Journal of the Mechanical Behavior of Biomedical Materials 172 (2025) 107171

Contents lists available at ScienceDirect

Journal of the Mechanical Behavior of Biomedical Materials

journal homepage: www.elsevier.com/locate/jmbbm

Evaluation of statistical methods to study flexural strength of dental CAD-CAM composites

Yousef Karevan [a], Christelle Sanchez [a,b], Adelin Albert [c], Amélie Mainjot [a,b,*]

[a] Dental Biomaterials Research Unit (d-BRU), Institute of Dentistry, University of Liège (ULiège), Liège, Belgium
[b] Department of Fixed Prosthodontics, Institute of Dentistry, University Hospital of Liège (CHU), Liège, Belgium
[c] Biostatistics and Research Method Center (B-STAT), University Hospital of Liège (CHU) and University of Liège (ULiège), Liège, Belgium

A R T I C L E I N F O

Keywords:
Prosthetic dentistry/prosthodontics
Quasi-brittle material
Polymer infiltrated ceramic network
Composite materials
Leave-K-Out cross-validation
Weibull analysis

A B S T R A C T

Determining the optimal statistical method is key for reliable interpretation of the flexural strength test. To date, statistical recommendations regarding the best approach and sample size are based on theoretical assumptions that may not hold in practice. Therefore, this study identified the optimal statistical approach for analyzing the flexural strength of computer-aided design/computer-aided manufacturing (CAD-CAM) composites using a real dataset. In this perspective, the flexural strength dataset of commercial CAD-CAM composites, Cerasmart 270 (CER), Katana Avencia (KAT), Grandio (GRN), and polymer-infiltrated ceramic network Vita Enamic (ENA) were used (10 blocks/material; 15 samples/block). Leave-K-out cross-validation was performed on this dataset with K = 3 to 9 blocks (a total of 967 scenarios per material). The goal was to compare seven common statistical methods: Normal distribution (1), Lognormal distribution (2), the 2-parameter Weibull distribution calculated by maximum likelihood estimation (MLE) (3) and least squares (LSQ) estimation (LSQ; with three different estimators 4-5-6); (4) 3-parameters Weibull distribution. These methods were assessed based on comparative performance, standalone performance, and lower tail prediction.

The results highlight that all statistical approaches have limited reliability with small sample sizes (e.g., n = 30) employed in dental materials research. Factors such as inter-block heterogeneity and physical characteristics of CAD-CAM composites could affect the efficacy of the statistical methods. The 2P-Weibull distribution was less prone to overestimating strength at low failure probability. LSQ with the mean estimator $(i/(n+1))$ comparatively outperformed others, particularly with small sample sizes. Weibull analysis would be more reliable with over 60 samples, but ideally more than 100 is recommended.

Abbreviations

AD^2	Anderson-Darling goodness of fit test	**LSQ**	Least square estimation
AD^2*	Adjusted Anderson-Darling goodness of fit test	**LSQ1**	Least square estimation with mean estimator $(i/n+1)$
CAD-CAM	Computer aided-design/computer aided manufacturing	**LSQ2**	Least square estimation with median estimator $(i-0.3)/(n+0.4)$
CDF	Cumulative density function	**LSQ3**	Least square estimation with Hazen estimator $(i-0.5)/n$
CER	Cerasmart 270	**MLE**	Maximum likelihood estimation
DF	Dispersed fillers	**MLE-2P**	MLE of 2P-Weibull distribution

(*continued on next column*)

(*continued*)

ENA	Vita Enamic	**MLE-3P**	MLE of 3P-Weibull distribution
GRN	Grandio Voco	**Norm**	Normal distribution
KAT	Katana Avencia	**PICN**	Polymer-infiltrated ceramic network
LKO-CV	Leave-K-out cross-validation	**PDF**	Probability density function
LogN	Lognormal distribution	**TS**	Training set
		VS	Validation set

1. Introduction

The adoption of CAD-CAM technology in prosthetic dentistry opened

* Corresponding author. Dental Biomaterials Research Unit (d-BRU) and Department of Fixed Prosthodontics, Institute of Dentistry, University of Liège (ULiège) and University of Liège Hospital (CHU), 45 Quai G. Kurth, Liège, 4020, Belgium.

E-mail address: a.mainjot@chuliege.be (A. Mainjot).

https://doi.org/10.1016/j.jmbbm.2025.107171
Received 12 May 2025; Received in revised form 24 July 2025; Accepted 18 August 2025
Available online 20 August 2025

the door for the development of various restorative materials (Liu et al., 2023). Among them, CAD-CAM composite blocks are especially promising. In comparison to ceramics, they provide faster milling, extended bur longevity, superior resistance to chipping, and the capacity for ultra-thin milling, favoring minimally invasive treatment (Chavali et al., 2017; Coldea et al., 2015; Curran et al., 2017; Lebon et al., 2015; Oudkerk et al., 2024; Pfeilschifter et al., 2018). Moreover, there is no need for firing procedures (Mainjot et al., 2016). CAD-CAM composites fall into two categories based on their microstructure: dispersed fillers (DF) and polymer-infiltrated ceramic network (PICN or "hybrid ceramic") materials. In DF, ceramic fillers are embedded in an organic matrix, which is polymerized at high temperatures, while PICN features a pre-sintered ceramic network infiltrated with organic monomers, then polymerized under high temperature and pressure (Mainjot et al., 2016).

The strength of dental materials is emphasized by manufacturers because it directly relates to load-bearing capacity (Wendler et al., 2017). Accordingly, in contrast to more fundamental concepts such as fracture toughness, which can be challenging to interpret, strength can be easily understood and applied by non-experts (Wendler et al., 2017). In this context, the "stronger" materials can be considered as better candidates for areas subjected to higher masticatory loads (Wendler et al., 2017). While dental materials are usually evaluated using mean flexural strength, different studies suggest that the strength at 1 % or 5 % failure probability is more important for clinical use (Lohbauer et al., 2002). Indeed, it is important that samples corresponding to a 1 % failure probability (for example, the weakest 100 out of 10,000 samples) are strong enough to avoid fractures in clinical scenarios. The example of 10,000 samples is based on the recommended size to reliably determine material strength at a 1 % failure probability (Le and Bažant, 2009). However, while determining this stress level purely by experiment is reliable and assumption-free, it is typically impractical due to resource constraints. Alternatively, one can assume a known parametric strength distribution and test its validity on a smaller, practical data set.

In this framework, finding the appropriate statistical distribution to describe the strength of materials is crucial. Normal, Weibull, and, to a lesser extent, Lognormal are three commonly used distributions in materials engineering. The Normal distribution is a common symmetric probability distribution known for its bell-shaped curve. The Lognormal distribution is used for positively skewed data. In contrast, the Weibull distribution with the Weibull modulus often seen in dental materials is employed to describe negatively skewed data. Fig. 1 shows schematically how each one of these distributions represents the scatter of dataset with a 200 MPa median (50th percentile) and a coefficient of variation of 10 %, which represents a typical scatter for flexural strength of CAD-CAM composites (Ducke and Ilie, 2021; Eldafrawy et al., 2023).

If the strength, denoted by X, follows the Normal distribution, then the cumulative density function (CDF)—which calculates the probability that the strength is less than or equal to a specific value of x—is expressed as:

$$F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right) \quad \text{Eq.(1)}$$

where x is the strength value, μ the mean of X, σ the standard deviation (SD) of X, and $\Phi(t)$ the Gaussian cumulative function given by $\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{u^2}{2}} du$.

In the Lognormal distribution, the natural logarithm of strength, noted log(X), follows the Normal distribution. The CDF of this distribution is:

$$F(x) = \Phi\left(\frac{\log(x)-\mu'}{\sigma'}\right) \quad \text{Eq.(2)}$$

where x is the strength value, μ' the mean of log(X) and σ' the standard deviation of log(X).

When X follows a Weibull distribution, the CDF writes:

$$F(x) = 1 - \exp\left[-\left(\frac{x}{\sigma_0}\right)^m\right] \quad \text{Eq.(3)}$$

where x is the strength value, m describes the shape of the distribution of strength X (Weibull modulus) and σ_0 the scale parameter at which the CDF of X is 63.2 % (characteristics strength).

1.1. Estimating the parameters of weibull distribution (m, σ_0)

There are different ways to estimate the parameters m and σ_0 of the Weibull distribution.

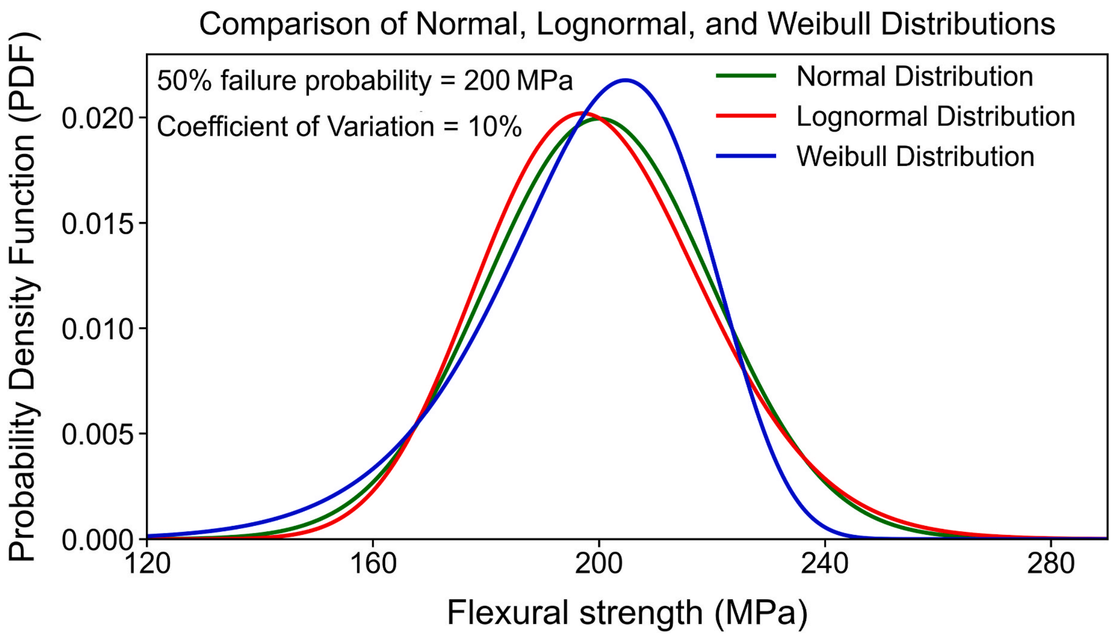

Fig. 1. Data scatter representation for a 200 MPa median with 10 % coefficient of variation in Normal, Lognormal, and Weibull Distributions.

1.1.1. Least square estimation method (LSQ)

Eq. (3) can be linearized by rearranging and taking the logarithm of both sides twice.

$$\log\left[\log\left[\frac{1}{1-F(x)}\right]\right] = m \log x - m \log \sigma_0 \qquad \text{Eq.(4)}$$

To get the exact F(x) at each stress level, an infinite number of samples would need to be tested, which is practically impossible (Khalili and Kromp, 1991). Therefore, the failure probability at different stress levels must be estimated. This is done by ranking samples based on their fracture strengths in ascending order and using estimator functions to assign each strength (stress) a failure probability F(i). The simplest forms of probability estimators, F(i) = (i-1)/n and F(i) = i/n, assign probabilities of occurrence of 0 % and 100 % to the lowest and highest data points, respectively. However, making absolute statements about the impossibility or certainty of an event is not supported in statistics (Datsiou and Overend, 2018). To address these limitations, several other estimators have been proposed, with three more commonly used (Eqs. (5)–(7)) (Benard and Bos-Levenbach, 1953; Hazen, 1914; Weibull, 1939).

$$F(i) = \frac{i}{n+1} \quad \text{mean estimator} \qquad \text{Eq.(5)}$$

$$F(i) = \frac{i-0{\cdot}3}{n+0{\cdot}4} \quad \text{median estimator} \qquad \text{Eq.(6)}$$

$$F(i) = \frac{i-0{\cdot}5}{n} \quad \text{Hazen estimator} \qquad \text{Eq.(7)}$$

where “i" is the rank of observation (strength) and n is total number of experimental observations.

LSQ estimates of m and σ_0 can then be obtained by applying ordinary linear regression (OLR) to the values $Y_i = \log\left[\log\left[\frac{1}{1-F(i)}\right]\right]$ on the log x_i (i = 1 … n) as given by Eq. (8). Thus, m is the slope of the regression line and σ_0 is equal to the exponential of (minus the intercept divided by the slope).

$$Y = m \cdot \log(x) - m \cdot \log(\sigma_0) \qquad \text{Eq.(8)}$$

1.1.2. Maximum likelihood estimation method (MLE)

The failure probability of a sample with strength x, belonging to a Weibull distribution with shape parameter (m) and scale parameter (σ_0), can be calculated using Eq. (9) (the probability density function (PDF) of the Weibull distribution):

$$f(x) = \frac{m}{\sigma_0} \times \left(\frac{x}{\sigma_0}\right)^{m-1} \times \exp\left[-\left(\frac{x}{\sigma_0}\right)^m\right] \qquad \text{Eq.(9)}$$

When considering the likelihood of observing a set of strength values $x_1, \ldots, x_n$, the individual probabilities associated with each strength should be multiplied (according to the joint probability of independent observations). The MLE method determines the population parameters (m, σ_0), which maximize the product of individual failure probabilities calculated by Eq. (9) (typically through an iterative procedure by computer). MLE offers several advantages over LSQ, such as no need to use F(i) and a significantly tighter confidence interval (Quinn and Quinn, 2010).

1.1.3. MLE of the 3-parameter weibull distribution

In many cases, including dental materials, it has been found that the 2-parameter Weibull distribution cannot properly fit the material's strength data at low failure probability (Le and Bažant, 2009). To circumvent this problem, the 3-parameter Weibull distribution can be an attractive choice. The 3-parameter Weibull distribution (Eq. (10)) is distinguished from the 2-parameter Weibull distribution through the incorporation of a threshold value (δ); all values below δ are assumed to have zero probability of occurrence. The 3-parameter Weibull distribution is less frequently used than the 2-parameter Weibull distribution, which can be due to its more complex computation procedure and less conservative determination of low failure probability. However, some studies have reported that 3P Weibull distribution outperforms 2P Weibull distribution in describing the flexural strength of materials (Han et al., 2009; Nohut, 2021; Yin et al., 2023).

$$F(x) = 1 - \exp\left[-\left(\frac{x-\delta}{\sigma_0}\right)^m\right] \textit{ with } x > \delta \qquad \text{Eq.(10)}$$

1.2. Leave-K-out cross-validation (LKO-CV)

Different strategies can be adopted to evaluate which statistical methods are a better choice to describe the strength of a material. The accuracy of a statistical method is determined by how well the estimated population aligns with the observed data (the closer the match, the better). For instance, a probability plot of the Weibull or Normal distribution can serve as a visual approach for this evaluation strategy.

However, it is well established in statistics and machine learning that a good fit to observed data does not guarantee that a model will generalize well. This concept is illustrated in Fig. 2. Suppose Fig. 2A represents the overall underlying population, while Fig. 2B and C shows models fitted to a limited sample drawn from this population. Although the more complex model provides a better fit to the sample compared to the simpler one (Fig. 2B vs. Fig. 2C), it exhibits inferior generalizability when applied to new, unseen data (Fig. 2E vs. Fig. 2D). Although, for simplicity, this example used only limited data and polynomial models, probability distributions themselves offer different degrees of flexibility. Whereas the Normal distribution can only model symmetric histograms, the Weibull distribution can model a wide range of histogram shapes depending on its Weibull modulus (shape parameter, Fig. 3).

Considering that the ultimate objective of statistical analysis is to be able to predict the behavior of unexplored data, if a large dataset is available, the leave-out or hold-out validation method can be used. Suppose the dataset is an optimal representative of the whole company's CAD-CAM composite product. The dataset is divided into training set (TS) and validation set (VS). The TS is statistically analyzed to estimate the characteristics of the overall dataset, much like researchers infer from the statistical analysis of limited samples. The VS is not used in the calculations but serves to evaluate the generalizability and effectiveness of the TS's statistical analysis. The reliable statistical approach is expected to properly represent the characteristics of the VS.

Splitting the dataset with n data into a TS and a VS can be done repeatedly and systematically by taking out K data as the VS in each iteration, with the remaining (n−K) data as the TS. The iteration continues until all combinations of selecting K data out of n are covered. This technique is called leave-K-out cross-validation (LKO-CV), and as multiple scenarios are analyzed, it enhances conclusion reliability by minimizing the risk of chance-based inferences (Yu, 2019). The above-mentioned processes are shown schematically in Fig. 4.

Currently there is a knowledge gap regarding the best statistical distribution that represents the strength of CAD-CAM composites. Some distributions have a greater physical meaning but still do not have a definitive physical basis to be used for every material (e.g., Weibull distribution). Indeed, to follow the Weibull distribution, a material must meet specific criteria, such as elasticity, presence of unimodal and non-interacting defects, and absence of R-curve behavior (in R-curve materials, the growth of a crack requires progressively more mechanical energy to be applied to the system until it reaches the plateau stage) (Barsoum, 2019; Danzer et al., 2007; Quinn and Quinn, 2010). That is possibly why many studies have found that the Weibull distribution does not adequately explain the observed scatter of strength in materials with properties that deviate from the assumptions (Basu et al., 2009; Belli and Lohbauer, 2021; Danzer et al., 2007; Nohut and Lu, 2012; Van Den Born et al., 1991). Accordingly, unless the superiority of using the Weibull

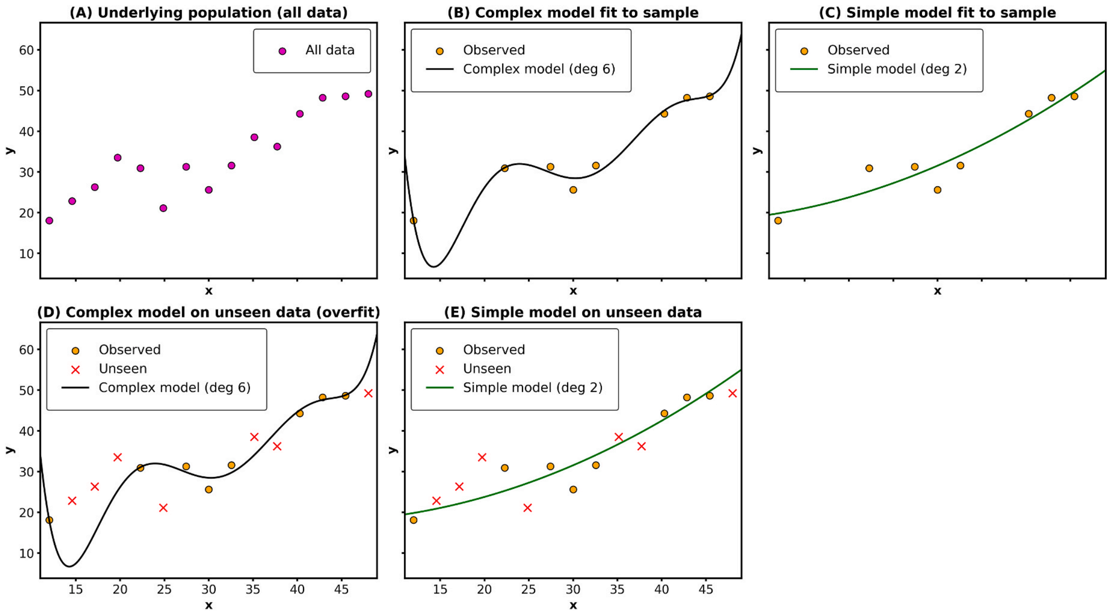

Fig. 2. The impact of model over-complexity on generalizability. (A) The true underlying relationship in the population; (B) Fit of a complex model to a training sample; (C) Fit of simple model to the same sample; (D) Generalization performance of the complex model on unseen data; (E) Generalization performance of the simple model on unseen data.

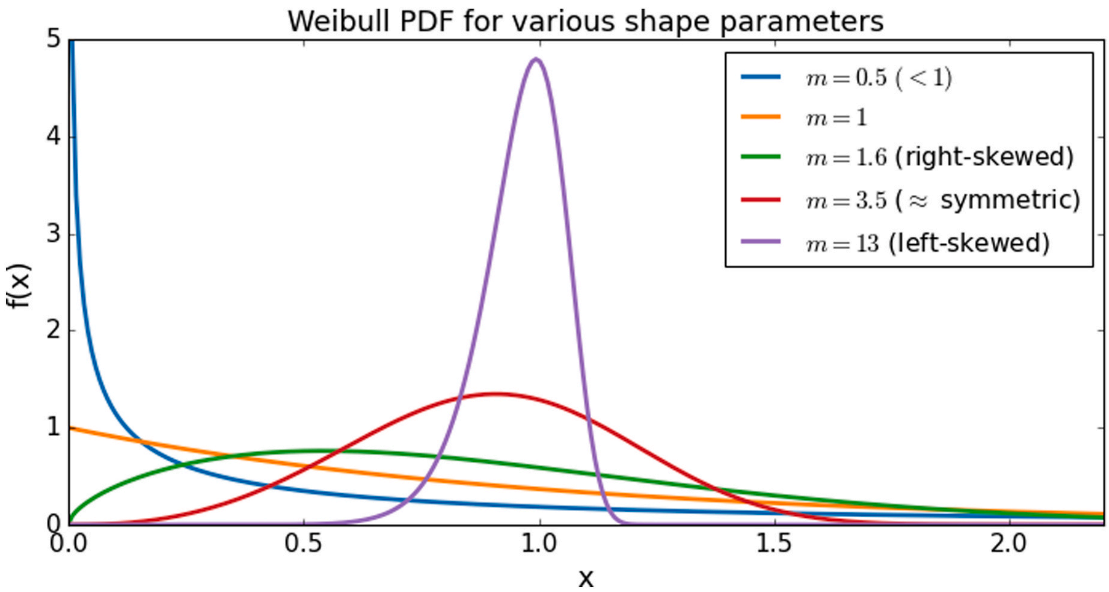

Fig. 3. The Weibull distribution is highly flexible: varying the shape parameter (m) produces a range of probability density function, f(x), forms from highly right-skewed to nearly symmetric or even left-skewed, with x as the variable and scale parameter $x_0 = 1$.

distribution compared to other alternatives is demonstrated, it should be regarded as being on equal footing with others.

Moreover, there is also a lack of consensus on the required sample size for reliable statistical analysis of the Weibull distribution, with recommendations varying from 20 to more than 100 (Barsoum, 2019; Ghelbere and Ilie, 2022; Quinn and Quinn, 2010; Trustrum and Jayatilaka, 1979; Xu et al., 2001). On a related note, the size of the tested data can also affect the choice of calculation approach for Weibull analysis. For instance, some authors reported that MLE outperforms LSQ regardless of the amount of data used (Khalili and Kromp, 1991), while others claim a sample size of at least 53 is needed for its superior performance (Wu et al., 2006).

Consequently, this study aims to use LKO-CV (the K range from 3 to 9 blocks) to analyze and compare the discussed statistical methods in terms of explaining the strength of CAD-CAM composites. This approach allows for comparing the efficacy of statistical methods across different sample sizes for statistical analysis, ranging from 1 block (L9O-CV; 15 flexural strength data points) to 7 blocks (L3O-CV; 105 flexural strength points). The comparison of the methods involves assessing their comparative and standalone predictive power, as well as their capability to reliably predict the strength at lower failure probabilities (VS).

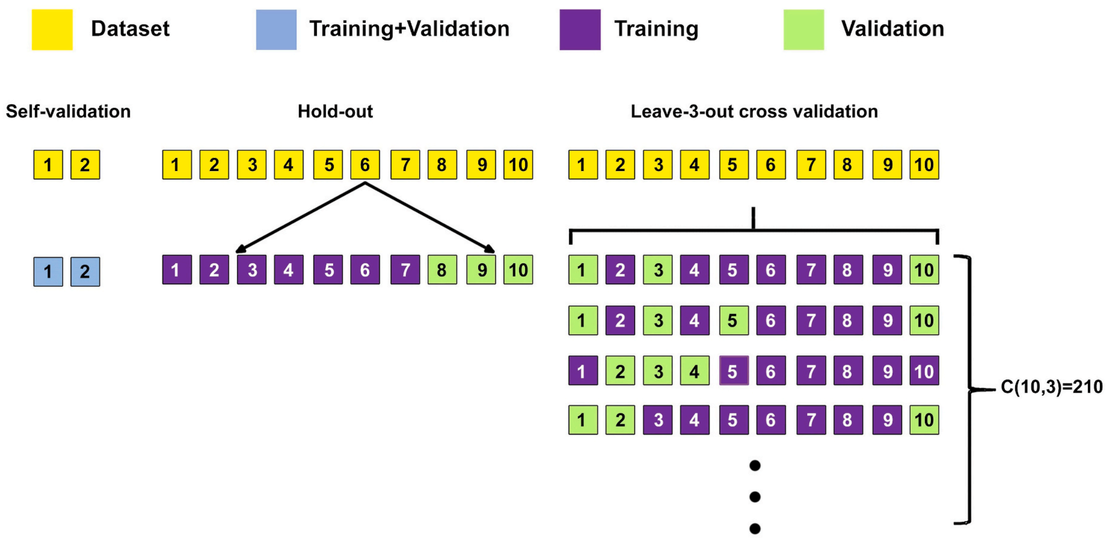

Fig. 4. Schematic representation of self-validation, hold-out method, and leave-K-out cross-validation techniques. In the context of this study, each square represents one block of CAD-CAM composites (15 specimens per block).

2. Materials and methods

2.1. Dataset preparation

This study used the dataset from Eldafrawy et al. (2023). Table 1 shows the compositions and lot numbers of the four commercially available composite materials studied in this work. Ten blocks of each CAD–CAM composite material were cut using a low-speed saw (Isomet; Buehler, Lake Bluff, IL, USA) with continuous water irrigation to produce bars with a dimension of 1.6 x 4.0 x 17 ± 0.1 mm in accordance with ISO 6872:2015 (International Organization for Standardization, 2015). The cutting process yielded 15 bars per block (n = 150 bars per material). The bars were then polished to a 20 µm finish using a diamond pad at 150 rpm under water (Struers, Ballerup, Denmark), following the recommendations of the same standard.

For the 3-point flexural strength test, the bars were subjected to a flexural testing device with a span length of 15 mm. The test was conducted on a computer-controlled universal testing machine (Instron model 5565, equipped with an extensometer) at a cross-head speed of 1 mm/min. The flexural strength (σ_f) of the bars was calculated using Eq. (11).

$$\sigma_f = \frac{3FL}{2ch^2} \quad \text{Eq.(11)}$$

where F represents the load at fracture, L the span, c the specimen width, and h the specimen height. The values of h and c were measured prior to testing each sample using a digital caliper (Mitutoyo).

Table 1
CAD-CAM blocks used in the study and their compositions.

Material	Composition		Manufacturer	Lot
	Organic matrix	Inorganic fillers		
Cerasmart 270 (CER)	Bis-MEPP + UDMA + DMA	Barium glass (300 nm) + silica (20 nm) (71 wt%)	GC corporation (Tokyo, Japan)	1908206 2103011
Grandio Voco (GRN)	UDMA + other DMA	Nanohybrid fillers (87 wt%) Size of fillers unknown	Voco Gmbh, (Cuxhaven Germany)	2145165 2212243
Katana Avencia (KAT)	UDMA + TEGDMA	Al_2O_3 (20 nm) + SiO_2 (40 nm) (62 wt%)	Kuraray Noritake (Tokyo, Japan)	L000473 L000484
Vita Enamic (ENA)	UDMA + TEGDMA	Glass-ceramic sintered network (86 wt%, 75 vol %)	Vita Zahnfabrik (Bad Sackingen, Germany)	65900 77100 65971

Data were completed according to manufacturers' information. SiO_2 Silicon oxide; Al_2O_3: Aluminum oxide; bis-GMA: bisphenol A glycidylmethacrylate; bis-EMA: Ethoxylated bisphenol A dimethacrylate; UDMA: urethane dimethacrylate; TEGDMA: triethylenglycol dimethacrylate; DMA: dimethacrylate; Bis-MEPP: 2,2-Bis(4-methacryloxypolyethoxyphenyl) propane.

2.2. LKO-CV

LKO-CV, with K ranging from 3 to 9 blocks, was applied to the dataset of each material (the total number of blocks for each material was 10, with each block containing 15 samples). This resulted in 10 distinct TS with 1 block (L9O-CV), 45 with 2 blocks (L8O-CV), 120 with 3 blocks (L7O-CV), 210 with 4 blocks (L6O-CV), 252 with 5 blocks (L5O-CV), 210 with 6 blocks (L4O-CV), and 120 with 7 blocks (L3O-CV). Fig. 5 schematically illustrates this process. LKO-CV were performed using the pandas library (The pandas development team, 2024), itertools, and collections libraries in Python (version 3.11.4, Anaconda Inc.).

2.3. Statistical analysis

The statistical analysis of generated training-validation pairs followed specific sequential steps: (1) perform statistical calculations on the TS; (2) assess the accuracy of the statistical inference through examining the degree to which the calculated populations fit the VS (goodness of fit test); and (3) aggregate all the conclusions drawn from training-validation pairs of the same sample size to derive the conclusion based on three criteria.

2.3.1. Flexural strength distribution determination

The summary of statistical approaches used in the present study is provided in Table 2. The parent population parameters were estimated assuming the dataset followed the Normal, Lognormal, and Weibull distributions. The Weibull parameters were calculated using five statistical methods, including LSQ (mean, median, Hazen estimator) and MLE (for two and three-parameter Weibull distributions). A detailed description of these statistical approaches is provided in introduction 1.2. MLEs were performed using the SciPy library (Virtanen et al.,

Fig. 5. Schematic of the leave-K-out cross-validation procedure (K = 9 to 3), corresponding to training sets of 1–7 blocks.

Table 2
Summary of statistical methods employed in the present study.

	Distribution	Method	Estimator	Abbreviation
1	Weibull	LSQ estimation	Mean rank (Eq. (5))	LSQ1
		LSQ estimation	Hazen's rank (Eq. (6))	LSQ2
		LSQ estimation	Median rank (Eq. (7))	LSQ3
		ML estimation (2P)	–	MLE-2P
		ML estimation (3P)	–	MLE-3P
2	Normal	ML estimation	–	Norm
3	Lognormal	ML estimation	–	LogN

LSQ: least-squares; ML: Maximum likelihood.

2020), and reliability package (Reid, 2020) in Python (version 3.11.4, Anaconda Inc.).

2.3.2. *Goodness of fit test (Anderson-darling test)*

The Anderson-Darling (AD^2) (Anderson and Darling, 1954) goodness-of-fit test was used to determine how effectively distributions can represent the strength data of the VS. This is achieved by measuring the agreement between the CDF of the theoretical distribution, derived from statistical analysis of the TS, and the empirical distribution observed in the VS. Eq. (12) can be used to calculate the AD^2 value:

$$AD^2 = -n - \sum_{i=1}^{n} \frac{(2\cdot i - 1)}{n} \times \{\log[F(i)] + \log[1 - F(n+1-i)]\} \qquad \text{Eq.(12)}$$

A smaller AD^2 value indicates a better fit between the statistical prediction for the VS, based on the TS, and the observed VS data. One key advantage of AD^2 is that it detects deviations in the distribution tails more effectively than alternatives such as the χ^2 test (Tiryakioğlu and Hudak, 2007).

2.3.3. ***The average rank method:*** *aggregating individual analyses*

The results of LKO-CV were analyzed using the average rank technique to determine the best calculation method; in the average rank technique, the calculation methods are ranked based on their AD^2, followed by assigning corresponding ranks. The statistical method with the lowest AD^2 value ranked one, and the one with the highest AD^2 value received rank seven. For each sample size, statistical methods were finally ranked by averaging their ranks across all TS with that sample size (Eq. (13)) (Brazdil and Soares, 2000).

$$\text{Final rank of calculation method (M)} = \frac{\sum_{i=1}^{N} r_i^M}{N} \qquad \text{Eq.(13)}$$

where r_i^x is the rank of method M for the i-th TS, and N the total number of TS.

2.3.4. *The success rate: aggregating individual analyses*

The success rate criterion was used to determine how often the output of statistical analysis can accurately describe the scatter of flexural strength of the VS. It quantifies the standalone effectiveness of statistical methodologies, whereas the average rank criterion assesses their comparative performance (one statistical approach may outperform others but still be irrelevant for describing the flexural strength distribution of the VS).

The first step in calculating the overall success rate was to determine the p-value (P_{AD}) associated with the adjusted AD^2 (AD^{2*}). For the Normal and Lognormal distributions, AD^{2*} was calculated using Eq. (14). Subsequently, Eqs. 15–18 were used for P_{AD} determination (D'Agostino and Stephens, 1986). For the Weibull distribution, the AD^{2*} was obtained through Eq. (19), after which Eq. (20) was employed to determine P_{AD} (Aslam, 2021). The statistical analysis fails to effectively predict the distribution of flexural strength in the VS when the P_{AD} value is less than $\alpha = 0.05$ (5 %). The overall success rate is defined as the ratio of successful predictions to the total number of predictions.

$$AD^{2*} = AD^2 \times \left(1 + \frac{0.75}{n} + \frac{2.25}{n^2}\right) \quad \text{Eq.(14)}$$

$$\text{If } AD^{2*} \geq 0\cdot6; \text{ then } P_{AD} = \exp\left[1\cdot2937 - 5\cdot709 \times (AD^{2*}) + 0\cdot0186 \times (AD^{2*})^2\right] \quad \text{Eq.(15)}$$

$$\text{If } 0{\cdot}34 \leq AD^{2*} \leq 0\cdot6; \text{ then } P_{AD} = \exp\left[0\cdot9177 - 4\cdot279 \times (AD^{2*}) - 1\cdot38 \times (AD^{2*})^2\right] \quad \text{Eq.(16)}$$

$$\text{If } 0{\cdot}2 \leq AD^{2*} \leq 0\cdot34; \text{ then } P_{AD} = 1 - \exp\left[-8\cdot318 + 42\cdot796 \times (AD^{2*}) - 59\cdot938 \times (AD^{2*})^2\right] \quad \text{Eq.(17)}$$

$$\text{If } AD^{2*} \leq 0\cdot2; \text{ then } P_{AD} = 1 - \exp\left[-13.436 + 101.14 \times (AD^{2*}) - 223.73 \times (AD^{2*})^2\right] \quad \text{Eq.(18)}$$

$$AD^{2*} = \left(1 + \frac{0{\cdot}2}{\sqrt{n}}\right) \cdot AD^2 \quad \text{Eq.(19)}$$

$$P_{AD} = \frac{1}{1 + \exp\left[-0{\cdot}1 + (1{\cdot}24 \times \ln AD^{2*}) + (4{\cdot}48 \times AD^{2*})\right]} \quad \text{Eq.(20)}$$

In Eqs. (14) and (19), n is the number of flexural strength data in VS.

2.3.5. *Optimistic vs. conservative prediction of lower tail*

The third criterion was used to determine whether the statistical analysis offers an optimistic or conservative outlook on the strength of CAD-CAM composites at low failure probabilities. The difference from the previous two criteria is that they provided insight into the effectiveness of the statistical approach in explaining the distribution throughout. Yet, the third criterion focused on a low failure probability. Now, the output of the statistical analysis was defined as "overly optimistic" when it predicted a 1 % chance of material failure at a specific stress level, while the empirical cumulative distribution function from the VS indicates a 3 % or higher chance of failure at that same stress level. The empirical cumulative distribution function, representing the CDF of experimental data in the VS, was calculated as previously described (Eqs. (5)–(7)). To apply this criterion, the statistical analysis on the TS was first used to determine the stress level associated with a 1 % failure probability. Then, using Eq. (7), the empirical cumulative distribution function of the VS was evaluated at that stress level; if it exceeded 3 %, the statistical method was considered overly optimistic. This concept is illustrated schematically in Fig. 6. The summary of materials and methods is illustrated in Fig. 7.

3. Results

Tables 3–6 show how well different statistical methods performed on the CER, GRN, KAT, and ENA datasets, according to average rank and success rate criteria. Regarding the average rank criterion, in the case of DF composites, when the TS is small (n ≤ 30), the LSQ estimation by regression analysis had better performance than the others in predicting the fracture behavior of the VS. Among the LSQ methods, LSQ1 was the top performer followed by LSQ2 and LSQ3. In contrast, the MLE-2P, MLE-3P, and Lognormal distribution had poor performance in describing the failure behavior of the VS. The Normal distribution occupies the middle ground. As the sample size increased, the predictive power of the MLE-2P and MLE-3P methods improved, narrowing the performance gap with alternatives that were better at predicting unseen data in smaller sample sizes. For ENA, the Normal distribution

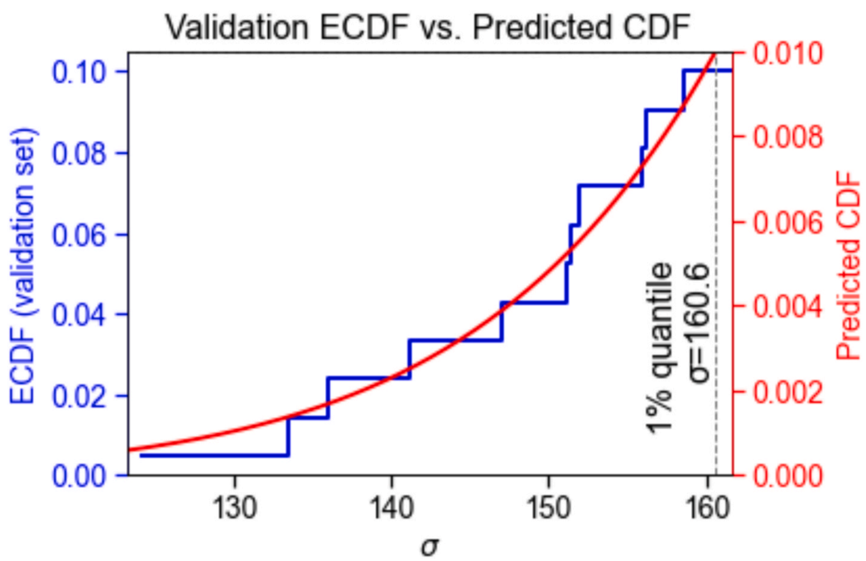

Fig. 6. Schema of the "overly optimistic" prediction of the MLE-2P method in a leave-3-out cross-validation (CER270; training blocks 2, 5, and 6). The statistical analysis of the training set predicted a 1 % failure probability at 160.6 MPa (predicted cumulative density function; red curve); the empirical cumulative density function of the validation set at 160.6 MPa exceeds 3 % (blue curve). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

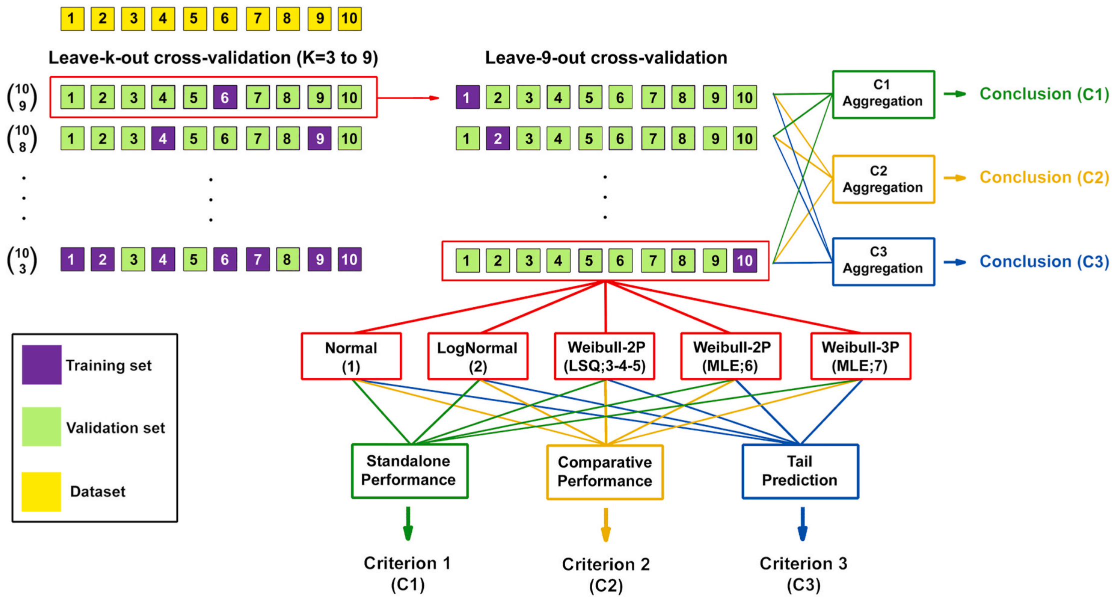

Fig. 7. Schematic of materials and methods: leave-K-out cross-validation (K = 9–3), with a detailed illustration of a leave-9-out scenario. Each scenario in leave-9-out cross-validation is analyzed using seven statistical methods, compared across three criteria, and results are integrated for each criterion.

Table 3
Results of the average rank method and success rate criteria for different statistical approaches (CER).

No. of blocks	Average rank							Success rate (%)						
	LSQ			MLE		Norm	LogN	LSQ			MLE		Norm	LogN
	1	2	3	2P	3P			1	2	3	2P	3P		
1	1.3	2.5	3.7	5.6	6.0	4.6	4.3	10.0	10.0	0.0	0.0	0.0	0.0	0.0
2	2.6	3.0	3.7	5.2	5.5	3.8	4.1	6.7	4.4	6.7	6.7	6.7	2.2	0.0
3	2.5	2.9	3.6	5.0	5.5	4.1	4.4	8.3	8.3	8.3	4.2	3.3	5.0	0.8
4	2.5	3.1	3.8	4.7	5.3	3.9	4.6	10.5	11.9	11.4	5.7	4.3	4.8	0.5
5	2.6	3.1	3.9	4.7	5.3	3.8	4.5	14.3	14.3	14.3	11.5	8.7	7.9	3.2
6	2.7	3.2	3.9	4.5	5.3	3.8	4.4	12.9	13.3	14.3	11.0	10.0	8.6	3.3
7	2.7	3.4	4.2	4.5	5.3	3.8	4.1	15.8	12.5	12.5	15.0	12.5	11.7	3.3

No. of blocks: Number of blocks used in leave-out cross-validation. Average rank indicates the comparative performance of methods in predicting validation sets behavior (based on the AD^2 goodness-of-fit test; lower is better); success rate (%) represents the percentage of successful predictions (P-value for $AD^{2*} \geq 0.05$). Methods: LSQ1–3 (2-parameter Weibull, least squares estimation using Eqs. (5)–(7) estimator), MLE-2P/3P (2-/3-parameter Weibull, maximum likelihood estimation), Norm (Normal), LogN (Lognormal).

Table 4
Results of the average rank method and success rate criteria for different statistical approaches (GRN).

No. of blocks	Average rank							Success rate (%)						
	LSQ			MLE		Norm	LogN	LSQ			MLE		Norm	LogN
	1	2	3	2P	3P			1	2	3	2P	3P		
1	3.4	3.3	3.1	4.9	5.4	3.8	4.1	0.0	0.0	10.0	10.0	0.0	0.0	0.0
2	3.4	3.4	3.7	4.2	4.7	3.9	4.8	6.77	6.7	8.9	6.7	2.2	0.0	0.0
3	3.5	3.3	3.7	3.9	4.3	4.2	5.2	11.7	14.2	15.8	18.3	14.2	7.5	0.0
4	3.4	3.3	3.6	3.9	4.1	4.2	5.5	18.6	20.5	21.0	19.5	17.1	6.2	1.0
5	3.3	3.3	3.8	3.9	4.0	4.2	5.5	23.0	25.4	22.6	22.2	19.8	10.3	2.0
6	3.3	3.4	3.9	3.9	3.9	4.2	5.4	27.1	27.6	26.7	27.1	25.2	13.8	7.1
7	3.3	3.5	4.1	4.0	4.0	3.9	5.2	30.0	33.3	33.3	32.5	28.3	20.8	10.8

No. of blocks: Number of blocks used in leave-out cross-validation. Average rank indicates the comparative performance of methods in predicting validation sets behavior (based on the AD^2 goodness-of-fit test; lower is better); success rate (%) represents the percentage of successful predictions (P-value for $AD^{2*} \geq 0.05$). Methods: LSQ1–3 (2-parameter Weibull, least squares estimation using Eqs. (5)–(7) estimator), MLE-2P/3P (2-/3-parameter Weibull, maximum likelihood estimation), Norm (Normal), LogN (Lognormal).

outperformed other methods when the sample size was 30 or greater; however, at a sample size of 15, LSQ1 and Lognormal outperformed the Normal distribution.

Regarding the success rate, with a small sample size the success rates were not greater than 10 %. The best results were achieved when there were over 100 data in the TS. In CER, KAT, GRN, and ENA, the highest success rates were 15.8 % with LSQ1, 15.8 % with MLE-2P, 34.2 % with LSQ2, and 19.2 % with Normal distribution, respectively (Tables 3–6).

Table 5
Results of the average rank method and success rate criteria for different statistical approaches (KAT).

No. of blocks	Average rank							Success rate (%)						
	LSQ			MLE		Norm	LogN	LSQ			MLE		Norm	LogN
	1	2	3	2P	3P			1	2	3	2P	3P		
1	2.5	3.3	3.9	4.2	5.5	4.1	4.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	3.1	3.1	3.7	4.2	4.7	4.2	4.9	8.9	6.7	4.4	4.4	2.2	0.0	0.0
3	3.4	3.4	3.9	3.7	4.3	4.2	5.1	5.8	7.5	6.7	5.0	5.0	0.0	0.0
4	3.5	3.6	3.9	3.6	3.8	4.3	5.4	9.5	8.1	7.1	7.1	6.2	1.4	0.0
5	3.6	3.6	3.9	3.4	3.8	4.3	5.3	10.3	9.5	9.1	12.3	10.7	2.4	0.0
6	3.7	3.7	3.9	3.4	3.7	4.3	5.3	12.4	11.4	10.5	12.9	10.5	1.4	1.0
7	3.8	3.8	3.9	3.4	3.8	4.2	5.2	9.2	11.7	13.3	15.8	14.2	4.2	0.0

No. of blocks: Number of blocks used in leave-out cross-validation. Average rank indicates the comparative performance of methods in predicting validation sets behavior (based on the AD^2 goodness-of-fit test; lower is better); success rate (%) represents the percentage of successful predictions (P-value for $AD^{2*} \geq 0.05$). Methods: LSQ1–3 (2-parameter Weibull, least squares estimation using Eqs. (5)–(7) estimator), MLE-2P/3P (2-/3-parameter Weibull, maximum likelihood estimation), Norm (Normal), LogN (Lognormal).

Table 6
Results of the average rank method and success rate criteria for different statistical approaches (ENA).

No. of blocks	Average rank							Success rate (%)						
	LSQ			MLE		Norm	LogN	LSQ			MLE		Norm	LogN
	1	2	3	2P	3P			1	2	3	2P	3P		
1	1.3	3.4	5.5	5.2	6.0	3.6	3.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	3.1	4.5	5.6	4.5	5.0	2.8	2.5	0.0	0.0	0.0	0.0	2.2	2.2	2.2
3	3.7	4.6	5.8	4.0	4.2	2.7	3.0	0.0	0.8	0.8	0.0	2.5	4.2	2.5
4	4.0	5.0	6.0	3.6	3.6	2.8	3.1	2.9	3.3	2.9	2.9	5.7	10.5	7.6
5	4.2	5.1	6.0	3.5	3.3	2.7	3.2	2.8	3.2	2.4	4.0	9.1	12.7	11.1
6	4.4	5.1	5.9	3.4	3.1	2.8	3.2	7.6	6.2	5.7	7.6	11.4	13.8	11.0
7	4.4	5.2	5.9	3.2	3.0	2.9	3.4	13.3	13.3	11.7	9.2	17.5	19.2	16.7

No. of blocks: Number of blocks used in leave-out cross-validation. Average rank indicates the comparative performance of methods in predicting validation sets behavior (based on the AD^2 goodness-of-fit test; lower is better); success rate (%) represents the percentage of successful predictions (P-value for $AD^{2*} \geq 0.05$). Methods: LSQ1–3 (2-parameter Weibull, least squares estimation using Eqs. (5)–(7) estimator), MLE-2P/3P (2-/3-parameter Weibull, maximum likelihood estimation), Norm (Normal), LogN (Lognormal).

As sample size increases, sampling variability and the gap between methods both decline for all materials (Figs. 8–11).

In terms of lower tail predictions (Tables 7–10), when considering all materials and sample sizes, the LSQ1 method was found to provide the most conservative outlook. On the other hand, the Lognormal approach significantly overestimated the flexural strength at the lower tail. The remaining approaches fell somewhere between these two extremes. Both the Normal distribution and MLE-3P methods were overly optimistic, while the remaining approach for calculating the zero-threshold Weibull distribution was considered conservative and reliable.

4. Discussion

In general, the Normal and 2-parameter Weibull distributions (particularly the LSQ regression method) are preferred choices for statistical analysis. The extent of the success or failure of statistical analysis seems to be significantly influenced by the properties of CAD-CAM composites. Factors such as mechanics of fracture, inter-block variability, composition, and microstructural features of CAD-CAM composites can play a pivotal role. For instance, from a fracture mechanics standpoint, the poor performance of the Lognormal aligns with prior suggestions that, because it is a multiplicative model, it's counterintuitive to use it to predict fracture in materials where cumulative damage causes failure (Le et al., 2011).

Due to the unknown parent distribution of the experimental data, which is further complicated by inter-block heterogeneity and sampling variability, providing an unequivocal explanation is challenging. However, the following discussion offers possible hypotheses linking CAD-CAM composite properties to the results of the statistical analysis.

Inter-block variability, as shown by Eldafrawy et al. (2023), might be an important factor affecting the outcome of statistical analysis. Specifically, (1) poor performance of MLE-3P at small sample sizes; and (2) better generalizability of LSQ1 over other LSQ methods might stem from inter-block variability.

Regarding the first point (the effect of inter-block variability on MLE-3P performance), with limited sample size, there can be overfitting of flexible distributions such as MLE-3P. Overfitting occurs when a flexible statistical model is so finely tuned to the TS that it fails to consider even subtle differences between TS and VS (caused by inter-block variability). In other words, MLE-3P could not generalize well to new, unseen data. By increasing the size of TS, which provides a more accurate representation of material properties, the overfitting issue of MLE-3P was circumvented to some extent (Tables 3–6). As a concluding remark, it can be suggested not to use the 3-parameter Weibull distribution in sample sizes smaller than 20 (Abernethy, 2006; Roos and Stawarczyk, 2012), and indeed, to exercise caution even when using more than 20 samples.

Regarding the second point about the effect of inter-block variability on optimal estimator for LSQ method, typically, the mean rank estimator, $(i/(n+1))$, performs better than other estimators (Tables 3–6). This finding contrasts with multiple statistical studies, which often, through Monte Carlo simulations, found that the Hazen rank or median rank estimator are better choices (Khalili and Kromp, 1991; Wu et al., 2001, 2006b). The plausible reason is that Monte Carlo simulations draw samples from one specified parent population, while CAD-CAM composites blocks can have distinct distributions. Thus, using the mean rank estimator, which results in a lower Weibull modulus - suggest higher scatter in data - can better capture the inter-block variability, leading to improved prediction.

Besides inter-block heterogeneity, another factor is the inconsistency between the properties of CAD-CAM composites and the physical assumptions under which material scatter follows the Weibull distribution. These violations include plastic deformation and multimodal flaw

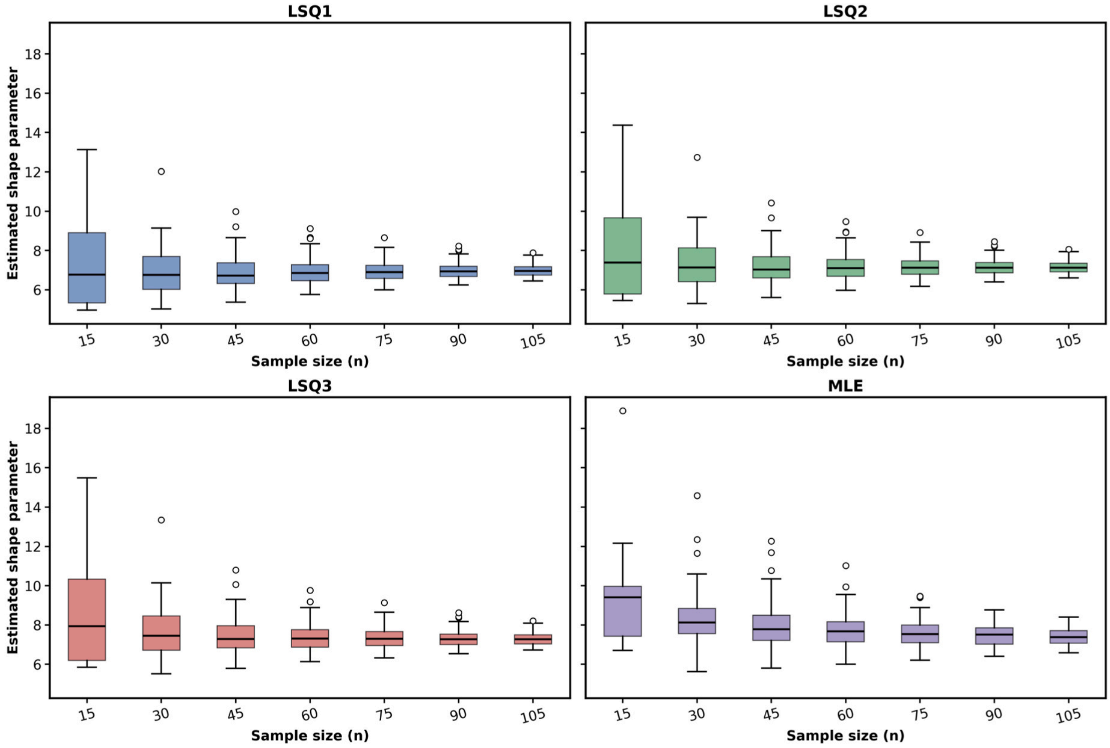

Fig. 8. Cerasmart 270; box plots of calculated 2-parameter Weibull modulus across different training set sample sizes, obtained by four estimation methods. LSQ1: least squares with mean estimator, $i/(n+1)$; LSQ2: least squares with median estimator, $(i-0.3)/(n+0.4)$; LSQ3: least squares with Hazen estimator, $(i-0.5)/n$; and MLE: maximum likelihood estimation.

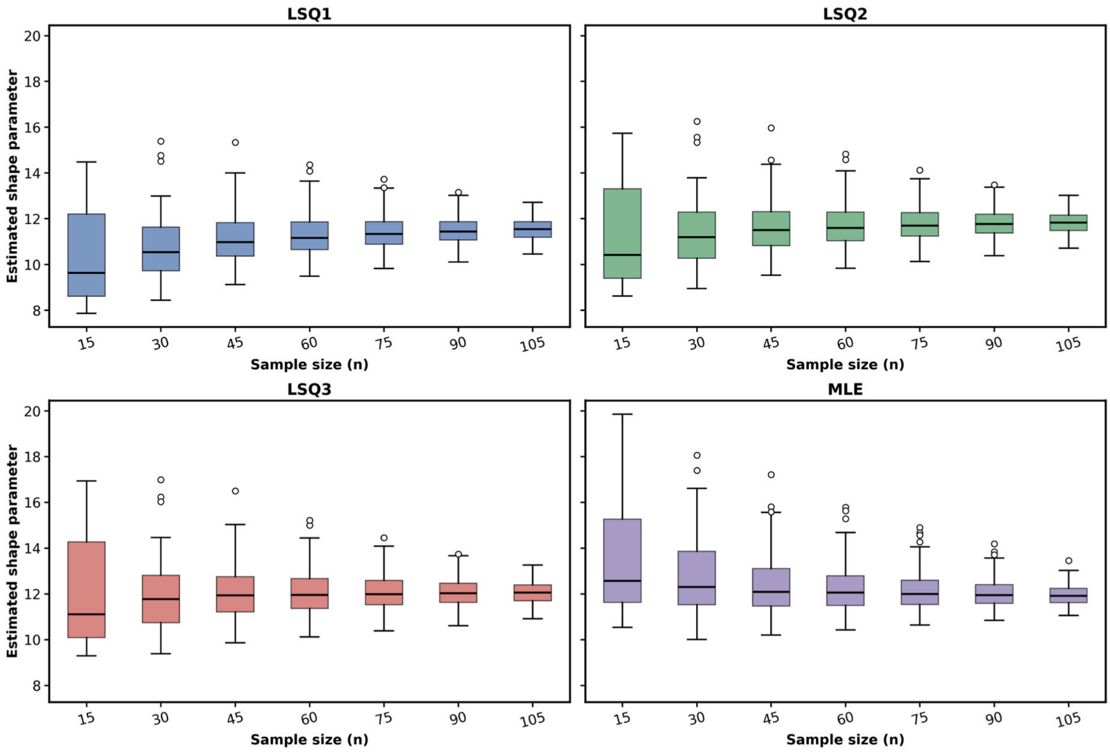

Fig. 9. Grandio Voco; box plots of calculated 2-parameter Weibull modulus across different training set sample sizes, obtained by four estimation methods. LSQ1: least squares with mean estimator, $i/(n+1)$; LSQ2: least squares with median estimator, $(i-0.3)/(n+0.4)$; LSQ3: least squares with Hazen estimator, $(i-0.5)/n$; and MLE: maximum likelihood estimation.

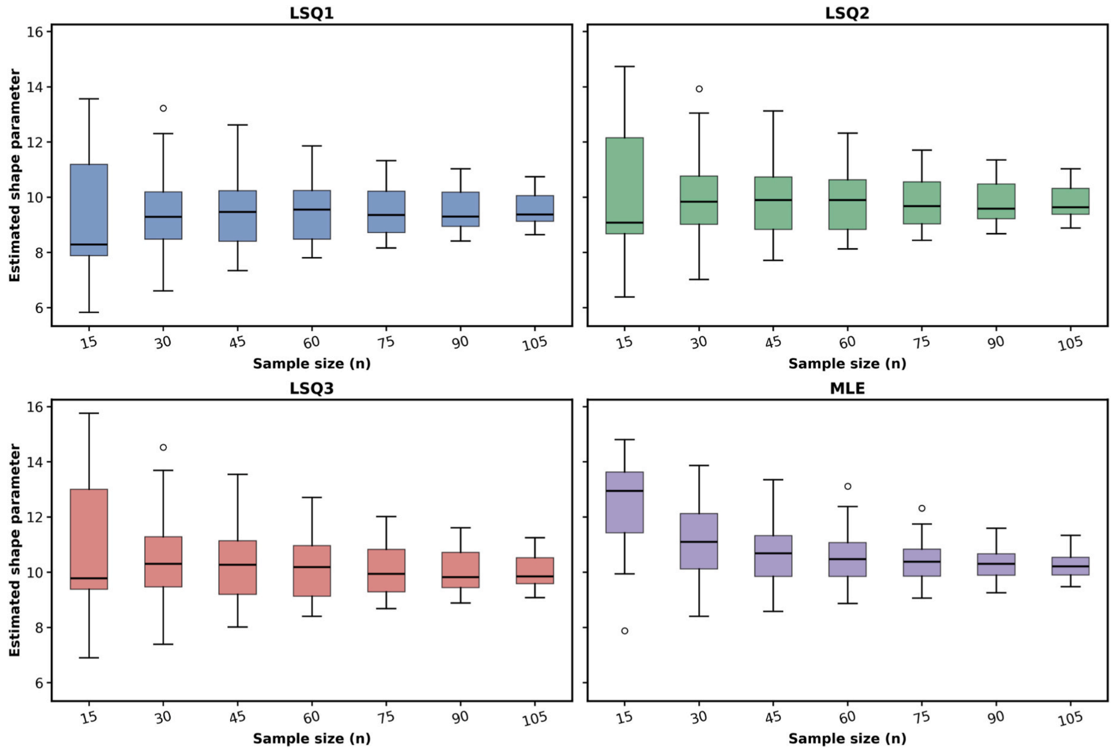

Fig. 10. Katana Avencia; box plots of calculated 2-parameter Weibull modulus across different training set sample sizes, obtained by four estimation methods. LSQ1: least squares with mean estimator, $i/(n+1)$; LSQ2: least squares with median estimator, $(i-0.3)/(n+0.4)$; LSQ3: least squares with Hazen estimator, $(i-0.5)/n$; and MLE: maximum likelihood estimation.

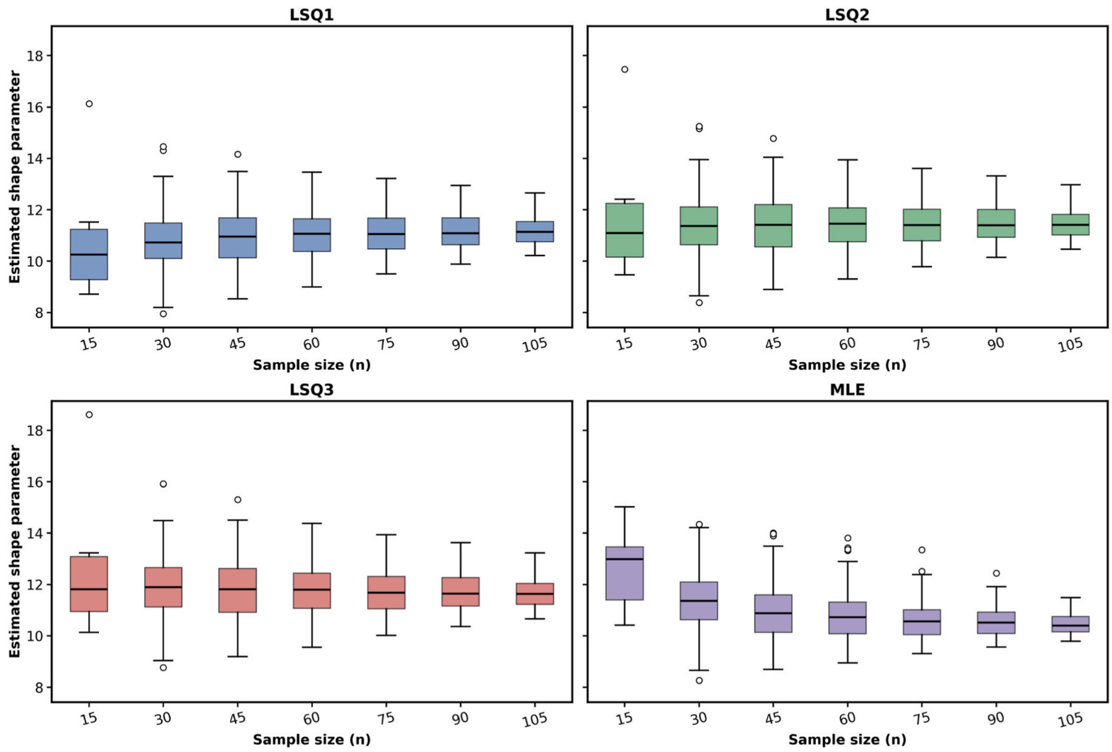

Fig. 11. Vita Enamic; box plots of calculated 2-parameter Weibull modulus across different training set sample sizes, obtained by four estimation methods. LSQ1: least squares with mean estimator, $i/(n+1)$; LSQ2: least squares with median estimator, $(i-0.3)/(n+0.4)$; LSQ3: least squares with Hazen estimator, $(i-0.5)/n$; and MLE: maximum likelihood estimation.

Table 7
The proportion of overoptimistic prediction for the lower tail flexural strength (with a 1 % failure probability, CER).

No. of blocks	LSQ1	LSQ2	LSQ3	MLE-2P	MLE-3P	Norm	LogN
1	20.0	20.0	20.0	40.0	50.0	50.0	60.0
2	6.67	8.89	11.1	15.6	28.9	35.6	46.7
3	5.00	5.83	8.33	15.8	30.0	35.8	50.9
4	2.38	2.86	5.24	8.57	21.0	25.7	41.0
5	2.38	3.97	6.75	9.13	19.4	24.6	39.3
6	0.95	1.90	2.86	2.86	11.0	15.7	29.5
7	0.83	1.67	2.50	3.33	6.67	33.3	53.3

No. of blocks: The number of blocks used as the training set in a leave-out cross-validation method. LSQ1 to LSQ3 denote least squares regression analysis of two-parameter Weibull distribution through Eqs. (5)–(7), respectively. MLE-2P and MLE-3P are maximum likelihood estimation of 2-parameters and 3 parameters Weibull distribution. Norm and LogN are Normal and lognormal distribution, respectively.

Table 8
The proportion of overoptimistic prediction for the lower tail flexural strength (with a 1 % failure probability, GRN).

No. of blocks	LSQ1	LSQ2	LSQ3	MLE-2P	MLE-3P	Norm	LogN
1	0.00	10.0	20.0	30.0	50.0	60.0	60.0
2	0.00	4.44	6.67	8.89	31.1	40.0	48.9
3	1.67	3.33	7.50	7.50	20.8	54.2	59.2
4	0.48	1.43	3.33	1.43	12.4	38.6	46.7
5	0.79	1.59	2.78	1.19	19.0	48.8	54.8
6	0.00	0.48	0.95	0.00	10.0	33.8	44.3
7	0.83	1.67	2.50	1.67	17.5	45.8	49.2

No. of blocks: The number of blocks used as the training set in a leave-out cross-validation method. LSQ1 to LSQ3 denote least squares regression analysis of two-parameter Weibull distribution through Eqs. (5)–(7), respectively. MLE-2P and MLE-3P are maximum likelihood estimation of 2-parameters and 3 parameters Weibull distribution. Norm and LogN are Normal and lognormal distribution, respectively.

Table 9
The proportion of overoptimistic prediction for the lower tail flexural strength (with a 1 % failure probability, KAT).

No. of blocks	LSQ1	LSQ2	LSQ3	MLE-2P	MLE-3P	Norm	LogN
1	20.0	30.0	30.0	30.0	40.0	50.0	60.0
2	6.67	6.67	8.89	11.1	22.2	53.3	55.6
3	4.17	5.83	10.8	6.67	23.3	54.2	58.3
4	1.43	3.81	5.71	1.90	13.8	41.4	52.9
5	4.76	7.54	9.13	7.14	19.0	44.0	56.3
6	1.90	3.33	4.76	2.86	11.4	33.3	46.7
7	5.83	8.33	12.5	7.50	15.0	43.3	55.8

No. of blocks: The number of blocks used as the training set in a leave-out cross-validation method. LSQ1 to LSQ3 denote least squares regression analysis of two-parameter Weibull distribution through Eqs. (5)–(7), respectively. MLE-2P and MLE-3P are maximum likelihood estimation of 2-parameters and 3 parameters Weibull distribution. Norm and LogN are Normal and lognormal distribution, respectively.

populations (Choi et al., 2019; Ducke and Ilie, 2021; Ghelbere and Ilie, 2022). Thus, it is expected that the parent distribution of flexural strength data does not fully conform to the Weibull distribution. As a result, even with a resource-intensive sample size of over 100, the top success rates remain relatively low, ranging from only 16 %–34 % (Tables 3–6). This limited generalizability, despite the reduced sampling variability (Figs. 8–11) can be attributed to the inconsistency of Weibull distribution assumptions with the properties of CAD-CAM composites.

It should be added that LSQ typically outperforms MLE with small samples. This finding is unsurprising, given that Monte Carlo simulations using m values representative of dental materials ($m > 3.5$, left-skewed) provide similar insight. It has been suggested that MLE only surpasses LSQ when sample sizes exceed ~20 (Trustrum and Jayatilaka, 1979). Bütikofer et al. also found that even for $n \geq 30$, LSQ often performs better than MLE, though their performance converges as sample size increases (Bütikofer et al., 2015). Of note, many studies have used 2-parameter Weibull distributions with $m < 3.5$, possibly because a wide range of phenomena tend to have such m values (Abubakar et al., 2024; Plana et al., 2022; Wais, 2017; Wendler et al., 2018). Since the performance of statistical methods depends on the underlying population parameters (Chu and Ke, 2012), insights from these studies should be interpreted with caution.

Table 10
The proportion of overoptimistic prediction for the lower tail flexural strength (with a 1 % failure probability, ENA).

No. of blocks	LSQ1	LSQ2	LSQ3	MLE-2P	MLE-3P	Norm	LogN
1	0.00	10.00	10.00	10.00	50.0	40.0	70.0
2	0.00	0.00	0.00	0.00	35.6	40.0	55.5
3	0.00	0.00	0.00	0.00	35.8	36.7	56.7
4	0.00	0.00	0.00	0.00	29.1	27.1	48.6
5	0.00	0.00	0.00	0.00	26.2	27.4	47.6
6	0.00	0.00	0.00	0.00	19.5	20.0	39.5
7	0.83	1.67	4.17	0.00	11.7	18.3	30.0

No. of blocks: The number of blocks used as the training set in a leave-out cross-validation method. LSQ1 to LSQ3 denote least squares regression analysis of two-parameter Weibull distribution through Eqs. (5)–(7), respectively. MLE-2P and MLE-3P are maximum likelihood estimation of 2-parameters and 3 parameters Weibull distribution. Norm and LogN are Normal and lognormal distribution, respectively.

Microstructure also seems to play a significant role. While DF composites flexural strength data are better aligned with the 2P-Weibull distribution, the flexural strength data of ENA materials with a PICN microstructure follow the Normal distribution for sample sizes of 30 or more (Table 6). This can be related to the quasi-brittle nature of ENA.

Quasi-brittle materials, whose microstructural inhomogeneities are not negligible compared to the specimen size, require a fracture model different from that used for perfectly brittle materials (Le and Bažant, 2009). The classical Weibull model assumes an infinite chain of statistically independent links, where failure is governed by the weakest link (Le and Bažant, 2009). However, quasi-brittle structures are characterized by a finite number of such links (Le and Bažant, 2009). Each link corresponds to a representative volume element, which is determined by the material's microstructure (Le and Bažant, 2009).

Bažant established that the strength of an individual representative volume element in quasi-brittle materials is approximately Normally distributed (Bažant et al., 2009). For small specimens containing only a few representative volume elements, the overall strength distribution remains nearly Normal (Bažant, 2019). However, as the specimen size increases and the number of representative volume elements exceeds approximately 10^4, the strength distribution converges toward the classical Weibull distribution (Bažant and Pang, 2006).

At intermediate specimen sizes, neither the Normal nor the Weibull distribution alone fully captures the strength behavior (Le and Bažant, 2009). The left segment (lower tail) of Weibull probability plot follows a Weibull distribution, while the right segment conforms with a Normal distribution (Le and Bažant, 2009). This issue results in a kink in the Weibull probability plot, as illustrated in Fig. 12 (Le and Bažant, 2009). The location of this grafting point between can shift depending on the material properties and specimen dimensions (Le and Bažant, 2009). This determines which distribution provides a better fit. In this study, ENA seems to align more closely with a Normal than with a 2P-Weibull distribution. Moreover, only ENA among the studied composites showed a kink in Weibull plots of TS with sample sizes of 75 or more.

The above-mentioned discussions compared the performance of different distributions in explaining data scattered across the entire range. However, in practice, the primary concern is the dependable

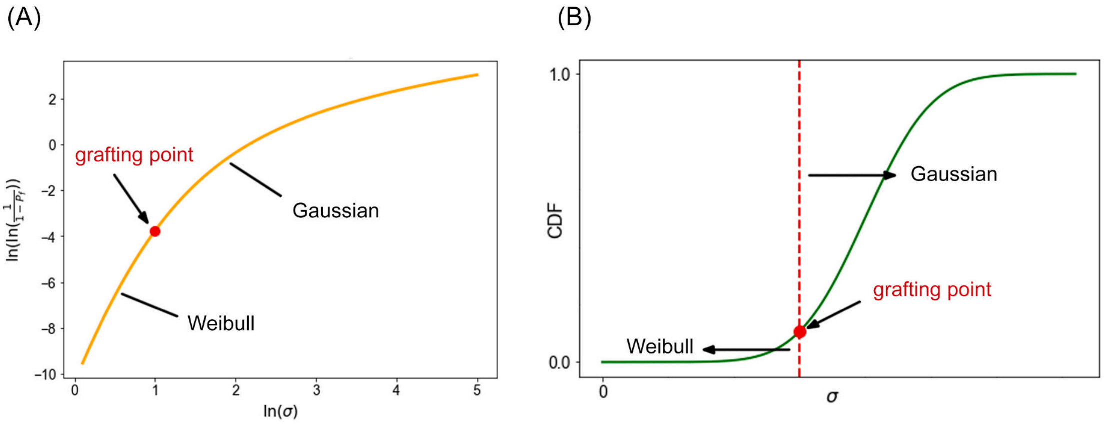

Fig. 12. The Weibull distribution probability plot (A) and cumulative density function (B) for quasi-brittle materials (Bažant, 2019).

estimation of strength at a low failure probability. LSQ1 consistently reduces unrealistic optimism, even with small sample sizes, outperforming all other methods. As sample sizes increase, other statistical methods for estimating 2P-Weibull parameters become conservative, but Lognormal, Normal, and 3P-Weibull remain overly optimistic. This demonstrates that the zero-threshold Weibull's conservative nature is especially needed for DF composites (Tables 7–10). In ENA, using either Normal or 2P-Weibull distribution can potentially lead to excessive overestimation or underestimation of flexural strength at low failure probability, respectively (contrary to DF composites, 2P-Weibull rarely overestimates ENA; Table 10). This observation is expected in quasi-brittle materials (Le and Bažant, 2009).

As discussed, different parameters such as sample size and CAD-CAM material properties affect which statistical distribution has the best performance. However, if the authors are restricted to choose a statistical method, LSQ1 seems to be the most effective one (Table 11). Moreover, the findings of the present study lend support to the recommendation that more than 100 samples are needed for reliable Weibull analysis (Barsoum, 2019; Nohut, 2014). Yet, given the vast array of materials, testing 100 samples may be impractical due to considerable time and cost. Thus, researchers might prefer testing fewer samples (e.g., 4 or 5, standalone performance in Tables 3–6), but testing 30 samples, which is common practice in dentistry, seems insufficient.

To the best of the authors' knowledge, this is the first study to use a large real dataset and the LKO-CV method to provide statistical recommendations for more reliable analysis. Therefore, addressing the limitations and understudied topics in this work is important for paving the way for future studies. The future perspectives of this work can possibly focus on two different aspects.

Table 11
The most successful method in each criterion for small (n = 15, 30), medium (n = 45 or 60), and large (n = 75, 90, 105) samples. Performance is determined based on taking the average of performance over a specified size.

Material	Data size	Average rank	Success rate	Lower tail outlook
CER	Small	LSQ1	LSQ1	LSQ1
	Medium	LSQ1	LSQ2	LSQ1
	Large	LSQ1	LSQ1	LSQ1
GRA	Small	LSQ1, LSQ2, LSQ3	LSQ3	LSQ1
	Medium	LSQ2	MLE-2P	LSQ1
	Large	LSQ1	LSQ2	LSQ1
KAT	Small	LSQ1	LSQ1	LSQ1
	Medium	LSQ1,LSQ2	LSQ1, LSQ2	LSQ1
	Large	MLE-2P	MLE-2P	LSQ1
ENA	Small	LSQ1	Norm, MLE-3P, LogN	LSQ1
	Medium	Norm	Norm	LSQ1, LSQ2, LSQ3, MLE-2P
	Large	Norm	Norm	MLE-2P

LSQ regression: Least-squares regression analysis; MLE: Maximum likelihood estimation.

Alternative statistical methods for estimating distribution parameters, such as the method of moments, should be explored. Additionally, variants of the Weibull distribution, such as the bimodal Weibull, as well as other relevant statistical distributions, may enhance the reliability of the analysis (Nohut and Lu, 2012; Orlovskaja et al., 2000). For instance, it has been shown that, under similar assumptions, the distribution of material strength depends on the flaw population: if the flaw population follows an inverse power-law distribution, the strength follows Weibull statistics. However, if the flaw population follows an exponential distribution, the strength is described by the Gumbel distribution (Alava et al., 2006).

Another area for improvement is testing methodology. This is because the testing protocol can influence statistical reliability. For the use of the Weibull distribution to be physically justified, a tensile stress-based fracture criterion must be met, and the testing configuration should produce a minimal stress gradient (Danzer et al., 2007; Lei et al., 2020). The 4-point bending test better fulfills this requirement than the 3-point test and, in addition, subjects a larger effective volume of the sample to maximum stress. This provides a more representative measure of the material's properties (Ilie et al., 2017). Additionally, if possible, chamfering the edges is recommended to reduce edge effects.

The practical significance and impact of statistical errors also need to be compared with other sources of experimental error in strength measurement, such as friction and twisting (Quinn and Morrell, 1991). Other objectives of future studies are to apply the same methodology to dental ceramics, light-cured composites, and 3D-printed composites.

At last, the authors wish to emphasize that Weibull distribution and Weibull fracture theory are not equivalent (Lei et al., 2020; Zok, 2017). Therefore, improving the reliability of statistical analysis does not justify applying the Weibull size-scaling relationship. The scaling law can be used if the material meets the underlying assumptions of Weibull fracture theory, with parameters determined accordingly (i.e., by testing materials under different configurations and using a Weibull phenomenological model) (Lei et al., 2020).

5. Conclusion

In practice, factors such as inter-block heterogeneity and the physical properties of computer-aided design/computer-aided manufacturing (CAD-CAM) composites can influence how well statistical methods perform. The following conclusions can be drawn from the present study:

- Least square (LSQ) estimation with mean estimation $(i/n+1)$ of the cumulative distribution function can be considered as the best choice among different calculation methods.
- Still, all evaluated statistical methods had limited reliability when applied to small sample sizes (e.g., n = 30) commonly used in dental materials research. Therefore, a sample size of more than 60 seems crucial for more dependable statistical analysis, and if possible, more than 100 tests should be conducted.
- For dispersed filler composites, 2-parameter Weibull distribution is most successful method in predicting unseen data to other alternatives for describing the flexural strength of composites when using LSQ regression.
- For Vita Enamic (ENA), Normal distribution can better describe the scatter of flexural strength; however, it overestimates ENA's strength at low failure probability.
- LSQ principle is generally providing better generalizability compared to maximum likelihood estimation.
- The Lognormal distribution is not recommended for statistical analysis CAD-CAM composites.

CRediT authorship contribution statement

Yousef Karevan: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Christelle Sanchez:** Writing – review & editing, Supervision. **Adelin Albert:** Writing – review & editing, Validation. **Amélie Mainjot:** Writing – review & editing, Validation, Supervision, Project administration.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors used ChatGPT to improve readability and language. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Declaration of competing Interest

The authors received no financial support for this work. Amélie Mainjot is married to the founder of the company MaJEB, which contributes to the development of PICN materials. The authors declare no other conflict of interest with respect to the authorship and/or publication of this article.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jmbbm.2025.107171.

Data availability

Data will be made available on request.

References

Abernethy, R.B., 2006. The new Weibull handbook: reliability & statistical analysis for predicting life, safety, risk, support costs, failures, and forecasting warranty claims, substantiation and accelerated testing, using Weibull, log normal. In: Abernethy, R. B. (Ed.), Crow-AMSAA, Probit, and Kaplan- Meier Models, 5. Fla, North Palm Beach.

Abubakar, H., Misiran, M., Sayed, A.A.I., Karaye, A.B., 2024. Optimization of weibull distribution parameters with application to short-term risk assessment and strategic investment decision-making. Stat. Optim. Inf. Comput. 12, 1684–1709. https://doi.org/10.19139/soic-2310-5070-2099.

Alava, M.J., Nukala, P.K.V.V., Zapperi, S., 2006. Statistical models of fracture. Adv. Phys. 55, 349–476. https://doi.org/10.1080/00018730300741518.

Anderson, T.W., Darling, D.A., 1954. A test of goodness-of-fit. J. Am. Stat. Assoc. 49, 765–769.

Aslam, M., 2021. A new goodness of fit test in the presence of uncertain parameters. Complex Intell. Syst. 7, 359–365. https://doi.org/10.1007/s40747-020-00214-8.

Barsoum, M., 2019. Fundamentals of Ceramics. CRC press.

Basu, B., Tiwari, D., Kundu, D., Prasad, R., 2009. Is Weibull distribution the most appropriate statistical strength distribution for brittle materials? Ceram. Int. 35, 237–246. https://doi.org/10.1016/j.ceramint.2007.10.003.

Bažant, Z.P., 2019. Design of quasibrittle materials and structures to optimize strength and scaling at probability tail: an apercu. Proc. R. Soc. Math. Phys. Eng. Sci. 475, 20180617. https://doi.org/10.1098/rspa.2018.0617.

Bažant, Z.P., Jia-Liang, L., Bazant, M.Z., 2009. Atomistic fracture and nano-macro transition for strength and lifetime statistics of quasibrittle structures. In: Presented at the 12th International Conference on Fracture. Ottawa Canada.

Bažant, Z.P., Pang, S.-D., 2006. Mechanics-based statistics of failure risk of quasibrittle structures and size effect on safety factors. Proc. Natl. Acad. Sci. 103, 9434–9439. https://doi.org/10.1073/pnas.0602684103.

Belli, R., Lohbauer, U., 2021. The breakdown of the Weibull behavior in dental zirconias. J. Am. Ceram. Soc. 104, 4819–4828. https://doi.org/10.1111/jace.17886.

Benard, A., Bos-Levenbach, E.C.B., 1953. The plotting of observations on probability paper. Statistica 7, 163–173.

Brazdil, P.B., Soares, C., 2000. A comparison of ranking methods for classification algorithm selection. In: López De Mántaras, R., Plaza, E. (Eds.), Machine Learning: ECML 2000, Lecture Notes in Computer Science. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 63–75. https://doi.org/10.1007/3-540-45164-1_8.

Bütikofer, L., Stawarczyk, B., Roos, M., 2015. Two regression methods for estimation of a two-parameter Weibull distribution for reliability of dental materials. Dent. Mater. 31, e33–e50. https://doi.org/10.1016/j.dental.2014.11.014.

Chavali, R., Nejat, A.H., Lawson, N.C., 2017. Machinability of CAD-CAM materials. J. Prosthet. Dent 118, 194–199. https://doi.org/10.1016/j.prosdent.2016.09.022.

Choi, B.-J., Yoon, S., Im, Y.-W., Lee, J.-H., Jung, H.-J., Lee, H.-H., 2019. Uniaxial/biaxial flexure strengths and elastic properties of resin-composite block materials for CAD/CAM. Dent. Mater 35, 389–401. https://doi.org/10.1016/j.dental.2018.11.032.

Chu, Y.-K., Ke, J.-C., 2012. Computation approaches for parameter estimation of weibull distribution. Math. Comput. Appl. 17, 39–47. https://doi.org/10.3390/mca17010039.

Coldea, A., Fischer, J., Swain, M.V., Thiel, N., 2015. Damage tolerance of indirect restorative materials (including PICN) after simulated bur adjustments. Dent. Mater. 31, 684–694. https://doi.org/10.1016/j.dental.2015.03.007.

Curran, P., Cattani-Lorente, M., Anselm Wiskott, H.W., Durual, S., Scherrer, S.S., 2017. Grinding damage assessment for CAD-CAM restorative materials. Dent. Mater. 33, 294–308. https://doi.org/10.1016/j.dental.2016.12.004.

D'Agostino, R.B., Stephens, M.A., 1986. Goodness-of-Fit-Techniques. MARCEL DEKKER, INC.

Danzer, R., Supancic, P., Pascual, J., Lube, T., 2007. Fracture statistics of ceramics – weibull statistics and deviations from Weibull statistics. Eng. Fract. Mech. 74, 2919–2932. https://doi.org/10.1016/j.engfracmech.2006.05.028.

Datsiou, K.C., Overend, M., 2018. Weibull parameter estimation and goodness-of-fit for glass strength data. Struct. Saf. 73, 29–41. https://doi.org/10.1016/j.strusafe.2018.02.002.

Ducke, V.M., Ilie, N., 2021. Aging behavior of high-translucent CAD/CAM resin-based composite blocks. J. Mech. Behav. Biomed. Mater. 115, 104269. https://doi.org/10.1016/j.jmbbm.2020.104269.

Eldafrawy, M., Karevan, Y., Nguyen, J.-F., Mainjot, A., 2023. Interblock and intrablock homogeneity of CAD-CAM composites mechanical properties. Dent. Mater. J. 42, 780–790. https://doi.org/10.4012/dmj.2023-090.

Ghelbere, R., Ilie, N., 2022. Prediction of uniaxial and biaxial flexural strengths of resin-based composites using the Weibull model. J. Mech. Behav. Biomed. Mater. 131, 105231. https://doi.org/10.1016/j.jmbbm.2022.105231.

Han, Z., Tang, L.C., Xu, J., Li, Y., 2009. A three-parameter Weibull statistical analysis of the strength variation of bulk metallic glasses. Scr. Mater. 61, 923–926. https://doi.org/10.1016/j.scriptamat.2009.07.038.

Hazen, A., 1914. Storage to be provided in impounding reservoirs for municipalwater supply. Trans. Am. Soc. Civ. Eng. 77, 1547–1550.

Ilie, N., Hilton, T.J., Heintze, S.D., Hickel, R., Watts, D.C., Silikas, N., Stansbury, J.W., Cadenaro, M., Ferracane, J.L., 2017. Academy of dental materials guidance—resin composites: Part I—mechanical properties. Dent. Mater. 33, 880–894. https://doi.org/10.1016/j.dental.2017.04.013.

International Organization for Standardization, 2015. ISO 6872: 2015. Dentistry—Ceramic Materials.

Khalili, A., Kromp, K., 1991. Statistical properties of Weibull estimators. J. Mater. Sci. 26, 6741–6752.

Le, J.-L., Bažant, Z.P., 2009. Strength distribution of dental restorative ceramics: finite weakest link model with zero threshold. Dent. Mater. 25, 641–648. https://doi.org/10.1016/j.dental.2008.11.011.

Le, J.-L., Bažant, Z.P., Bazant, M.Z., 2011. Unified nano-mechanics based probabilistic theory of quasibrittle and brittle structures: I. Strength, static crack growth, lifetime and scaling. J. Mech. Phys. Solid. 59, 1291–1321. https://doi.org/10.1016/j.jmps.2011.03.002.

Lebon, N., Tapie, L., Vennat, E., Mawussi, B., 2015. Influence of CAD/CAM tool and material on tool wear and roughness of dental prostheses after milling. J. Prosthet. Dent 114, 236–247. https://doi.org/10.1016/j.prosdent.2014.12.021.

Lei, W.-S., Zhang, P., Yu, Z., Qian, G., 2020. Statistics of ceramic strength: use ordinary Weibull distribution function or Weibull statistical fracture theory? Ceram. Int. 46, 20751–20768. https://doi.org/10.1016/j.ceramint.2020.05.024.

Liu, X., Yao, X., Zhang, R., Sun, L., Zhang, Z., Zhao, Y., Zhang, T., Yan, J., Zhang, Y., Wu, X., Li, B., 2023. Recent advances in glass-ceramics: performance and toughening

mechanisms in restorative dentistry. J. Biomed. Mater. Res. B Appl. Biomater. jbm.b. 35334. https://doi.org/10.1002/jbm.b.35334.
Lohbauer, U., Petschelt, A., Greil, P., 2002. Lifetime prediction of CAD/CAM dental ceramics. J. Biomed. Mater. Res. 63, 780–785. https://doi.org/10.1002/jbm.10468.
Mainjot, A.K., Dupont, N.M., Oudkerk, J.C., Dewael, T.Y., Sadoun, M.J., 2016. From artisanal to CAD-CAM blocks: state of the art of indirect composites. J. Dent. Res. 95, 487–495. https://doi.org/10.1177/0022034516634286.
Nohut, S., 2021. Three-parameter (3P) weibull distribution for characterization of strength of ceramics showing R-Curve behavior. Ceram. Int. 47, 2270–2279. https://doi.org/10.1016/j.ceramint.2020.09.067.
Nohut, S., 2014. Influence of sample size on strength distribution of advanced ceramics. Ceram. Int. 40, 4285–4295. https://doi.org/10.1016/j.ceramint.2013.08.093.
Nohut, S., Lu, C., 2012. Fracture statistics of dental ceramics: discrimination of strength distributions. Ceram. Int. 38, 4979–4990. https://doi.org/10.1016/j.ceramint.2012.02.093.
Orlovskaja, N., Peterlik, H., Steinkellner, W., Kromp, K., 2000. Prediction of strength of recrystallized siliconcarbide from pore size measurement Part I the bimodality of the distribution. J. Mater. Sci. 35, 699–705. https://doi.org/10.1023/a:1004705300886.
Oudkerk, J., Herman, R., Eldafrawy, M., Wulfman, C., Ernst, M., Vanheusden, A., Mainjot, A., 2024. Intraoral wear of PICN CAD-CAM composite restorations used in severe tooth wear treatment: 5-year results of a prospective clinical study using 3D profilometry. Dent. Mater. 40, 1056–1063. https://doi.org/10.1016/j.dental.2024.05.015.
Pfeilschifter, M., Preis, V., Behr, M., Rosentritt, M., 2018. Edge strength of CAD/CAM materials. J. Dent. 74, 95–100. https://doi.org/10.1016/j.jdent.2018.05.004.
Plana, D., Fell, G., Alexander, B.M., Palmer, A.C., Sorger, P.K., 2022. Cancer patient survival can be parametrized to improve trial precision and reveal time-dependent therapeutic effects. Nat. Commun. 13. https://doi.org/10.1038/s41467-022-28410-9.
Quinn, G.D., Morrell, R., 1991. Design data for engineering ceramics: a review of the flexure test. J. Am. Ceram. Soc. 74, 2037–2066. https://doi.org/10.1111/j.1151-2916.1991.tb08259.x.
Quinn, J.B., Quinn, G.D., 2010. A practical and systematic review of Weibull statistics for reporting strengths of dental materials. Dent. Mater. 26, 135–147. https://doi.org/10.1016/j.dental.2009.09.006.
Reid, M., 2020. MatthewReid854/reliability: v0.5.1. https://doi.org/10.5281/ZENODO.3938000.
Roos, M., Stawarczyk, B., 2012. Evaluation of bond strength of resin cements using different general-purpose statistical software packages for two-parameter Weibull statistics. Dent. Mater. 28, e76–e88. https://doi.org/10.1016/j.dental.2012.04.013.
The pandas development team, 2024. pandas-dev/pandas: pandas. https://doi.org/10.5281/ZENODO.10537285.
Tiryakioğlu, M., Hudak, D., 2007. On estimating Weibull modulus by the linear regression method. J. Mater. Sci. 42, 10173–10179. https://doi.org/10.1007/s10853-007-2060-5.
Trustrum, K., Jayatilaka, A.D.S., 1979. On estimating the Weibull modulus for a brittle material. J. Mater. Sci. 14, 1080–1084. https://doi.org/10.1007/BF00561290.
Van Den Born, I.C., Santen, A., Hoekstra, H.D., De Hosson, J.ThM., 1991. Mechanical strength of highly porous ceramics. Phys. Rev. B 43, 3794–3796. https://doi.org/10.1103/PhysRevB.43.3794.
Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., Van Der Walt, S.J., Brett, M., Wilson, J., Millman, K.J., Mayorov, N., Nelson, A.R.J., Jones, E., Kern, R., Larson, E., Carey, C.J., Polat, İ., Feng, Y., Moore, E.W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E.A., Harris, C.R., Archibald, A.M., Ribeiro, A. H., Pedregosa, F., Van Mulbregt, P., , SciPy 1.0 Contributors, Vijaykumar, A., Bardelli, A.P., Rothberg, A., Hilboll, A., Kloeckner, A., Scopatz, A., Lee, A., Rokem, A., Woods, C.N., Fulton, C., Masson, C., Häggström, C., Fitzgerald, C., Nicholson, D.A., Hagen, D.R., Pasechnik, D.V., Olivetti, E., Martin, E., Wieser, E., Silva, F., Lenders, F., Wilhelm, F., Young, G., Price, G.A., Ingold, G.-L., Allen, G.E., Lee, G.R., Audren, H., Probst, I., Dietrich, J.P., Silterra, J., Webber, J.T., Slavič, J., Nothman, J., Buchner, J., Kulick, J., Schönberger, J.L., De Miranda Cardoso, J.V., Reimer, J., Harrington, J., Rodríguez, J.L.C., Nunez-Iglesias, J., Kuczynski, J., Tritz, K., Thoma, M., Newville, M., Kümmerer, M., Bolingbroke, M., Tartre, M., Pak, M., Smith, N.J., Nowaczyk, N., Shebanov, N., Pavlyk, O., Brodtkorb, P.A., Lee, P., McGibbon, R.T., Feldbauer, R., Lewis, S., Tygier, S., Sievert, S., Vigna, S., Peterson, S., More, S., Pudlik, T., Oshima, T., Pingel, T.J., Robitaille, T.P., Spura, T., Jones, T.R., Cera, T., Leslie, T., Zito, T., Krauss, T., Upadhyay, U., Halchenko, Y.O., Vázquez-Baeza, Y., 2020. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat. Methods 17, 261–272. https://doi.org/10.1038/s41592-019-0686-2.
Wais, P., 2017. A review of Weibull functions in wind sector. Renew. Sustain. Energy Rev. 70, 1099–1107. https://doi.org/10.1016/j.rser.2016.12.014.
Weibull, W., 1939. A statistical theory of the strength of materials. Ingeniörvetenskapsakademiens Handl 151.
Wendler, M., Belli, R., Petschelt, A., Mevec, D., Harrer, W., Lube, T., Danzer, R., Lohbauer, U., 2017. Chairside CAD/CAM materials. Part 2: flexural strength testing. Dent. Mater. 33, 99–109. https://doi.org/10.1016/j.dental.2016.10.008.
Wendler, M., Belli, R., Valladares, D., Petschelt, A., Lohbauer, U., 2018. Chairside CAD/CAM materials. Part 3: cyclic fatigue parameters and lifetime predictions. Dent. Mater. 34, 910–921. https://doi.org/10.1016/j.dental.2018.03.024.
Wu, D., Zhou, J., Li, Y., 2006a. Methods for estimating Weibull parameters for brittle materials. J. Mater. Sci. 41, 5630–5638. https://doi.org/10.1007/s10853-006-0344-9.
Wu, D., Zhou, J., Li, Y., 2006b. Unbiased estimation of Weibull parameters with the linear regression method. J. Eur. Ceram. Soc. 26, 1099–1105. https://doi.org/10.1016/j.jeurceramsoc.2005.01.044.
Wu, Dongfang, Li, Y., Zhang, J., Chang, L., Wu, Dihua, Fang, Z., Shi, Y., 2001. Effects of the number of testing specimens and the estimation methods on the Weibull parameters of solid catalysts. Chem. Eng. Sci. 56, 7035–7044. https://doi.org/10.1016/S0009-2509(01)00340-2.
Xu, Y., Cheng, L., Zhang, L., Yan, D., You, C., 2001. Optimization of sample number for Weibull function of brittle materials strength. Ceram. Int. 27, 239–241. https://doi.org/10.1016/S0272-8842(00)00056-0.
Yin, H., He, X., Song, G., Zheng, Y., Bai, Y., 2023. Statistical analysis of the strength reliability of the MAX phases. J. Am. Ceram. Soc. 106, 1321–1332. https://doi.org/10.1111/jace.18829.
Yu, C.H., 2019. Resampling methods: concepts, applications, and justification. https://doi.org/10.7275/9CMS-MY97.
Zok, F.W., 2017. On weakest link theory and Weibull statistics. J. Am. Ceram. Soc. 100, 1265–1268. https://doi.org/10.1111/jace.14665.