negative likelihood ratios (LR−), which indicate how much a negative test result decreases the probability of a condition being present. Lower negative likelihood ratios lead to greater reductions in posttest probability.

- *Diagnostic OR*: LR+/LR−; ratio of the odds of the test being positive if the subject has the disease relative to the odds of the test being positive if the subject does not have the disease.

Other Measures of Reliability of a Diagnostic Test

Cohen's Kappa

Cohen's kappa is a measure of agreement between two dependent categorical samples, and is used whenever one wants to know if there is an agreement between two raters. Variable measured by the two raters should be a nominal variable.[45,46]

If it is an ordinal variable and two or more raters, Kendall's tau can be used. If it is a metric variable and more than two raters, interclass correlation can be used.

It is important to understand that the Cohen's kappa coefficient can only tell how reliably both raters are measuring the same thing (i.e., reliability). It does not tell what both the two raters are measuring is the right thing (i.e., validity).

The table of Landis and Koch (1977) can be used as a guide to interpret Cohen's kappa.[47]

Kappa	
>0.8	Almost perfect
>0.6	Substantial

(*Continued*)

(*Continued*)

Kappa	
>0.4	Moderate
>0.2	Fair
0–0.2	Slight
<0	Poor

Example: To assess the interrater reliability between two radiologists in lung nodules from chest radiographs.

Two radiologists independently review a set of 100 chest X-ray images to determine the presence or absence of lung nodules. The results are recorded in a contingency table:

	Radiologist 2: Nodule present	Radiologist 2: Nodule absent	Total
Radiologist 1: Nodule present	30	10	40
Radiologist 1: Nodule absent	20	40	60
Total	50	50	100

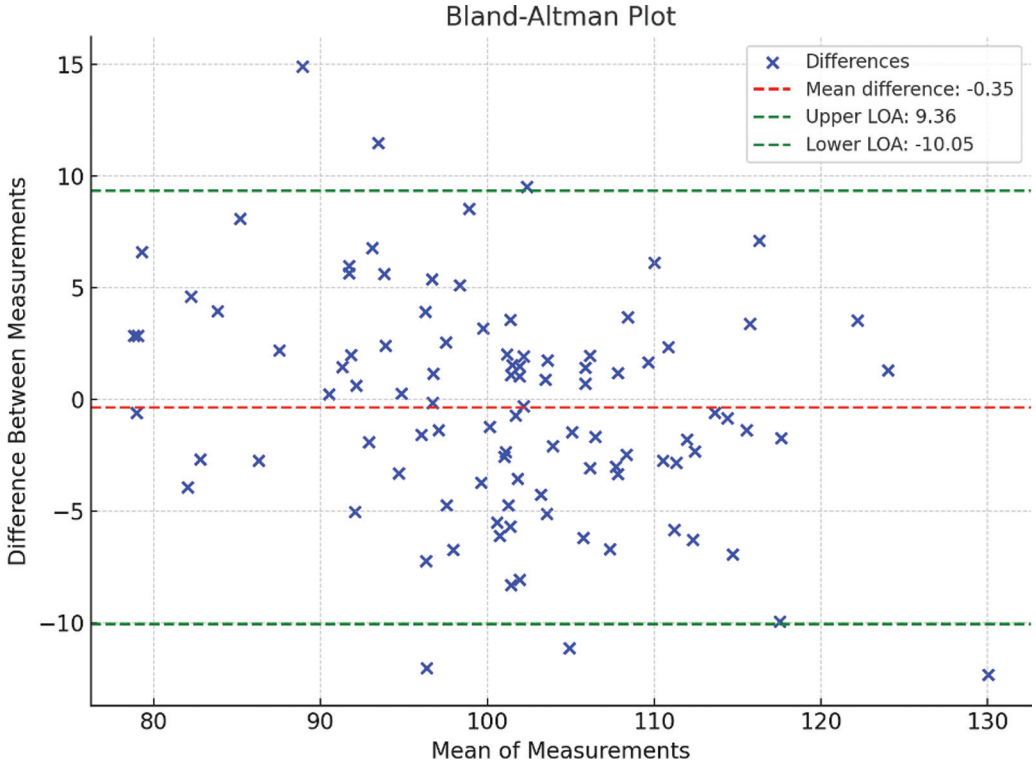

Fig. 9 Graph representing a Bland–Altman plot used to compare two measurement methods and assess the agreement between them; *X*-axis represents the average of two measurements for each subject; *Y*-axis represents the difference between two measurements for each subject. Most points should lie within the limits of agreement [LOA] (green dashed lines), which is a sign of good agreement between the two methods. Mean difference (red line) indicates whether there is a systematic difference (or bias) between the methods. In this case, the bias is small (−0.35), implying the two methods are generally similar, though with some variability. Upper limit of agreement (LOA): Calculated as the mean difference plus 1.96 times the standard deviation (SD) of the differences; Lower limit of agreement (LOA): Calculated as the mean difference minus 1.96 times the SD of the differences. These limits indicate the range within which 95% of the differences between the two methods are expected to fall.

Calculation of Cohen's kappa:

Cohen's kappa (κ) is used to measure the agreement between the two radiologists, correcting for the agreement that would occur by chance.

1. *Observed agreement (P_o)*: Proportion of instances where the radiologists agree.
   $P_o = (30 + 40)/100 = 0.7$
2. *Expected agreement (P_e)*: Proportion of agreement expected by chance.
   $P_e = ([40 \times 50]/100) + ([60 \times 50]/100) = (0.2) + (0.3) = 0.5$
3. Cohen's kappa (κ):
   $\kappa = (P_o - P_e)/(1 - P_e) = (0.7 - 0.5)/(1 - 0.5) = 0.4$

Interpretation:

- A κ value of 0.40 indicates moderate agreement between the two radiologists. While there is some level of agreement beyond what would be expected by chance, the level of agreement is not very high.
- This suggests that there might be a need for further training or more standardized diagnostic criteria to improve consistency between radiologists.

SE of Cohen's kappa is a measure of the precision of the estimated value. Smaller SE means more precise is the estimate and larger SE indicates less precision.

Weighted Cohen's Kappa

In case of an ordinal variable, that is, a variable with a ranking, it is important that the gradations are also considered. A difference between "very good" and "average" is greater than between "very good" and "good." This deviation is also included in the calculation of weighted Cohen's kappa. The differences can be weighted linearly or quadratically.[48,49]

Fleiss Kappa

Fleiss kappa is a measure of agreement between three or more raters for a nominal variable.[50]

Kendall's Tau

Kendall's tau is a nonparametric measure of ordinal association and is used to assess the strength of relationship between two or more ordinal variables. It ranges from 0 to 1 with values close to 1 indicating a strong association, and those close to 0 indicating a weak or no association.[51]

Intraclass Correlation

Intraclass correlation (ICC) is a statistical measure that quantifies the degree of consistency among observations that are made on the same individuals or objects. ICC is used to assess the reliability and consistency of measurements taken by different raters, assessors, or instruments.

ICC is particularly useful when evaluating the reliability of a diagnostic test across multiple raters (interrater reliability) or repeated measurements (intrarater reliability).[52–54]

Types of ICC:

1. Single measure ICC: Single measure on each subject.
2. Two-way random effect ICC: Multiple measurements on each subject, and there are multiple raters.
3. Two-way mixed effect ICC: Multiple measurements on each subject, and there are multiple raters as well as multiple subjects.

ICC values can range from 0 to 1, with 1 indicating perfect consistency among observations, and 0 indicating no consistency. In general, ICC values above 0.75 are considered to be good, and those above 0.9 are considered to be excellent.

One important point to keep in mind while using such measures of agreement is that both the raters should not have the same level of experience. One of the raters should be a novice (with lesser years of experience) and the other an expert (with larger number of years of experience). This helps to ensure the reproducibility of findings when applied to the population.

Bland–Altman Plot

Bland–Altman plot (also known as a difference plot) is used to assess the agreement between two methods of measurement or between repeated measurements from the same test (continuous variables). It provides a graphical representation of the differences between two sets of measurements (►**Fig. 9**). They are often used to compare a new diagnostic test with a gold standard test or to evaluate the repeatability of measurements from the same test. The plot shows the mean difference between the two measurement methods (bias) and the limits of agreement (typically $\pm$ 1.96 standard deviations from the mean difference). If the differences between the two methods are close to zero and fall within the limits of agreement, the two methods are said to be in good agreement and thus reliable. Large biases or wide limits of agreement suggest poor agreement, indicating that the test may not be reliable. For example, comparing a new blood glucose monitoring device with an established standard can be done using a Bland–Altman plot to visually assess if the two methods agree sufficiently.[55]

Conclusion

Advanced statistical techniques are indispensable for radiologists aiming to extract meaningful insights from complex data, enhance diagnostic accuracy, and drive evidence-based practice. By mastering correlation and causality, regression analysis, survival curves, and diagnostic test evaluation, radiologists can navigate the complexities of clinical and research environments effectively. This comprehensive guide equips radiologists with the tools needed to conduct rigorous analyses, interpret findings accurately, and contribute to advancing medical knowledge and patient care in the field of radiology.

Note
Work done in: Department of Imaging Sciences and Interventional Radiology, Sree Chitra Institute of Medical Sciences, Trivandrum

Authors' Contributions
All the authors were involved in the procedure, data collection, and manuscript revision.

Funding
None.

Conflict of Interest
None declared.

Acknowledgments
None.

References

1 Bewick V, Cheek L, Ball J. Statistics review 7: correlation and regression. Crit Care 2003;7(06):451–459
2 Zaniletti I, Larson DR, Lewallen DG, Berry DJ, Maradit Kremers H. How to distinguish correlation from causation in orthopaedic research. J Arthroplasty 2023;38(04):634–637
3 Mukaka MM. Statistics corner: a guide to appropriate use of correlation coefficient in medical research. Malawi Med J 2012;24(03):69–71
4 Schober P, Boer C, Schwarte LA. Correlation coefficients: appropriate use and interpretation. Anesth Analg 2018;126 (05):1763–1768
5 de Winter JC, Gosling SD, Potter J. Comparing the Pearson and Spearman correlation coefficients across distributions and sample sizes: a tutorial using simulations and empirical data. Psychol Methods 2016;21(03):273–290
6 Rovetta A. Raiders of the lost correlation: a guide on using Pearson and Spearman coefficients to detect hidden correlations in medical sciences. Cureus 2020;12(11):e11794
7 Chen S, Ghadami A, Epureanu BI. Practical guide to using Kendall's τ in the context of forecasting critical transitions. R Soc Open Sci 2022;9(07):211346
8 Bonett DG. Point-biserial correlation: interval estimation, hypothesis testing, meta-analysis, and sample size determination. Br J Math Stat Psychol 2020;73(Suppl 1):113–144
9 Shimonovich M, Pearce A, Thomson H, Keyes K, Katikireddi SV. Assessing causality in epidemiology: revisiting Bradford Hill to incorporate developments in causal thinking. Eur J Epidemiol 2021;36(09):873–887
10 Provost LP. Commentary: establishing causality in quality improvement studies. Pediatr Qual Saf 2023;8(03):e653
11 Ali P, Younas A. Understanding and interpreting regression analysis. Evid Based Nurs 2021;24(04):116–118
12 Streiner DL. Control or overcontrol for covariates? Evid Based Ment Health 2016;19(01):4–5
13 Hazra A, Gogtay N. Biostatistics series module 10: brief overview of multivariate methods. Indian J Dermatol 2017; 62(04):358–366
14 Kim HY. Statistical notes for clinical researchers: simple linear regression 1 - basic concepts. Restor Dent Endod 2018;43(02): e21

15 Casson RJ, Farmer LDM. Understanding and checking the assumptions of linear regression: a primer for medical researchers. Clin Exp Ophthalmol 2014;42(06):590–596
16 Riley RD, Snell KI, Ensor J, et al. Minimum sample size for developing a multivariable prediction model: part II - binary and time-to-event outcomes. Stat Med 2019;38(07):1276–1296
17 Marill KA. Advanced statistics: linear regression, part II: multiple linear regression. Acad Emerg Med 2004;11(01):94–102
18 Alexopoulos EC. Introduction to multivariate regression analysis. Hippokratia 2010;14(Suppl 1):23–28
19 Sperandei S. Understanding logistic regression analysis. Biochem Med (Zagreb) 2014;24(01):12–18
20 Stoltzfus JC. Logistic regression: a brief primer. Acad Emerg Med 2011;18(10):1099–1104
21 Schober P, Vetter TR. Logistic regression in medical research. Anesth Analg 2021;132(02):365–366
22 van Ginkel JR. Standardized regression coefficients and newly proposed estimators for [formula: see text] in multiply imputed data. Psychometrika 2020;85(01):185–205
23 Yang K, Tu J, Chen T. Homoscedasticity: an overlooked critical assumption for linear regression. Gen Psychiatr 2019;32(05): e100148
24 Vatcheva KP, Lee M, McCormick JB, Rahbar MH. Multicollinearity in regression analyses conducted in epidemiologic studies. Epidemiology (Sunnyvale) 2016;6(02):227
25 Abd ElHafeez S, D'Arrigo G, Leonardis D, Fusaro M, Tripepi G, Roumeliotis S. Methods to analyze time-to-event data: the Cox regression analysis. Oxid Med Cell Longev 2021;2021; 1302811
26 Andrade C. Survival analysis, Kaplan-Meier curves, and Cox regression: basic concepts. Indian J Psychol Med 2023;45(04): 434–435
27 Deo SV, Deo V, Sundaram V. Survival analysis-part 2: Cox proportional hazards model. Indian J Thorac Cardiovasc Surg 2021;37(02):229–233
28 Zeng Z, Gao Y, Li J, et al. Violations of proportional hazard assumption in Cox regression model of transcriptomic data in TCGA pan-cancer cohorts. Comput Struct Biotechnol J 2022; 20:496–507
29 Bellera CA, MacGrogan G, Debled M, de Lara CT, Brouste V, Mathoulin-Pélissier S. Variables with time-varying effects and the Cox model: some statistical concepts illustrated with a prognostic factor study in breast cancer. BMC Med Res Methodol 2010;10:20
30 Stolberg HO, Norman G, Trop I. Survival analysis. AJR Am J Roentgenol 2005;185(01):19–22
31 Clark TG, Bradburn MJ, Love SB, Altman DG. Survival analysis part I: basic concepts and first analyses. Br J Cancer 2003;89 (02):232–238
32 Park SH, Han K, Park SY. Mistakes to avoid for accurate and transparent reporting of survival analysis in imaging research. Korean J Radiol 2021;22(10):1587–1593
33 Rai S, Mishra P, Ghoshal UC. Survival analysis: a primer for the clinician scientists. Indian J Gastroenterol 2021;40(05): 541–549
34 Leung KM, Elashoff RM, Afifi AA. Censoring issues in survival analysis. Annu Rev Public Health 1997;18:83–104
35 Rich JT, Neely JG, Paniello RC, Voelker CCJ, Nussenbaum B, Wang EW. A practical guide to understanding Kaplan-Meier curves. Otolaryngol Head Neck Surg 2010;143(03): 331–336
36 Dudley WN, Wickham R, Coombs N. An introduction to survival statistics: Kaplan-Meier analysis. J Adv Pract Oncol 2016;7(01):91–100
37 Bland JM, Altman DG. The logrank test. BMJ 2004;328(7447): 1073
38 Dormuth I, Liu T, Xu J, Yu M, Pauly M, Ditzhaus M. Which test for crossing survival curves? A user's guideline. BMC Med Res Methodol 2022;22(01):34
39 Hajian-Tilaki K. Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. Caspian J Intern Med 2013;4(02):627–635
40 Nahm FS. Receiver operating characteristic curve: overview and practical use for clinicians. Korean J Anesthesiol 2022;75 (01):25–36
41 Florkowski CM. Sensitivity, specificity, receiver-operating characteristic (ROC) curves and likelihood ratios: communicating the performance of diagnostic tests. Clin Biochem Rev 2008;29(Suppl 1, Suppl 1)S83–S87
42 Safari S, Baratloo A, Elfil M, Negida A. Evidence based emergency medicine; part 4: pre-test and post-test probabilities and Fagan's nomogram. Emergency (Tehran) 2016;4(01): 48–51
43 Monaghan TF, Rahman SN, Agudelo CW, et al. Foundational statistical principles in medical research: sensitivity, specificity, positive predictive value, and negative predictive value. Medicina (Kaunas) 2021;57(05):503
44 Parikh R, Mathai A, Parikh S, Chandra Sekhar G, Thomas R. Understanding and using sensitivity, specificity and predictive values. Indian J Ophthalmol 2008;56(01):45–50
45 Tang W, Hu J, Zhang H, Wu P, He H. Kappa coefficient: a popular measure of rater agreement. Shanghai Jingshen Yixue 2015;27 (01):62–67
46 Li M, Gao Q, Yu T. Kappa statistic considerations in evaluating inter-rater reliability between two raters: which, when and context matters. BMC Cancer 2023;23(01):799
47 Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33(01):159–174
48 Warrens MJ. Cohen's linearly weighted kappa is a weighted average. Adv Data Anal Classif 2012;6:67–79
49 Ranganathan P, Pramesh CS, Aggarwal R. Common pitfalls in statistical analysis: measures of agreement. Perspect Clin Res 2017;8(04):187–191
50 Gwet KL. Large-sample variance of Fleiss generalized kappa. Educ Psychol Meas 2021;81(04):781–790
51 Ma Y. On inference for Kendall's τ within a longitudinal data setting. J Appl Stat 2012;39(01):2441–2452
52 Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 2016;15(02):155–163
53 Liu J, Tang W, Chen G, Lu Y, Feng C, Tu XM. Correlation and agreement: overview and clarification of competing concepts and measures. Shanghai Jingshen Yixue 2016;28(02):115–120
54 Liljequist D, Elfving B, Skavberg Roaldsen K. Intraclass correlation - a discussion and demonstration of basic features. PLoS One 2019;14(07):e0219854
55 Riffenburgh RH, Gillen DL. Techniques to aid analysis. In Statistics in Medicine. 4th Ed.2020:631–649

