Optometry & Contact Lenses · Vol. 5, No. 1, pp. 16-37, 2025
https://doi.org/10.54352/dozv.JEKY2435
Published online in English as Open Access: 30.12.2024

Advancing Statistical Literacy in Eye Care: A Series for Enhanced Clinical Decision-Making

Part 1: Introduction to Statistical Tools for Eye Care Research

Daniela Oehring[1,2], Pedro Miguel Serra[1,3]

[1] Ph.D. • [2] Faculty of Health, University of Plymouth, Plymouth, UK • [3] Ophthalmology Clinic Vista Sanchez Trancon, Spain

Received 31 October 2024; accepted 28 November 2024

Abstract

Purpose. Advancements in eye and vision care hinge on the rigorous application of research and the precise interpretation of clinical data. However, the field of Eye and Vision Research (EVR) frequently encounters research waste attributed to methodological flaws and improper statistical analyses, undermining the validity of studies and inefficiently utilising substantial financial resources. This paper, the first instalment in the series "Advancing Statistical Literacy in Eye Care: A Series for Enhanced Clinical Decision-Making," aims to address these challenges by enhancing the statistical literacy of eye care professionals.

Material and Methods. Through a comprehensive narrative literature review and the generation of simulated clinical datasets, this study identifies essential statistical concepts, common pitfalls, and best practices pertinent to EVR. The literature review used multiple databases, including PubMed, Scopus, and Web of Science, focusing on peer-reviewed articles and professional textbooks relevant to statistical methodologies. Simulated datasets reflecting realistic clinical measurements, such as pupil diameter, refractive error, central corneal thickness, and intraocular pressure, were created using Python to illustrate key statistical principles and their applications.

Results. The paper explores fundamental statistical concepts, including data types (nominal, ordinal, metric), data preparation techniques, handling missing data and outliers, and applying descriptive statistics. Additionally, it explores data distribution characteristics, normality assessment, and data transformation methods to ensure robust and reliable statistical analyses. By bridging theoretical knowledge with practical examples, this instalment seeks to equip eye care professionals with the tools to critically evaluate research, integrate evidence-based practices, and contribute meaningfully to the scientific community.

Conclusion. This study establishes a foundational framework to enhance statistical literacy among eye care professionals by exploring essential statistical concepts and best practices in EVR. By addressing common methodological flaws and improper analyses, it aims to reduce research waste and improve the validity of studies. Ultimately, this initiative is expected to promote more accurate data interpretation, better clinical decision-making, and improved patient care in the field of eye and vision health.

Keywords

Statistical Literacy, Eye Care, Clinical Decision-Making, Eye and Vision Research, Descriptive Statistics, Data Analysis, Research Methodology

Introduction

Advancements in eye and vision care depend on rigorous research and precise clinical data interpretation. As eye care professionals navigate an ever-expanding body of scientific literature, understanding and applying appropriate statistical methods cannot be overstated. Statistical literacy is essential for critically evaluating research, integrating evidence-based practices, and contributing to the scientific community.

Evidence-based practice is essential to eye and vision care, guiding diagnosis, treatment, and overall patient care through high-quality research supported by sound statistical methods[1-5] However, the field faces significant challenges due to studies often plagued by methodological flaws, and improper statistical analysis, which undermine research validity and contribute to unnecessary use of resources on studies that fail to yield reliable findings (research waste).[3-5]

Research waste in ophthalmology is a severe issue, leading to inefficiencies and resource waste.[5] For instance, only 22.4 % of phase III ophthalmology trials cite systematic reviews as a justification for the study, missing opportunities to build on existing evidence and resulting in redundant research.[6] In surgical studies, the problem is exacerbated by poor methodological rigour, often resulting in incomplete or invalid findings. Estimates suggest that up to 85 % of global health research, representing a substantial portion of the $ 200 billion annual expenditure, may be wasted due to non-publication, unclear reporting and lack of systematic review use in study design.[7-9] In EVR, this waste commonly stems from inadequate sample sizes, inappropriate statistical tests, or misinterpreting results.[7-9] A review of ophthalmic literature found a substantial portion of published studies contained statistical inaccuracies, further eroding research credibility and slowing the progression of evidence-based practice.[8-10]

Improving statistical literacy among researchers and clinicians is crucial to addressing these issues. A solid understanding of statistical principles and correct application can enhance research quality, reduce waste, and support evidence-based patient care through reliable and valid findings.

Methods

This article is the first instalment in a five-part series designed to enhance statistical literacy among eye care professionals. To accomplish this, a two-pronged methodological approach was employed: a comprehensive literature review and the creation of simulated datasets to exemplify key statistical concepts pertinent to Eye and Vision Research (EVR).

A narrative literature review was conducted to identify essential statistical concepts, common pitfalls, and best practices in applying statistical methods within EVR. The literature search was performed across multiple databases, including PubMed, Scopus, and Web of Science, covering publications up to October 2024.

Keywords and Search Terms: The search utilised combinations of terms such as “vision science“, “statistical methods“, “data analysis“, “clinical research“, “biostatistics“, “eye care“, “research methodology“, “data distribution“, “missing data“, “outliers“, and “statistical literacy“.

Inclusion Criteria: We included peer-reviewed articles and professional textbooks that focused on statistical methods applicable to ophthalmology and optometry, studies highlighting common statistical errors in EVR, guidelines on best practices for data analysis in clinical research, and articles emphasising the importance of statistical literacy in eye care.

Exclusion Criteria: Articles not directly related to statistical methods in eye care, non-English publications, conference abstracts without full texts, and studies lacking methodological details were excluded.

Data Extraction and Synthesis: Relevant articles were initially selected based on title and abstract screening, followed by full-text reviews. Key information extracted included statistical concepts discussed, common errors identified, recommendations for best practices, and implications for clinical decision-making. The findings were synthesised thematically to provide a comprehensive overview of fundamental statistical concepts, data types, data preparation, and descriptive statistics relevant to eye care professionals.

Simulated Data Generation: To provide practical examples and visual illustrations of the statistical concepts discussed, we generated simulated datasets using Python (V3.12.4). The datasets were crafted to reflect realistic clinical data commonly encountered in EVR, focusing on pupil diameter (n = 1000), refractive error (n = 1000), and central corneal thickness and intraocular pressure (CCT, IOP, n = 1000). The data were modelled to follow a given distribution with parameters (mean and standard deviation) reflective of typical clinical observations (pupil size,[12-17] refractive error,[18-24] CCT and IOP[25,26,27,28,29,30]).

Data Generation Process: Data generation and analysis were performed using Python and libraries such as NumPy (V2.0.1)[31] for numerical computations, pandas[32] for data manipulation, and matplotlib[33] and seaborn[34] for data visualisation. The simulated datasets were generated using statistical models appropriate for each variable. For example, normal distributions were used for pupil diameter, while refractive error data were modelled to exhibit skewness akin to real-world clinical data.

The simulated datasets are available through the Open Science Framework (OSF) at DOI: 10.17605/OSF.IO/6EYMW,[35] ensuring transparency and allowing readers to replicate the analyses presented.

As the study involved simulated data and a review of existing literature, no ethical approval was required. The simulated data do not represent real patient information, eliminating patient confidentiality and privacy concerns.

By combining a thorough literature review with the creation of simulated clinical datasets, the article bridges theoretical knowledge and practical application. This integrative

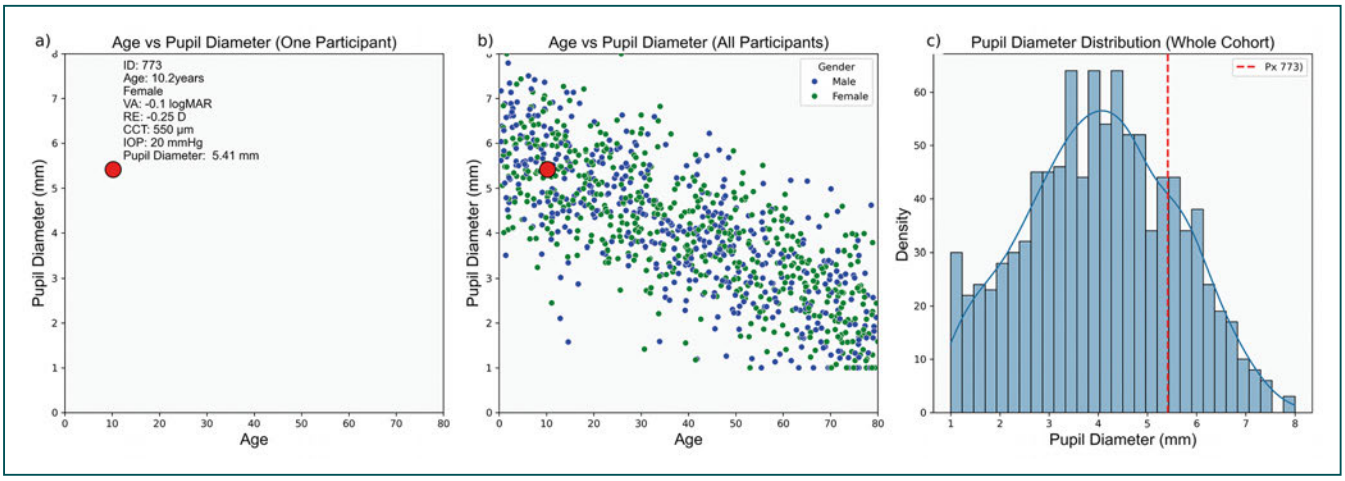

Figure 1: Visualisation of age versus pupil diameter for (a) one participant and (b)all participants from a simulated cohort. Gender differences are depicted, providing insights into how pupil diameter distribution varies with age. (c) A histogram of pupil diameter distribution across a simulated cohort. The histogram represents pupil diameters for 1,000 individuals, with the horizontal axis showing size in millimetres and the vertical axis indicating the number of individuals. Each bar corresponds to a 0.25 mm diameter bin. A vertical dashed line marks an individual's pupil size, while a thin blue line outlines the overall distribution shape.

approach is intended to enhance understanding and facilitate the application of statistical tools in clinical decision-making in eye care.

Fundamental statistical concepts

Interpreting cohort findings begins with collecting individual data points from a sample representing the target population. Each data point, such as pupil diameter (Figure 1), provides specific information about a single subject. When aggregated, these data points create a dataset that reflects the characteristics and variability of the broader population.

Biological measurements inherently carry uncertainty and variability due to individual biological differences, environmental influences, and measurement errors. Uncertainty reflects confidence in measurement accuracy and conclusions, while variability pertains to natural data fluctuations. Thus, data collection alone is insufficient. Statistics provides the tools to quantify uncertainty and assess variability, enabling researchers to distinguish between random fluctuations and meaningful patterns. By applying statistical techniques, researchers can make informed inferences about the population, assess the reliability of the findings and estimate the likelihood that observed effects are due to chance.

Data is not monolithic, varying by the nature of what is measured and the measurement technique used. Each data type requires specific analytical methods. In EVR, data is primarily empirical and derived from observation and experimentation. It can be categorised as nominal (e. g., type of cataract), ordinal (e. g., visual acuity), and metric (e. g., axial length), Table 1. Additionally, data can be defined by the values a variable can assume; continuous data spans a range limited only by measurement precision, while discrete data is restricted to specific, separate values, typically representing counts or whole numbers.

Data Preparation

Preparing the data for analysis is essential for accurate results and involves exploration, transformation and validation. Data exploration allows the researcher to understand how the data will be used and determine how to clean, structure and organise it. Transformation includes structuring the data, organising it relationally, and normalising it by removing redundancies. Cleaning addresses irregularities, including missing values, inaccuracies and outliers.[36]

Implementing validation rules, such as range and consistency checks, during data entry reduces error rates. Regular audits identify anomalies, and duplicate detection prevents double-counting of data points. Effective data management, including Standard Operating Procedures (SOPs) for data collection, entry, and storage, ensures consistency and reliability. Training personnel and using Electronic Data Capture (EDC) systems with validation tools further improve data quality. Additionally, statistical quality control measures, such as control charts and process capability analysis, monitor data collection over time, detect shifts or trends that may signal quality issues and ensure adherence to quality standards.[36,37]

Missing Data

Missing data is a common research issue that can introduce bias, reduce statistical power, and compromise study generalisability. In clinical research, missing data may result from patient dropouts, non-compliance, technical issues during data collection, or data entry errors. Addressing missing data requires understanding the underlying mechanisms and applying suitable statistical methods.[38,39] Missing data falls into three categories: Missing Completely At Random (MCAR), Missing At Random (MAR) and Missing Not At Random (MNAR),[40] Table 2. The choice of method depends

Table 1: Classification of Data Types and Their Applications in Ophthalmic Research

Type/Scale	Description	Examples/Application
Nominal Data	Characterised by discrete categories or labels with no inherent order or ranking, nominal data often captures qualitative distinctions crucial to clinical practice. **Key Characteristics:** **No Order:** Categories are simply names or identifiers; there is no intrinsic ranking or hierarchy. **No Distance:** The „distance" between categories is undefined and carries no quantitative meaning.	Eye colour (blue, brown, green); Type of cataract (nuclear, cortical, posterior subcapsular); Presence/absence of a specific ocular condition **Statistical Approach:** Analysis often focuses on counts, frequencies, proportions, and associations between nominal variables and other data types.
Ordinal Data	Ordinal data closes the gap between qualitative and quantitative information by representing categories with a natural order or ranking. **Key Characteristics:** **Ordered Categories:** Categories have a clear sequence or progression, but the intervals between them may not be equal. **Unequal Distances:** The difference between categories is not necessarily consistent or quantifiable.	Visual acuity scores (20/20, 20/40, 20/200); Severity scales for dry eye disease (mild, moderate, severe); Grading scales, Likert scales **Statistical Approach:** Median, percentiles, and non-parametric tests are commonly used for ordinal data analysis.
Metric Data	Encompassing numerical measurements with meaningful and quantifiable intervals, metric data underpins a vast array of ophthalmic research. **Key Characteristics:** **Meaningful Intervals:** The difference between any two values on the scale is consistent and carries quantitative significance. **Further Classification:** Metric data is further divided into: 1. **Interval Data:** Possesses a consistent scale but lacks a true zero point. 2. **Ratio Data:** Features a consistent scale and a true zero point, representing the absence of the measured attribute. This allows for meaningful ratios and a wider range of statistical analyses.	**Interval Data:** Temperature in Celsius or Fahrenheit, and some psychometric scales used in vision research such as, Decibels in visual field testing). Visual acuity in LogMAR. **Ratio Data:** Intraocular pressure, axial length, and corneal thickness. **Statistical Approach:** Metric data opens the door to a broad spectrum of statistical techniques, including means, standard deviations, correlations, and parametric tests.

on the type of missingness and study context, Table A1. Figure 2 presents a step-by-step approach for addressing missing data.

Outliers

Outliers are data points that deviate significantly from the dataset and may arise from measurement errors, data entry mistakes, or true extreme values due to biological variability or rare conditions. Identifying and appropriately managing outliers is essential to maintaining data integrity and valid statistical analysis.[45,46] Table 3 summarises various methods for outlier detection along with their applications and limitations.

After detecting outliers, they should be carefully examined and managed. The first step is verification, involving checks for data entry errors or measurement inaccuracies, such as typographical mistakes or instrument calibration issues. Determining if the outlier is clinically plausible or represents an extreme physiological condition is also important.[45,46]

Researchers have several options for handling outliers. Retaining them is appropriate if they represent valid observations, though this may increase variability and affect statistical robustness. Conversely, excluding outliers is justified if they are errors or do not represent the studied population, with predefined exclusion criteria to prevent bias. Alternatively, data transformations (e.g., logarithmic) can reduce the influence of outliers without removal, and robust methods, such

Table 2: Types of Missing Data Mechanisms and Their Implications in Clinical Research[40]

Type	Definition	Mathematical Representation	Implications	Clinical Example
Missing Completely at Random (MCAR)	The probability of missingness is independent of both observed and unobserved data.	$P(M = 1 \mid X,Y) = P(M = 1)$ M is the missingness indicator, meaning M = 1 if data is missing and M = 0 if data is present. X represents the observed data. Y represents the unobserved or missing data.	Data are a random subsample of the original data. Analyses remain unbiased when using only complete cases. This assumption allows for unbiased statistical estimates, though it is generally challenging to satisfy in real-world data.	Missing visual acuity measurements due to random equipment malfunction affecting all patients equally.
Missing at Random (MAR)	The probability of missingness depends only on the observed data and not on the unobserved data	$P(M = 1 \mid X,Y) = P(M = 1 \mid X)$	Missingness can be accounted for by conditioning on observed variables. This assumption is less restrictive than MCAR and is often more applicable in real-world data, allowing researchers to make valid inferences using appropriate statistical methods, such as multiple imputation or maximum likelihood, without introducing bias from the missing data pattern.	Patients with severe diabetic retinopathy are less likely to return for follow-up visits, and their severity is recorded in previous visits.[41,42]
Missing Not at Random (MNAR)	The probability of missingness depends on unobserved data itself (most challenging scenario to address).	$P(M = 1 \mid X,Y) = P(M = 1 \mid Y)$	Missingness is related to the missing values themselves. In MNAR situations, standard methods like multiple imputation or maximum likelihood are generally insufficient on their own because they rely on the assumption that missingness depends on observed data. Addressing MNAR data often requires additional modelling or sensitivity analysis and, in some cases, external data or assumptions to make the data analysis valid.	Patients experiencing severe side effects (unrecorded because they dropped out) are more likely to discontinue participation in a clinical trial.[43,44]

as least absolute deviations (LAD) or M-estimators, minimise the influence of extreme values on the analysis.[48]

Outliers can significantly impact clinical research and practice. They may sometimes represent novel findings, such as rare clinical presentations that warrant further investigation. For instance, an unusually early onset of age-related macular degeneration (AMD) could reveal insights into atypical disease progression or unique risk factors. However, outliers can also distort statistical analyses, leading to erroneous conclusions. Extreme values can skew parameter estimates, compromising study validity. This is particularly relevant in risk stratification, where identifying outliers can help identify high-risk patients who may need special interventions. In clinical settings, outliers often underscore important considerations. For example, in visual field testing, a subset of glaucoma patients with unusually rapid progression may indicate non-compliance with the treatment or a more aggressive disease variant requiring closer monitoring.[49-52] Likewise, an unusually long axial length in biometric measurements may suggest pathological myopia, prompting further investigation or treatment adjustments.[53,54]

Descriptive Statistics

Descriptive statistics are fundamental tools that summarise and organise data without making inferences or predictions, providing a snapshot of a dataset's key features. In EVR, where

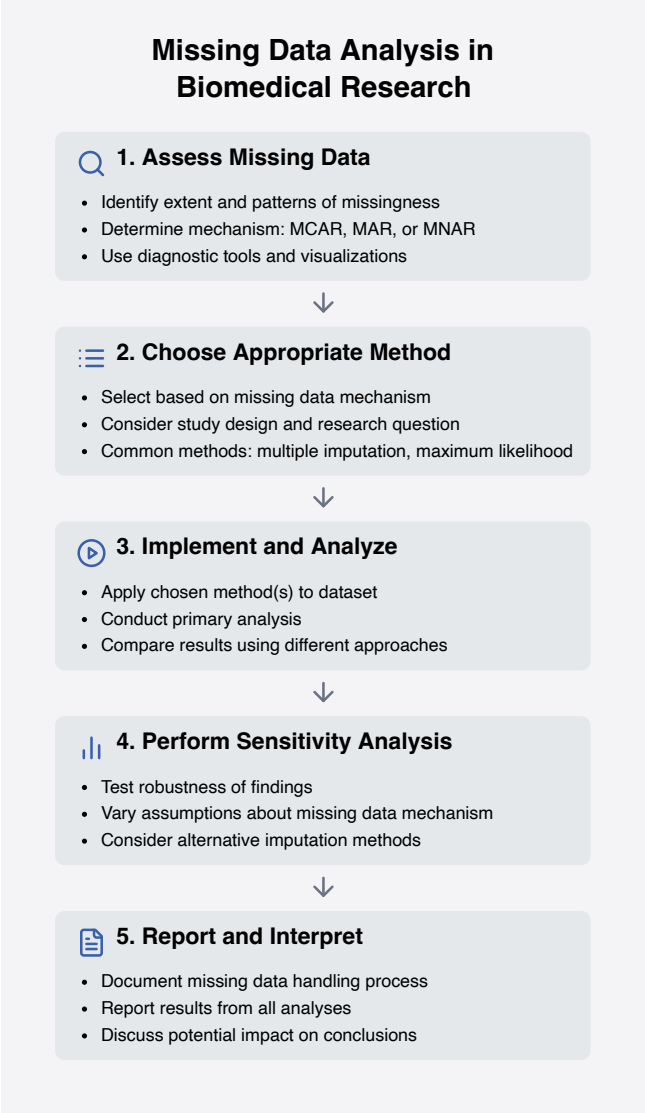

Figure 2: Step-by-step approach for handling missing data in biomedical research (© Daniela Oehring 2024)

data ranges from biometric measurements to patient-reported outcomes, descriptive statistics help distil complex information into clear, interpretable metrics.

Measures of Central Tendency

Measures of central tendency identify the central point around which data values cluster. The primary measures are the mean, median, and mode, each providing unique insights, especially when analysing ophthalmic data with specific distribution characteristics, Table 4 and Figure 3.

In multicentre studies or meta-analyses, where data originates from various sources with different sample sizes, the weighted mean can account for these differences among studies, yielding a more accurate overall estimate. When data is log-normally distributed or involves growth rate, such as bacterial counts[60] in ophthalmic infections, the geometric mean provides a more suitable measure of central tendency.

Measures of Dispersion

Measures of dispersion quantify variability within a dataset, complementing central tendency measures by providing insight into data homogeneity, Table 5. Homogeneity refers to the degree of similarity or uniformity among data points.[61] In clinical settings, understanding population variability is essential for interpreting individual measurements. For instance, IOP data for individuals aged 0 to 80 tends to cluster around an age-specific mean, Figure 4. However, central tendency measures alone do not capture the variability within the cohort. Measures of dispersion are essential in estimating the range of values an individual might exhibit, influenced by the data scale and the nature of the measurements.

Data Distribution Characteristics

Data distribution is critical for selecting appropriate statistical methods and interpreting results accurately, as it describes how data points are spread across various values in a dataset. It provides a statistical overview of frequency and likelihood for measurements, guiding expectations and informing clinical decision-making.[69] Understanding the distribution helps determine if assumptions like normality hold, influencing the validity of results. Key properties include central tendency, dispersion, skewness, and kurtosis, which together characterise the shape, spread, and extremities of the data. Skewness measures asymmetry around the mean, indicating if data clusters more on one side, while kurtosis quantifies the sharpness of a distribution's peak, reflecting its propensity to produce outliers. Figure 5 illustrates various distribution scenarios, and Table 6 summarises the types of skewness and kurtosis with clinical examples.

Skewness can be categorised based on thresholds as approximately symmetric (skewness between −0.5 and 0.5), moderately skewed (between −1 and −0.5 or between 0.5 and 1) and highly skewed (below −1 or above 1). These thresholds may vary by field of study, so it is important to consider skewness in the context of data and analysis goals. Skewed data may violate assumptions of parametric statistical tests, which generally assume normal (symmetrical) distributions. Data transformation (e.g., logarithmic) or non-parametric statistical methods may be necessary for valid results. Kurtosis helps assess outlier risk; for instance, in clinical trials, a leptokurtic distribution in outcomes may prompt additional scrutiny of extreme values, ensuring they are genuine observations, not errors. It also informs statistical choices, as some tests are more robust to kurtosis deviations.

Data Normality

Understanding data distribution is fundamental in statistics, as it guides the selection of statistical tests and the validity of inferences drawn from the data.[70] The normal, or Gaussian, distribution is particularly central to statistical theory and practice, especially in EVR. A symmetrical, bell-shaped curve

Table 3: Methods to detect outliers [47]

Type	Characteristics
z-scores or Standard Scores	Measure how many standard deviations a data point is from the mean. The z-score is calculated as: $z = (x - \mu) / \sigma$ z is the z-score of the data point. X is the value of the data point. μ is the mean of the dataset. σ is the standard deviation of the dataset. **Interpretation of Z-scores** • A **z-score of 0** means the data point is exactly at the mean. • **Positive z-scores** indicate that the data point is above the mean. • **Negative z-scores** indicate that the data point is below the mean. • **Z-scores close to +1 or −1** mean the data point is within one standard deviation of the mean, while scores beyond ±2 or ±3 standard deviations are typically considered outliers, depending on the context. This method assumes the data follow a normal distribution, which limits its applicability to normally distributed datasets.
Modified z-score	The median absolute deviation (MAD) is a more robust alternative for non-normally distributed data. The modified z-score is calculated as: $z_{modified} = [0.6745 (x_i - \text{median})] / \text{MAD}$ $z_{modified}$ is the modified z-score of the data point. x_i is the value of the data point. Median is the median of the dataset. MAD (Median Absolute Deviation) is calculated as the median of the absolute deviations from the median: $\text{MAD} = \text{Median}(\lvert x_i - \text{Median} \rvert)$ The constant 0.6745 serves as a scaling factor that adjusts the MAD to be comparable to the standard deviation for large normal distributions, ensuring that modified z-scores align more closely with traditional z-scores in well-behaved data. This method is less sensitive to extreme values and better suited for skewed distributions **Interpretation of Modified Z-scores** • A modified z-score of 0 indicates a value equal to the median. • Higher absolute modified z-scores indicate values further from the median. • Generally, a modified z-score of greater than 3.5 is often used as a threshold for identifying potential outliers. Modified z-scores are especially useful when the data may include outliers or does not meet the assumption of normality. This approach provides a more stable measure of variability for identifying unusual observations in non-normal distributions.
Boxplot	Visually displays outliers as points exceeding 1.5 times the interquartile range (IQR) from the quartiles. Boxplots are particularly useful in exploratory data analysis, providing a simple and visual way to detect outliers. Boxplots, though effective for visualising data distribution and detecting outliers, have limitations. They rely on the 1.5× IQR rule, which may mislabel points as outliers in skewed or small datasets and does not distinguish the severity of outliers. This approach assumes symmetry and may overlook multivariate outliers or ignore nuances in distribution shape. While boxplots are valuable for quick exploratory analysis, additional methods are often necessary to reliably identify true outliers, particularly in complex datasets.

Continued on next page

(Continuation) ***Table 3:*** *Methods to detect outliers* [47]

Type	Characteristics
Grubbs' test	Can be used to detect a single outlier in a univariate dataset. The test statistic is defined as: $G = (\|x_i - \bar{x}) / s$ x_i is the value of the suspected outlier. $\bar{x}$ is the sample mean. s is the sample standard deviation. The test compares the G value to a critical value from Grubbs' distribution table at a chosen significance level (often 0.05). If G exceeds this critical value, the suspected point is considered an outlier. Grubbs' test also assumes normality, making it less effective for non-normally distributed data. For datasets with multiple outliers, repeated application of the test is needed, which may reduce its accuracy.
Mahalanobis distance*	Mahalanobis distance is tailored for settings where variables are interrelated, making it ideal for detecting outliers in complex datasets typical in vision sciences. This distance calculation utilises the entire data structure through its covariance, providing a more accurate spatial assessment compared to simpler distance measures like Euclidean. Practical for both clinical diagnostics and research, Mahalanobis distance helps identify unusual data patterns effectively and is accessible via standard statistical software. The Mahalanobis distance $D^2 = (x - \mu)^T \Sigma^{-1} (x - \mu)$ x is the vector of the data point's values. μ is the mean vector of the dataset. Σ is the covariance matrix of the dataset. $(x - \mu)^T$ is the transpose of the deviation vector.

Table 4: *Measures of Central Tendency and Their Applications in Ophthalmic Research* [55]

Measure	Description	Application
Mean (Arithmetic Average) Applicable for metrical data with/without natural zero	The mean is calculated by summing all data points and dividing by the number of observations. It is sensitive to extreme values (outliers), which can significantly influence the mean in small sample sizes.	**Visual Acuity Scores:** When aggregating visual acuity measurements, when using the logarithm of the minimum angle of resolution (logMAR) scale, the mean provides a precise average that accommodates the logarithmic nature of visual acuity data.[56] **Central Corneal Thickness (CCT):** Mean CCT values are important in assessing risks for diseases like corneal ectasia and glaucoma, where deviations from the mean can have clinical significance.[57]
Median Applicable for metrical data with/without natural zero as well as ordinal data	The median is the middle value when data are ordered sequentially. It is robust against outliers and skewed data distributions, making it valuable in ophthalmic research where data may not be normally distributed.	**Intraocular Pressure (IOP):** IOP readings often exhibit skewness due to the presence of glaucomatous eyes with elevated pressures. The median provides a more representative central value in such cases.[58]
Mode Applicable for all data scales: metric, ordinal, nominal	The mode represents the most frequently occurring value in a dataset. It is particularly useful for categorical or discrete data.	**Classification of Refractive Errors:** Identifying the mode in a population can highlight the most common refractive error (e.g., myopia) within a specific demographic.[59]

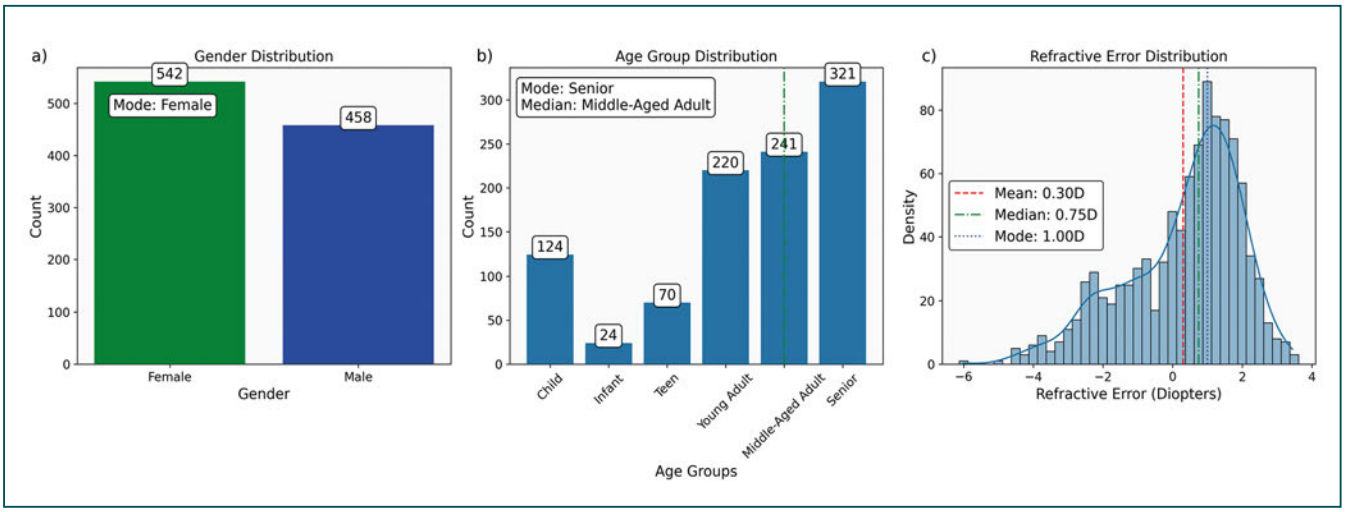

Figure 3: Visual representation of simulated health data (n=1000) showcasing three key central tendency measures (mean, median, and mode). The graphs show (a) the distribution of gender (nominal), (b) age groups (ordinal), and (c) refractive error (metric) within the cohort. The graph shows the distribution of refractive error in the sample, with the bar heights indicating the number of individuals with specific refractive errors. The mean is represented by the red dashed line (approximately 0.3 D), the median (0.75D) and the mode at (1.0D).

characterises this continuous distribution centred around the mean with a defined standard deviation, **Figure 6**.

The normal distribution applies to continuous interval or ratio scale data. In EVR, many biometric measurements, such as corneal thickness, axial length and visual acuity in logMAR units, are ratio-scaled continuous data and can be modelled using the normal distribution,[78] provided they meet criteria for symmetry and lack of significant kurtosis. The normal distribution is foundational in statistical analysis, underpinning several fundamental methods, especially parametric tests like t-tests, analysis of variance (ANOVA), and linear regression, which assume normally distributed data or residuals. Violation of this assumption may lead to inaccurate r esults.

A related and critical concept is the Central Limit Theorem (CLT), which states that as the sample size increases, the distribution of sample means approaches normality, regardless of the population's original distribution. This principle enables researchers to use normal-based inferences with large samples, even if the underlying data are non-normal.[79]

The normal distribution is also essential for statistical inference, particularly when calculating confidence and prediction intervals. These intervals allow for accurate estimation of population parameters and prediction of future observations. They rely on normal distribution properties, which ensure reliable inferences. Lastly, the normal distribution facilitates standardisation, where z-scores (**Table 3**) measure the distance of a data point from the mean in standard deviation units. This standardisation compares across different scales or distributions.

Assessing Normality

Various graphical and statistical methods can be used to assess normality, each with distinct strengths and limitations. Combining graphical and statistical approaches (**Table 7**) provides a more robust assessment of data normality.[70]

Histograms: Histograms provide a straightforward representation of data frequency distributions. An asymmetric, bell-shaped curve generally indicates normality; however, interpreting histograms can be subjective, particularly with smaller samples where the distribution shape may be unclear. Additionally, histograms may not reveal subtle deviations from normality.

Q-Q (Quantile-Quantile) Plots: Q-Q plots compare the quantiles of the sample data to those of a theoretical normal distribution, providing insight into data distribution and identifying outliers, skewness and kurtosis. The x-axis represents theoretical quantiles from a reference distribution (e.g., normal distribution), while the y-axis shows the quantiles of the sample data. Each point corresponds to a quantile of the dataset plotted against a quantile of the reference distribution. To assess normality, the (Euclidean) distance between each data point and the corresponding point on the diagonal (representing the ideal case for a normal distribution) is calculated in **Figures 8a** and **8b**.

Typical Q-Q plot scenarios include:

- Straight Line: If the points align closely along the diagonal, the sample data likely follows a normal distribution.
- Upward or Downward Curvature: Indicates right (upward) or left (downward) skewness, respectively.
- S-shaped Curve: Suggests heavy tails in the data, indicating more extreme values than expected in a normal distribution.
- Outliers: Points far from the line represent outliers or significant deviations from the expected distribution.

P-P (Probability-Probability) Plots: P-P plots display the cumulative probabilities of the sample data (y-axis) against those of a normal distribution (x-axis). Each data point on a P-P plot represents the probability of a sample data point

Table 5: Measures of Dispersion and Their Applications in Ophthalmic Research[61]

Measure	Description	Clinical Scenario Example
Range Applicable for metrical, ordinal data	The range is the simplest measure, calculated as the difference between the maximum and minimum values.	**Axial Length Variations:** Reporting the range of axial lengths in a myopic population can inform the potential variability in surgical outcomes for procedures like implantable Collamer lens insertion.[62,63]
Variance and Standard Deviation (SD) Applicable only for normally distributed, metrical data	Variance represents the average squared deviation from the mean in a normally distributed data set (see chapter 5.4), while the SD is the square root of variance, providing dispersion in the same units as the data. **Single SD (1σ):** Approximately 68 % of the data falls within one standard deviation above and below the mean, about 34% to either side. **Double SD (2σ):** Approximately 95 % of the data falls within two standard deviations above and below the mean, about 47.5% to either side. **Triple SD (3σ):** Approximately 99.7 % of the data falls within three standard deviations above and below the mean. This translates to about 49.85 % to either side.	**Retinal Thickness Measurements:** SD is essential when assessing the variability of retinal nerve fibre layer (RNFL) thickness, aiding in the early detection of glaucoma progression.[64,65]
Coefficient of Variation (CV) Applicable only for normally distributed, metrical data	The CV is a normalised measure of dispersion, calculated as the SD divided by the mean, often expressed as a percentage.	**Consistency of Surgical Outcomes:** In studies comparing surgical techniques, the CV can highlight which method yields more consistent refractive outcomes postoperatively.[66]
Interquartile Range (IQR) Applicable for metrical, ordinal data	The IQR spans the middle 50 % of data, between the 25th (Q1) and 75th (Q3) percentiles, and is less affected by outliers.	**Patient-Reported Outcome Measures (PROMs):** IQR is useful in summarising PROMs like visual function questionnaires, where data may be ordinal and skewed.[67,68]
Advanced Considerations		
Bootstrapping for Confidence Intervals	For small sample sizes or non-normally distributed data, bootstrapping methods can estimate the variability and construct confidence intervals for central tendency measures.	

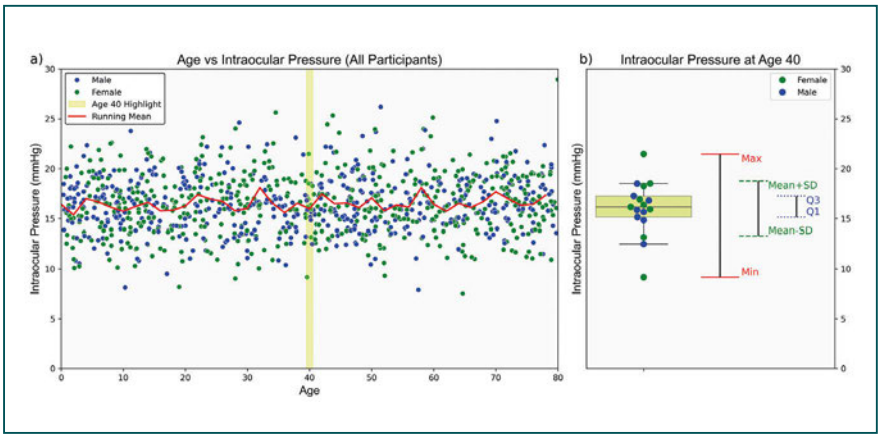

Figure 4: The relationship between age and intraocular pressure (IOP) in a generated dataset shows IOP variation across ages for male and female participants. (a) IOP measurements for all participants, with individuals aged 40 highlighted; the red line represents the (running) mean IOP per year. (b) A box plot detailing IOP at age 40, illustrating key measures of dispersion, including range (distance between maximum and minimum, red), standard deviation (SD, green), and interquartile range (IQR, distance between Q1 and Q3, blue).

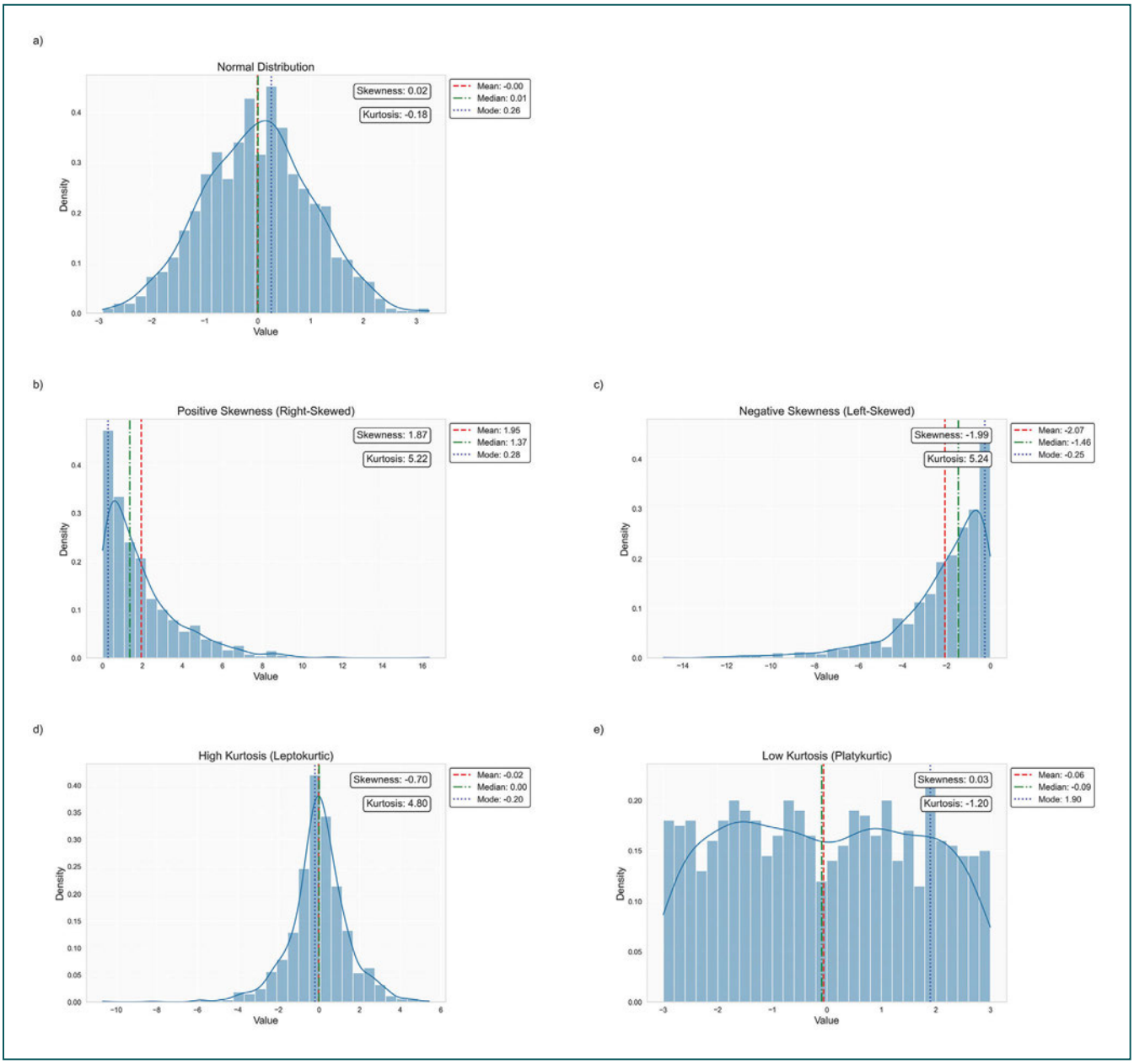

Figure 5: Set of different distributions representing variations in distribution skewness and kurtosis. (a) Normal distribution with skewness and kurtosis is close to zero. (b and c) Right skewed (positive Skewness) and Left skewed (negative skewness), respectively. (d and e) High Kurtosis and Low Kurtosis distributions.

Table 6: Summary of skewness and kurtosis[70]

Skewness	Characteristics	Clinical Scenario
Type		
Skewness = 0	Indicates a symmetrical distribution where the left and right tails are balanced, Figure 5a.	Visual acuity measurements in a large, healthy population may approximate a symmetrical distribution, as most values cluster around a central point, with fewer extremes on either side.[71]
Positive Skewness (Right-Skewed Distribution) Skewness > 0	The tail of the distribution extends to the right, indicating that higher values are less frequent, Figure 5b.	IOP measurements often exhibit right skewness, as most patients have normal pressures, while a few have elevated levels. Recognising this helps clinicians monitor those with high IOP for potential risk factors.[72]
Negative Skewness (Left-Skewed Distribution) Skewness < 0	The tail extends to the left, suggesting lower values are less frequent, Figure 5c.	Some visual field indices, where most patients achieve high scores, show left skewness. Rare low scores may indicate underlying pathology, guiding clinicians to investigate potential issues in affected patients.[73]
Kurtosis		
Mesokurtic Kurtosis = 3 (or Excess Kurtosis = 0)	This indicates a mesokurtic distribution, which has the same kurtosis as a normal distribution, Figure 5a.	—
Leptokurtic (High Kurtosis) Kurtosis > 3 (or Excess Kurtosis > 0)	A leptokurtic distribution has a sharp peak and heavy tails, indicating that data are closely clustered around the mean but with a higher likelihood of extreme values, Figure 5d.	Measurements of endothelial cell counts in the cornea may show leptokurtic distributions, where most values are near the mean, but occasional significant deviations occur. High kurtosis alerts researchers to the potential impact of outliers on statistical analyses and clinical interpretations.[74,75]
Platykurtic (Low Kurtosis) Kurtosis < 3 (or Excess Kurtosis < 0)	A platykurtic distribution features a flatter peak and thinner tails, suggesting a more uniform data spread with fewer extreme values, Figure 5e.	A platykurtic distribution might indicate a wide range of common values with less propensity for outliers in cases such as refractive error measurements in a diverse population.[76,77]

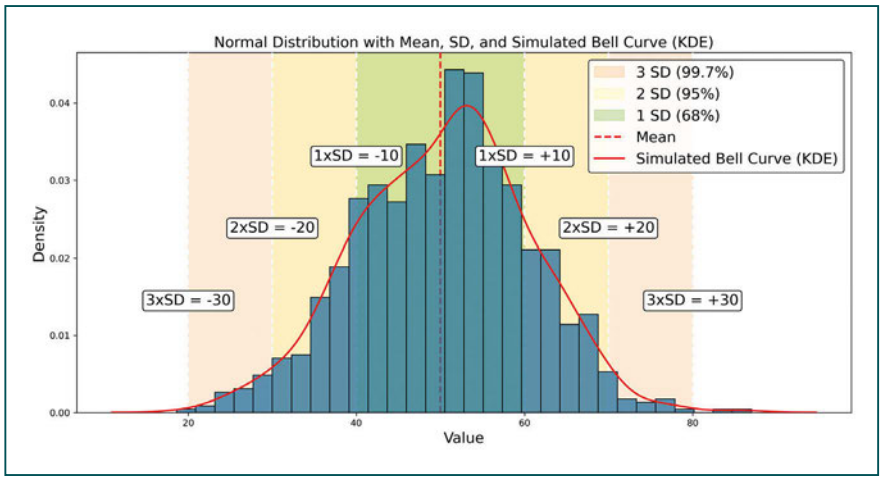

Figure 6: Normal Distribution with l abelled standard deviations (1σ, 2σ, 3σ), illustrating the proportion of data falling within each range. The simulated bell curve (red) demonstrates the concept of normality in clinical data distribution.

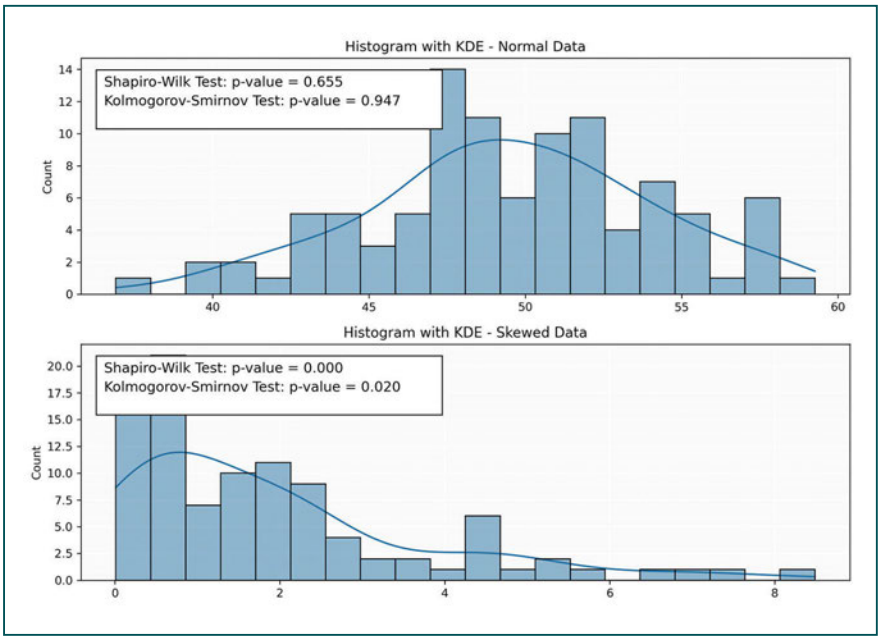

Figure 7: Histograms for Normality Assessment Histograms of normally distributed and skewed data, with accompanying kernel density estimates (KDE) and p-values from the Shapiro-Wilk and Kolmogorov-Smirnov tests. These highlight the visual and statistical assessment of normality in clinical data.

plotted against its corresponding probability from a theoretical distribution. Deviations from the line suggest violations of normality, similar to those described for Q-Q plots. While Q-Q plots provide information about the shape of the distribution, P-P plots assess how well the overall probabilities align between the theoretical and sample distributions. Additionally, P-P plots are less sensitive to deviations in the tails of the distribution compared to Q-Q plots.

Statistical Tests: While statistical tests (Table 7) provide valuable insights into the normality of data, they have limitations. A primary challenge is their sensitivity to sample size; in small samples, these tests may lack the power to detect non-normality (low power), while in large samples, they may identify trivial deviations as significant (over-sensitivity). Consequently, relying solely on statistical tests can be misleading, especially when combined with other tests assuming normality. This highlights the importance of employing graphical and statistical methods to understand the data distribution comprehensively.

Clinical Implications and Meaningfulness of Normality

Assumption of normality significantly impacts clinical research and practice, directly influencing the validity of statistical tests. Incorrectly assuming normality can lead to invalid results, ultimately influencing clinical decisions. For instance, diagnostic thresholds, such as reference ranges for retinal nerve fibre layer thickness, are based on the normal distri-

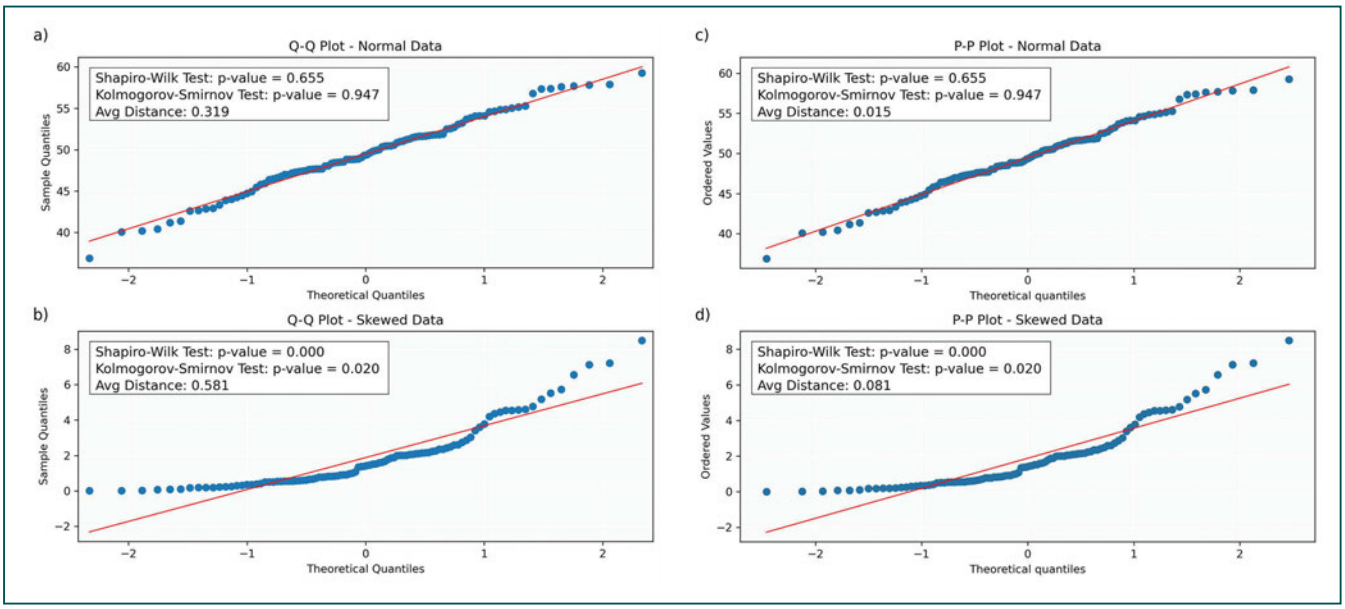

Figure 8: (a and b) Quantile-Quantile plots for normal and skewed data, (c and d) Probability-Probability plots for normal and skewed data.

Table 7: Statistical Tests for Normality in Clinical Data[70]

Test	Description	Limitations
Shapiro-Wilk Test	The Shapiro-Wilk test assesses the null hypothesis that the data are normally distributed. It is particularly effective for small to moderate sample sizes (typically < 50 samples), and it is considered one of the most powerful tests for detecting departures from normality. A p-value less than the chosen significance level (e. g., 0.05) indicates that the data deviate significantly from a normal distribution.[80]	The Shapiro-Wilk test can be overly sensitive to even slight deviations from normality in large datasets, leading to false positives where minor departures from normality are flagged as significant.
Kolmogorov-Smirnov-Test	The Kolmogorov-Smirnov test compares the empirical distribution function of the sample data to the cumulative distribution function of a specified distribution, such as the normal distribution. It is more appropriate for larger samples.	Like the Shapiro-Wilk test, the K-S test may be too sensitive in large samples, detecting deviations from normality that may not significantly affect the results of parametric tests. Additionally, this test may have low power in smaller samples, failing to detect non-normality when it exists.
Anderson-Darling Test	This test is an extension of the K-S test but gives more weight to the tails of the distribution, making it more sensitive to deviations in the extremes. It is suitable for detecting both general departures from normality and specific tail deviations.	Like other normality tests, it may be overly sensitive in large samples, leading to significant results even for minor deviations that do not impact overall analysis.
Lilliefors-Test	A variation of the K-S test, the Lilliefors test is designed for situations where the mean and variance of the population are unknown. It is often used as an alternative to the Shapiro-Wilk test when the sample size is moderate to large.	The Lilliefors test shares similar limitations with other tests, particularly its sensitivity to small deviations in large datasets.
D'Agostino-Pearson Test	This test assesses whether the skewness and kurtosis of the data differ significantly from that of a normal distribution. It combines two tests – one for skewness and one for kurtosis – making it suitable for detecting both asymmetry and heavy or light tails.	This test assumes a reasonably large sample size (at least 20–50 data points) for accurate results and, like other tests, may flag minor deviations in large samples as significant.

bution to identify pathological conditions accurately.[81,82] Deviations from this assumption can result in misclassification of patients, affecting diagnosis and treatment.

In personalised medicine, understanding the distribution of biomarkers, like IOP or central corneal thickness, enables clinicians to tailor interventions to individual patient profiles. A comprehensive understanding of how these biomarkers are distributed within populations informs individualised treatment plans, accounting for patient variability and ensuring more precise, targeted therapeutic approaches.[83,84]

The importance of ensuring normality in clinical data is particularly evident in treatment efficacy studies. For instance, confirming that IOP reduction measurements are normally distributed in a clinical trial evaluating a new glaucoma medication is critical for valid comparisons using parametric tests.[85] Accurate statistical analysis facilitates evidence-based decisions regarding the treatment's efficacy, ultimately determining its adoption in clinical practice.

Data transformation

When data significantly deviate from normality, advanced techniques can help approximate normality and enhance the accuracy of statistical analyses. One commonly used strategy is data transformation. For instance, logarithmic transformations effectively address right-skewed data with positive values, such as reducing skewness in IOP measurements that contain high outliers, **Figure 9a**.[58,72] Similarly, square root transformations help count data or variables following a Poisson distribution, such as the number of microaneurysms in diabetic retinopathy studies, **Figure 9b**.[86,87] Another powerful technique is the Box-Cox transformation, which systematically identifies the optimal power transformation to achieve normality, providing a more formalised approach to data normalisation, **Figure 9c**.[88]

Some ophthalmic data may not follow a standard distribution at all. In cases of mixed distributions, specialised approaches are required. For instance, bimodal distributions,

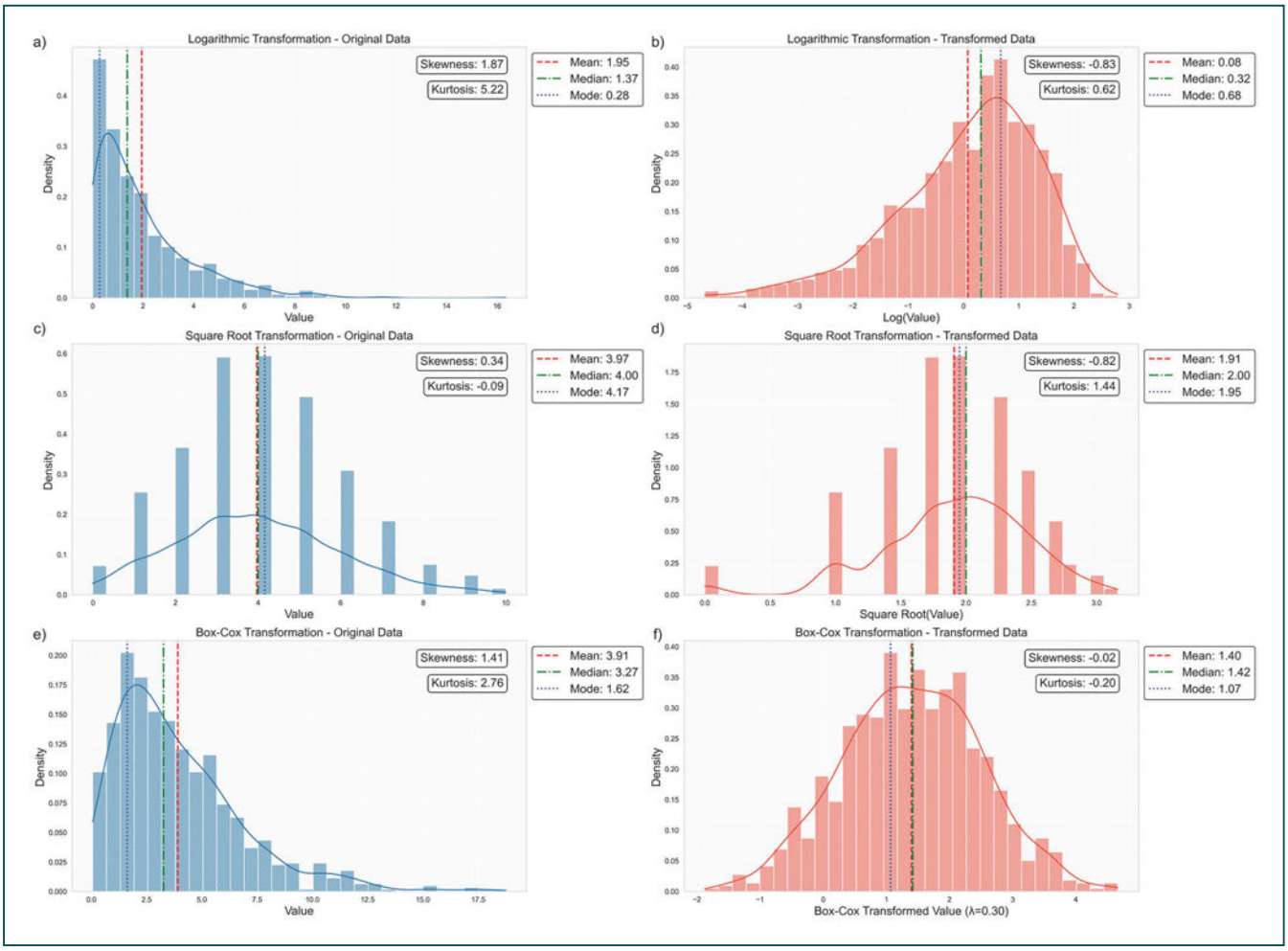

Figure 9: *Original and Transformed Data Distributions. a), c), and e) depict the original Exponential, Poisson, and Gamma distributions. Panels b), d), and f) show the corresponding transformed distributions using logarithmic, square root, and Box-Cox transformations. Each histogram includes a Kernel Density Estimate (KDE) and annotations for mean (red dashed line), median (green dash-dot line), mode (blue dotted line), skewness, and kurtosis, highlighting how each transformation normalises the data distribution.*

characterised by two distinct peaks, can occur in refractive error measurements, reflecting separate peaks for myopia and hyperopia.[76,77] Similarly, multimodal distributions may appear in populations with subgroups, such as age-related variations in lens opacity, requiring techniques like finite mixture models to model the data accurately.[89,90] When traditional transformations are insufficient to achieve normality or the data's structure inherently defies such transformations, non-parametric methods provide a robust alternative. This aligns with best statistical practices, which recommend tailoring the analytical approach to the data's characteristics rather than conforming the data to a preferred test.[91] Non-parametric methods are invaluable in these scenarios, offering flexibility and reliability when parametric assumptions are violated. The subsequent articles in this series will delve deeper into the specific statistical tests suitable for these complex scenarios, ensuring a comprehensive understanding of both parametric and non-parametric methodologies.

Reporting Descriptive Statistics in Publications or Trial Reports

Descriptive statistics summarise the key features of a dataset, providing essential insights into patient demographics, clinical outcomes, and critical variables. By summarising data concisely, descriptive statistics help clinicians and researchers understand the general characteristics of their study population and highlight significant trends.

Reporting Standards for Different Types of Research: Adhering to established reporting standards ensures clarity, reproducibility, and transparency across various research disciplines. Different studies, such as randomised controlled trials, observational studies, qualitative research, and meta-analyses, require specific reporting guidelines (e.g., CONSORT for clinical trials,[92] STROBE for observational studies,[93] PRISMA for systematic reviews[94]). Researchers should familiarise themselves with and follow the appropriate guidelines for their study design to enhance the quality and credibility

of their reports. To effectively report descriptive statistics, researchers should:

- Ensure transparency in data handling, such as how missing data were addressed or outliers were managed.
- Clearly describe the data types (e. g., nominal, ordinal, continuous).
- Report measures of central tendency alongside variability.
- Specify sample sizes for each variable.
- Justify the choice of statistical measures (e. g., using median due to skewed distribution).

Best practices for reporting missing data involve distinguishing between missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR). Researchers should employ appropriate statistical techniques (e. g., multiple imputations or maximum likelihood estimation) and report the extent and pattern of missing data, along with sensitivity analyses conducted to assess the results' robustness.

Outlier identification and management should be transparent and systematic. Standard methods include z-scores, modified z-scores (for non-normal data), and graphical techniques like boxplots. Researchers must evaluate outliers to determine their nature, whether they are data errors, valid physiological extremes, or measurement inaccuracies. Any decisions regarding the handling of outliers should be documented clearly to allow for critical evaluation and replication.

Data visualisation is essential for conveying distribution shape and variability, helping to identify issues such as skewness or outliers. Common methods include histograms, boxplots and Q-Q plots.

Transparency in reporting is crucial for reproducibility. Researchers should include graphical representations of data distributions and discuss normality assumptions and any corrective measures taken.

Descriptive statistics should be linked to clinical relevance, demonstrating their impact on clinical outcomes and decision-making. For instance, defining normative ranges for clinical measurements can assist in diagnosing conditions, while variability in treatment outcomes can inform clinical efficacy.

Conclusion

Rigorous and transparent reporting of descriptive statistics is essential for the integrity of clinical research. By following best practices – such as reporting central tendency and dispersion measures, properly managing outliers and missing data, and providing clear visualisations - researchers can enhance their findings' reliability and clinical relevance, ultimately improving patient care and outcomes in ophthalmic research.

Conflict of Interest Declaration

The authors declare that they have no affiliations with or involvement in any organisation or entity with any financial interest in the subject matter or materials discussed in this manuscript.

Funding Statement

This article did not receive a specific grant from public, commercial, or not-for-profit funding agencies.

Author Contributions

DO was the principal author and initiator of the article and was responsible for its conception, drafting, and writing. PMS reviewed the article and was responsible for reviewing and editing it for intellectual content, grammar, and coherence.

Corresponding author

Prof. Dr. Daniela Oehring

E-Mail:
daniela.oehring@plymouth.ac.uk

Literatur

1 Ho, K. C., Stapleton, F., Wiles, L., Hibbert, P., Alkhawajah, S., White, A., Jalbert, I. (2019). Systematic review of the appropriateness of eye care delivery in eye care practice. BMC Health Serv. Res.,19, 646.

2 Chalmers, I., Glasziou, P. BMJ Blogs. (2016). Is 85% of health research really 'wasted'? British Medical Journal http://blogs bmj com/bmj/2016/01/14/paulglasziou-and-iain-chalmers-is-85-of-health-research-really-wasted/ Accessed. 2016;10.

3 Bykov, K., Patorno, E., D'Andrea, E., He, M., Lee, H., Graff, J. S., Franklin, J. M. (2022). Prevalence of Avoidable and Bias-Inflicting Methodological Pitfalls in Real-World Studies of Medication Safety and Effectiveness. Clin. Pharmacol. Ther., 111, 209-217.

4 Chapman, S. J., Aldaffaa, M., Downey, C. L., Jayne, D G. (2019). Research waste in surgical randomized controlled trials. Br. J. Surg., 106, 1464-1471.

5 Glasziou, P. P., Chalmers, I. (2018). Research waste is still a scandal – an essay by Paul Glasziou and Iain Chalmers. BMJ,363, k, 4645.

6 Torgerson, T., Evans, S., Johnson, B. S., Vassar, M. (2020). The use of systematic reviews to justify phase III ophthalmology trials: an analysis. Eye (Lond). 34, 2041-2047.

7 Okumura, Y. (2016). Reducing Research Waste Through Good Reporting Practices. J. Epidemiol., 26, 397-398.

8 Ramke, J., Kuper, H., Limburg, H., Kinloch, J., Zhu, W., Lansingh, V. C., Congdon, N., Foster, A., Gilbert, C. E. (2018). Avoidable Waste in Ophthalmic Epidemiology: A Review of Blindness Prevalence Surveys in Low and Middle Income Countries 2000-2014. Ophthalmic Epidemiol. 25, 13-20.

9 Kumar, A. Statistical Errors in Ophthalmic Research. Indian Journal of Clinical and Experimental Ophthalmology. 3, 241-242.

10 Bunce, C., Czanner, G. (2022). Common Statistical Issues in Ophthalmic Research. Ophthalmic Epidemiology: CRC Press; pp. 75-84.

11 Van Rossum, G., Drake, Jr. F. L. (1995). Python reference manual. Centrum voor Wiskunde en Informatica Amsterdam.

12 Bitsios, P., Prettyman, R., Szabadi, E. (1996). Changes in autonomic function with age: a study of pupillary kinetics in healthy young and old people. Age Ageing. 25, 432-438.

13 Kiel, M., Grabitz, S. D., Hopf, S., Koeck, T., Wild, P. S., Schmidtmann I., Lackner, K. J., Münzel, T., Beutel, M. E., Pfeiffer, N., Schuster, A. K.(2022). Distribution of Pupil Size and Associated Factors: Results from the Population-Based Gutenberg Health Study. J. Ophthalmol., 9520512.

14 Kohnen, E. M., Zubcov, A. A., Kohnen, T. (2004). Scotopic pupil size in a normal pediatric population using infrared pupillometry. Graefes Arch. Clin. Exp. Ophthalmol., 242, 18-23.

15 Lu, R., Zhang, X., Shi, J. (2021). Tonic pupil size and its variability are associated with fluid intelligence in adolescents aged 11-14 years. Psych. J., 10, 20-32.

16 Salati, C., Salvetat, M. L., Zeppieri, M., Brusini, P. (2007). Pupil size influence on the intraocular performance of the multifocal AMO-Array intraocular lens in elderly patients. Eur. J. Ophthalmol., 17, 571-578.

17 Winn, B., Whitaker, D., Elliott, D. B., Phillips, N. J. (1994). Factors affecting light-adapted pupil size in normal human subjects. Invest. Ophthalmol. Vis. Sci., 35, 1132-1137.

18 Bomotti, S., Lau, B., Klein, B. E. K., Lee, K. E., Klein, R., Duggal, P., Klein, A. P. (2018). Refraction and Change in Refraction Over a 20-Year Period in the Beaver Dam Eye Study. Invest. Ophthalmol. Vis. Sci., 59, 4518-4524.

19 Goldblum, D., Brugger, A., Haselhoff, A., Schmickler, S. (2013). Longitudinal change of refraction over at least 5 years in 15,000 patients. Graefes Arch. Clin. Exp. Ophthalmol., 251, 1431-1436.

20 Ip, J. M., Kifley, A., Rose, K. A., Mitchell, P. (2007). Refractive findings in children with astigmatic parents: the Sydney Myopia Study. Am. J. Ophthalmol., 144, 304-306.

21 Sawada, A., Tomidokoro, A., Araie, M., Iwase, A., Yamamoto, T., Tajimi Study G. (2008). Refractive errors in an elderly Japanese population: the Tajimi study. Ophthalmology, 115, 363-370 e3.

22 Wesemann W. [Analysis of spectacle lens prescriptions shows no increase of myopia in Germany from 2000 to 2015]. Ophthalmologe, 115, 409-417.

23 Williams, K. M., Bertelsen, G., Cumberland, P., Wolfram, C., Verhoeven, V. J., Anastasopoulos, E., Buitendijk, G. H., Cougnard-Grégoire, A., Creuzot-Garcher, C., Erke, M. G., Hogg, R., Höhn, R., Hysi, P., Khawaja, A. P., Korobelnik, J. F., Ried, J., Vingerling, J. R., Bron, A., Dartigues, J. F., Fletcher, A., Hofman, A., Kuijpers, R. W., Luben, R. N., Oxele, K., Topouzis, F., von Hanno, T., Mirshahi, A., Foster, P. J., van Duijn, C. M., Pfeiffer, N., Delcourt, C., Klaver, C. C., Rahi, J., Hammond, C. J.; European Eye Epidemiology (E(3)) Consortium. (2015). Increasing Prevalence of Myopia in Europe and the Impact of Education. Ophthalmology, 122, 1489-1497.

24 Williams, K. M., Verhoeven, V. J., Cumberland, P., Bertelsen, G., Wolfram, C., Buitendijk, G. H., Hofman, A., van Duijn, C. M., Vingerling, J. R., Kuijpers, R. W., Höhn, R., Mirshahi, A., Khawaja, A. P., Luben, R. N., Erke, M. G., von Hanno, T., Mahroo, O., Hogg, R., Gieger, C., Cougnard-Grégoire, A., Anastasopoulos, E., Bron, A., Dartigues, J. F., Korobelnik, J. F., Creuzot-Garcher, C., Topouzis, F., Delcourt, C., Rahi, J., Meitinger, T., Fletcher, A., Foster, P. J., Pfeiffer, N., Klaver, C. C., Hammond, C. J. (2015). Prevalence of refractive error in Europe: the European Eye Epidemiology (E(3)) Consortium. Eur. J. Epidemiol., 30, 305-315.

25 Hashemi, H., Heydarian, S., Aghamirsalim, M., Yekta, A., Hashemi, A., Sajadi, M., Khabazkhoob M. (2023). Distribution and associated factors of intraocular pressure in the older population: Tehran Geriatric Eye Study. Int. J. Ophthalmol., 16, 418-426.

26 Hoffmann, E. M., Aghayeva, F., Wagner, F. M., Fiess, A., Nagler, M., Munzel, T., Wild, P. S., Beutel, M. E., Schmidtmann, I., Lackner, K. J., Pfeiffer, N., Schuster, A. K. (2022). Intraocular Pressure and Its Relation to Ocular Geometry: Results From the Gutenberg Health Study. Invest. Ophthalmol. Vis. Sci., 63, 40.

27 Hoffmann, E. M., Lamparter, J., Mirshahi, A., Elflein, H., Hoehn, R., Wolfram, C., Lorenz, K., Adler, M., Wild, P. S., Schulz, A., Mathes, B., Blettner, M., Pfeiffer, N. (2013). Distribution of central corneal thickness and its association with ocular parameters in a large central European cohort: the Gutenberg health study. PLoS One, 8, e66158.

28 Ma, D., Wei, S., Sun, Y., Li, S. M., An, W. Z., Hu, J. P., Cao, K., Yang, X. H., Lin, C. X., Guo, J. Y., Li, H., Fu, J., Wang, N. (2021). Distribution of IOP and its relationship with refractive error and other factors: the Anyang University Students Eye Study. Int. J. Ophthalmol., 14, 554-559.

29 Suzuki, S., Suzuki, Y., Iwase, A., Araie, M. (2005). Corneal thickness in an ophthalmologically normal Japanese population. Ophthalmology. 112, 1327-1336.

30 Wang, Y. X., Xu, L., Wie, W. B., Jonas, J. B. (2011). Intraocular pressure and its normal range adjusted for ocular and systemic parameters. The Beijing Eye Study 2011. PLoS One, 13, e0196926.

31 Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., Del Río, J. F., Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke C., Oliphant, T. E. (2015). Array programming with NumPy. Nature, 585, 357-362.

32 McKinney, W. (2015). Pandas, python data analysis library. URL http://pandas pydata org., 3-15.

33 Barrett, P., Hunter, J., Miller, J. T., Hsu, J. C., (2025). matplotlib – A Portable Python Plotting Package. (Shopbell, L., Britton, M. C., Ebert, eds.) R. Astronomical data analysis software and systems XIV. ASP Conference Series, 347, 91-95.

34 Waskom, M., Botvinnik, O., O'Kane, D., Hobson, P., Lukauskas, S., Gemperline, D. C., Augspurger, T., Halchenko, Y., Cole, J. B., Warmenhoven, J., De Ruiter, J., Pye,, C., Hoyer, S., Vanderplas, J., Villalba, S., Kunter, G., Quintero, E., Bachant, P., Martin, M., Meyer, K. (2017). mwaskom/seaborn, v0. 8.1, Zenodo.

35 Oehring, D., Serra, P. M. (2024). Advancing Statistical Literacy in Eye Care: A Series for Enhanced Clinical Decision-Making. 1 ed. OSF.

36 Hellerstein, J. M., Heer, J., Kandel, S. (2018). Self-Service Data Preparation: Research to Practice, IEEE Data Eng Bull., 41, 23-34.

37 Kordon, A. K. (2020). Data Preparation. Applying Data Science. 221-249.

38 Newman, D. A. (2014). Missing Data. Organizational Research Methods, 17, 372 - 411.

39 Kang, H. (2013). The prevention and handling of the missing data. Korean J Anesthesiol., 64, 402-406.

40 Rubin, D. B. (1976). Inference and Missing Data, Psychometrika, 63, 3, 581-592.

41 Suresh, R., Yu, H. J., Thoveson, A., Swisher, J., Apolinario, M., Zhou, B., Shah, A. R., Fish, R. H., Wykoff, C. C. (2020). Loss to Follow-Up Among Patients With Proliferative Diabetic Retinopathy in Clinical Practice. Am. J. Ophthalmol., 215, 66-71.

42 Vengadesan, N., Ahmad, M., Sindal, M. D., Sengupta, S. (2017). Delayed follow-up in patients with diabetic retinopathy in South India: Social factors and impact on disease progression. Indian. J. Ophthalmol., 65, 376-384.

43 Elegbede, A., Andrei, A., Holen, K. D. (2009). Reconsenting patients with cancer on clinical trials: Does added risk influence continued participation? Journal of Clinical Oncology, 27, 15 suppl., e15610.

44 Nantz, E. , Liu-Seifert, H., Skljarevski, V. (2009). Predictors of premature discontinuation of treatment in multiple disease states. Patient Prefer Adherence, 3, 31-43.

45 Hansen, J., Ahern, S., Earnest, A. (2023). Evaluations of statistical methods for outlier detection when benchmarking in clinical registries: a systematic review. BMJ Open, 13, e069130.

46 Gress, T. W., Denvir, J., Shapiro, J. I. (2018). Effect of removing outliers on statistical inference: implications to interpretation of experimental data in medical research. Marshall J. Med., 4, 9.

47 Prasad, S. (2014). Advanced Statistical Methods. Publisher, Springer Nature Singapore.

48 Deneshkumar, V., Senthamaraikannan, K., Manikandan, M. (2014). Identification of Outliers in Medical Diagnostic System Using Data Mining Techniques. International journal of statistics and applications. 4, 241-248.

49 Forster, J. E., MaWhinney, S., Ball, E. L., Fairclough, D. (2012). A varying-coefficient method for analyzing longitudinal clinical trials data with nonignorable dropout. Contemp. Clin. Trials, 33, 378-385.

50 Anderson, A. J., Bedggood, P. A., George Kong, Y. X., Martin, K. R., Vingrys, A. J. (2017). Can Home Monitoring Allow Earlier Detection of Rapid Visual Field Progression in Glaucoma? Ophthalmology, 124, 1735-1742.

51 Crabb, D. P., Garway-Heath, D. F. (2012). Intervals between visual field tests when monitoring the glaucomatous patient: wait-and-see approach. Invest. Ophthalmol. Vis. Sci., 53, 2770-2776.

52 Verma, S., Nongpiur, M. E., Atalay, E., Wie, X., Husain, R., Goh, D., Perera, S. A., Aung, T. (2017). Visual Field Progression in Patients with Primary Angle-Closure Glaucoma Using Pointwise Linear Regression Analysis. Ophthalmology, 124, 1065-1071.

53 Zhang, S., Chen, Y., Li, Z., Wang, W., Xuan, M., Zhang, J., Hu, Y., Chen, Y., Xiao, O., Yin, Q., Zheng, Y., He, M., Han, X. (2024). Axial Elongation Trajectories in Chinese Children and Adults With High Myopia. JAMA Ophthalmol., 142, 87-94.

54 Flores-Moreno, I., Puertas, M., Almazan-Alonso, E., Ruiz-Medrano, J., Garcia-Zamora, M., Vega-Gonzalez, R., Ruiz-Moreno, J. M. (2022). Pathologic myopia and severe pathologic myopia: correlation with axial length. Graefes Arch. Clin. Exp. Ophthalmol., 260, 133-140.

55 Adamson, K. A., Prion, S. K. (2013). Making Sense of Methods and Measurement: Measures of Central Tendency. Clinical Simulation in Nursing, 9, e617–e618.

56 Holladay, J. T. (19997). Proper method for calculating average visual acuity. J. Refract. Surg., 13, 388-391.

57 Agarwal, L., Agrawal, N., Badhu, B. P., Lavaju, P. (2019). Central corneal thickness and intraocular pressure in patients of primary open angle glaucoma and normal population in Nepalese population: A hospital based study. Nepal J. Ophthalmol., 11, 46-54.

58 Cui, Y., Yang, X., Zhang, G., Guo, H., Zhang, M., Zhang, L., Zeng, J., Liu, Q., Zhang, L., Meng, Q. (2019). Intraocular Pressure in General and Diabetic Populations From Southern China: the Dongguan Eye Study. Invest Ophthalmol. Vis. Sci., 60, 761-769.

59 Xu, L., Li, J., Cui, T., Hu, A., Fan, G., Zhang, R., Yang, H., Sun, B., Jonas, J. B. (2005). Refractive error in urban and rural adult Chinese in Beijing. Ophthalmology, 112, 1676-1683.

60 Robertson, A. H. (1932). Averaging Bacterial Counts. J. Bacteriol., 23, 123-134.

61 Rayat, C. S. (2018). Measures of Dispersion. In: Statistical Methods in Medical Research (eds. Rayat, C. S.) Springer Nature, pp. 47-52.

62 Chen, X., Li, L., Rao, J., Chen, Y. X., Gao, Y., Huang, R. X., Zhou, Q. Z. (2023). Long-term observation on safety and visual quality of implantable collamer lens V4c implantation for myopia correction: a 5-year follow-up. Int. J. Ophthalmol., 16, 1123-1129.

63 Liu, H., Li, F. F., Xia, H. J., Zhou, J. (2021). Visual quality after implantation of trifocal intraocular lenses in highly myopic eyes with different axial lengths. Int J Ophthalmol., 14, 371-377.

64 Wu, J. H., Moghimi, S., Walker, E., Nishida, T., Liebmann, J. M., Fazio, M. A, Girkin, C. A., Zangwill, L. M., Weinreb, R. N. (2024). Long-term variability of retinal nerve fibre layer thickness measurement in patients with glaucoma of African and European descents. Br. J. Ophthalmol., 108, 1094-100.

65 San Pedro, M. J. B., Sosuan, G. M. N., Yap-Veloso, M. I. R. (2024). Correlation of Macular Ganglion Cell Layer + Inner Plexiform Layer (GCL + IPL) and Circumpapillary Retinal Nerve Fiber Layer (cRNFL) Thickness in Glaucoma Suspects and Glaucomatous Eyes. Clin. Ophthalmol., 8, 2313-2325.

66 Razmjoo, H., Atarzadeh, H., Kargar, N., Behfarnia, M., Nasrollahi, K., Kamali, A. (2017). The Comparative Study of Refractive Index Variations between Haigis, Srk/T and Hoffer-Q Formulas Used for Preoperative Biometry Calculation in Patients with the Axial Length >25 mm. Adv. Biomed. Res., 6, 78.

67 Robertson, A. O, Tadic, V., Cortina-Borja, M., Rahi, J., (2021). Child Vision Pg. Feasibility of using patient-reported outcome measures with visually impaired children/young people attending paediatric ophthalmology clinics. Arch. Dis. Child., 106, 687-692.

68 Hepworth, L. R., Rowe, F. J., Burnside, G. (2019). Development of a patient reported outcome measures for measuring the impact of visual impairment following stroke. BMC Health Serv. Res., 19, 348.

69 Baek, H. J., Kim, H. S., Kim, N., Choi, Y. J., Kim, Y. J. (2012). Percent change of perfusion skewness and kurtosis: a potential imaging biomarker for early treatment response in patients with newly diagnosed glioblastomas. Radiology, 264, 834-843.

70 Redmond, C. K., Colton, T. (2001)). Biostatistics in Clinical Trials. Wiley.

71 Cheng, C. Y., Hsu, W. M., Liu, J. H., Tsai, S. Y., Chou, P. (2003). Refractive errors in an elderly Chinese population in Taiwan: the Shihpai Eye Study. Invest. Ophthalmol. Vis., 44, 4630-4638.

72 Colton T, Ederer F. The distribution of intraocular pressures in the general population. Surv. Ophthalmol., 25, 123-129.

73 Chauhan, B. C., Henson, D. B. (1987). The distribution of visual field scores in a normal population. Ophthal. Physiol. Opt., 7,109-115.

74 Schimmelpfennig, B. H. (1984). Direct and indirect determination of nonuniform cell density distribution in human corneal endothelium. Invest. Ophthalmol. Vis. Sci., 25, 223-229.

75 Ono, T., Mori, Y., Nejima, R., Iwasaki, T., Miyai, T., Miyata, K. (2021). Corneal endothelial cell density and morphology in ophthalmologically healthy young individuals in Japan: An observational study of 16842 eyes. Sci. Rep., 11, 18224.

76 Fricke, T.. R, Keay, L., Resnikoff, S., Tahhan, N., Koumbo, O., Paudel, P., Ayton, L. N., Britten-Jones, A. C., Kweon, S., Li, J. C. H., Lee, L., Wagner, P., Weng, R., Beranger, B., Olivier, J. (2023). Improving population-level refractive error monitoring via mixture distributions. Ophthalmic Physiol. Opt., 43, 445-453.

77 Cumberland, P. M., Bao, Y., Hysi, P. G., Foster, P. J., Hammond, C. J., Rahi, J. S.; UK Biobank Eyes & Vision Consortium.(2015). Frequency and Distribution of Refractive Error in Adult Life: Methodology and Findings of the UK Biobank Study. PLoS One., 10, e0139780.

78 Ojaimi, E., Rose, K. A., Morgan, I. G., Smith, W., Martin, F. J., Kifley, A., Robaei, D., Mitchell, P. (2005). Distribution of ocular biometric parameters and refraction in a population-based study of Australian children. Invest. Ophthalmol. Vis. Sci. 46, 2748-2754.

79 Adams, W. J. (1974). The life and times of the central limit theorem. Kaedmon Publishing Company.

80 Shapiro, S. S., Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). Biometrika., 52, 591-611.

81 Lingham, G., Lee, S. S., Charng, J., Clark, A., Chen, F. K., Yazar, S., Macckey, D, A. (2021). Distribution and Classification of Peripapillary Retinal Nerve Fiber Layer Thickness in Healthy Young Adults. Transl. Vis. Sci. Technol., 10, 3.

82 Yanni, S. E., Wang, J., Cheng, C. S., Locke, K.I, Wen, Y., Birch, D. G., Birch, E. E. (2013). Normative reference ranges for the retinal nerve fiber layer, macula, and retinal layer thicknesses in children. Am. J. Ophthalmol., 155, 354-360 e1.

83 Wolfs, R. C., Klaver, C. C., Vingerling, J. R., Grobbee, D. E., Hofman, A., de Jong, P. T. (1997). Distribution of central corneal thickness and its association with intraocular pressure: The Rotterdam Study. Am. J. Ophthalmol., 123, 767-772.

84 Ramm, .L, Spoerl, E., Pillunat, L. E., Terai, N. (2020). Is the Corneal Thickness Profile Altered in Diabetes Mellitus? Curr. Eye Res., 45, 1228-1234.

85 Heijl, A., Leske, M. C., Bengtsson, B., Hyman, L., Bengtsson, B., Hussein, M. Early Manifest Glaucoma Trial Group. (2002). Reduction of intraocular pressure and glaucoma progression: results from the Early Manifest Glaucoma Trial. Arch. Ophthalmol., 120, 1268-1279.

86 Tavakoli, M., Pourreza-Shahri, R., Pourreza, H. R., Mehdizadeh, A., Banaee, T., Toosi, M. T. B. A. (2013). complementary method for automated detection of microaneurysms in fluorescein angiography fundus images to assess diabetic retinopathy. Pattern Recognit., 46, 2740-2753.

87 Chen, S. J., Chou, P., Lee, A. F., Lee, F. L., Hsu, W. M., Liu, J. H., Tung, T. H. (2010). Microaneurysm number and distribution in the macula of Chinese type 2 diabetics with early diabetic retinopathy: a population-based study in Kinmen, Taiwan. Acta Diabetol., 47, 35-41.

88 Blackie, C. A., Harris, W. F. (1997). Refraction and keratometry: departures from and transformations toward multivariate normality. Optom. Vis. Sci., 74, 452-458.

89 Manouchehri, N., Rahmanpour, M., Bouguila, N., Fan, W. (2019). Learning of Multivariate Beta Mixture Models via Entropy-based component splitting. IEEE Symposium Series on Computational Intelligence (SSCI)., 2825-2832.

90 Quintero, F. O. L, Contreras-Reyes, J. E. (2017). Estimation for finite mixture of simplex models: applications to biomedical data. Statistical Modelling. 18, 129-148.

91 Fife, D. A. (2019). The Eight Steps of Data Analysis: A Graphical Framework to Promote Sound Statistical Analysis. Perspectives on Psychological Science, 15, 1054-1075.

92 Moher, D., Schulz, K. F., Altman, D., Group, C. (2001). The CONSORT Statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. Lancet, 357, 1191-1194.

93 von Elm, E., Altman, D. G., Egger, M., Pocock, S. J., Gotzsche, P. C., Vandenbroucke, J. P.; STROBE-Initiative. (2008). The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: guidelines for reporting observational studies. Internist (Berl). 688-693.

94 Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G.; Prisma Group (2010). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Int. J. Surg., 8, 336-341.

95 van der Valk, R., Schouten, J. S., Webers, C. A., Hendrikse, F., Prins, M. H. (2006). Changes in glaucoma treatment and achieved IOP after introduction of new glaucoma medication. Graefes Arch. Clin. Exp. Ophthalmol., 244, 1267-1272.

96 Bengtsson, B., Leske, M.C., Hyman, L., Heijl, A.; Early Manifest Glaucoma Trial Group (2007). Fluctuation of intraocular pressure and glaucoma progression in the early manifest glaucoma trial. Ophthalmology, 114, 205-209.

97 Chevret, S., Seaman, S., Resche-Rigon, M. (2005). Multiple imputation: a mature approach to dealing with missing data. Intensive Care Med., 41, 348-350.

98 Saffari, S. E., Volovici, V., Ong, M. E. H., Goldstein, B. A., Vaughan, R., Dammers, R., Steyerberg, E. W., Liu, N. (2022). Proper Use of Multiple Imputation and Dealing with Missing Covariate Data. World Neurosurg., 161, 284-290.

99 Huque, M. H., Carlin, J. B., Simpson, J. A., Lee, K. J. A (2018). Comparison of multiple imputation methods for missing data in longitudinal studies. BMC Med. Res. Methodol., 18, 168.

100 Rosato, R., Pagano, E., Testa, S., Zola, P., di Cuonzo, D. (2021). Missing data in longitudinal studies: Comparison of multiple imputation methods in a real clinical setting. J. Eval. Clin. Pract., 27, 34-41.

101 Janssen, K. J., Donders, A. R., Harrell, F. E., Jr., Vergouwe, Y., Chen, Q., Grobbee, D. E., Moons, K. G.(2010). Missing covariate data in medical research: to impute is better than to ignore. J. Clin. Epidemiol., 63, 721-727.

102 Cao, Y., Allore, H., Vander, Wyk, B., Gutman, R. (2022). Review and evaluation of imputation methods for multivariate longitudinal data with mixed-type incomplete variables. Stat. Med., 41, 5844-5876.

103 Laird, N. M., Lange, N., Stram, D. O. (1987). Maximum likelihood computations with repeated measures: application of the EM algorithm. Journal of the American Statistical Association, 82, 97-105.

104 Anderson, A. J., Johnson, C. A. (2013). How useful is population data for informing visual field progression rate estimation? Invest. Ophthalmol. Vis. Sci., 54, 2198-2206.

105 Daza, E. J., Hudgens, M. G., Herring, A. H. (2017). Estimating inverse-probability weights for longitudinal data with dropout or truncation: The xtrccipw command. Stata. J., 17, 253-278.

106 Shardell, M., Hicks, G. E., Ferrucci, L. (2015). Doubly robust estimation and causal inference in longitudinal studies with dropout and truncation by death. Biostatistics, 16, 155-168.

107 Kong, F., Chen, Y. F., Jin K. (2009). A bias correction in testing treatment efficacy under informative dropout in clinical trials. J. Biopharm. Stat., 19, 980-1000.

108 Liu, G. F., Liu, F., Mehrotra, D. V. (2020). Model Averaging Using Likelihoods That Reflect Poor Outcomes for Clinical Trial Dropouts. Statistics in Biopharmaceutical Research. 12, 79 - 89.

Appendix

Table A1: Summary of Methods for Handling Missing Data

Method	Applicability	Advantages	Limitations	Clinical Application
Complete Case Analysis	Listwise or Case wise Deletion. Applicable for MCAR	Simplifies analysis	Reduces sample size; potential bias if not MCAR	Studies with truly random missing data but sufficient sample size. In studies with small sample sizes, such as those focused on rare ocular diseases, removing cases with missing data can severely compromise the study's conclusions.
Single Imputation	MCAR or when missingness is minimal	Easy to implement	Underestimates variability; potential bias	Preliminary analyses; not recommended for final analysis
a) Mean or Median Imputation	Replaces missing values with the mean (for normally distributed data) or median (for skewed data) of observed values.	Simple and easy to implement.	Underestimates variability, leading to biased standard errors and confidence intervals and ignoring relationships between variables.	This method may obscure true associations, particularly in studies measuring average retinal thickness where relationships between variables are critical.
b) Regression Imputation	Predicts missing values based on regression models using other observed variables.	Incorporates relationships between variables, improving the quality of imputed values.	Does not account for uncertainty in predictions, treating imputed values as known and potentially leading to overconfidence in estimates.	This method might introduce bias in clinical research, as in studies of glaucoma progression where over-reliance on imputed values could misguide clinical interpretations.

Continued on next page

(Continuation) ***Table A1:*** *Summary of Methods for Handling Missing Data*

Method	Applicability	Advantages	Limitations	Clinical Application
c) Last Observation Carried Forward (LOCF)	This method, common in longitudinal studies, replaces missing values with the last observed measurement.	Maintains sample size and is simple to apply.	Assumes no change over time, which is often unrealistic and can bias results.	LOCF is problematic for progressive conditions like glaucoma, where IOP may change significantly over time, making this approach inappropriate [95,96]
Multiple Imputation	Robust method that accounts for the uncertainty in imputed values by generating multiple datasets with different plausible estimates for missing data. Applicable for MAR	Preserves variable relationships; valid inferences	Computationally intensive	**Process:** 1. **Imputation:** Generate mm complete datasets by imputing missing values multiple times from their predictive distribution. 2. **Analysis:** Perform statistical analysis on each dataset separately. 3. **Pooling:** Combine the results using Rubin's rules to derive overall estimates and standard errors. Imputing missing OCT measurements based on observed patient characteristics.[97-102]
Maximum Likelihood (EM)	Estimate model parameters by maximising the likelihood function using all available data, accounting for the missing data structure. Applicable for MAR	Efficient estimates under correct model	Requires correct model specification	In glaucoma studies with intermittent follow-up data, the EM algorithm can be used to estimate progression rates using all available information, providing accurate estimates even with incomplete datasets.[103,104]
Weighting Methods (IPW)	Weighting methods, such as Inverse Probability Weighting (IPW), adjust for missing data by assigning weights to observed cases based on the probability of being observed. Applicable for MAR	Utilises all data; reduces bias	Requires modeling missingness probabilities	In longitudinal studies of contact lens comfort, where dropout rates may depend on prior discomfort levels, IPW can adjust for differential dropout and improve the accuracy of results.[105,106]
Model-Based Approaches	Model-based approaches explicitly model the missing data mechanism. Applicable for MNAR	Models the missing data mechanism explicitly	These models require strong assumptions about the missing data mechanism, which are often difficult to verify. Misspecification can lead to biased results.	Handling dropout due to unobserved adverse effects in ocular medication trials.[49,107,108]