Aus der Poliklinik für Zahnärztliche Prothetik
der Ludwig-Maximilians-Universität München

Direktor: Prof. Dr. med. dent. Daniel Edelhoff

Monolithische Zirkonoxid-Keramiken: Prüfmethoden und Auswertungsmethoden der Festigkeitswerte

Dissertation zum Erwerb des Doktorgrades der Zahnmedizin
an der Medizinischen Fakultät der
Ludwig-Maximilians-Universität zu München

vorgelegt von
Christine Sabrina Schatz
aus Bayreuth
2017

Mit Genehmigung der Medizinischen Fakultät der
Universität München

Berichterstatter: PD Dr. Dipl. Ing. (FH) Bogna Stawarczyk, MSc
Mitberichterstatter: Prof. Dr. Christoph Benz
Prof. Dr. Andrea Wichelhaus

Betreuung durch den promovierten Mitarbeiter:
PD Dr. Dipl. Ing. (FH) Bogna Stawarczyk, MSc

Dekan:
Prof. Dr. med. dent. Reinhard Hickel

Tag der mündlichen Prüfung: 07.12.2017

Eidesstattliche Versicherung

Schatz, Christine Sabrina

Name, Vorname

Ich erkläre hiermit an Eides statt,

dass ich die vorliegende Dissertation mit dem Thema

Monolithische Zirkonoxid-Keramiken:
Prüfmethoden und Auswertungsmethoden der Festigkeitswerte

selbständig verfasst, mich außer der angegebenen keiner weiteren Hilfsmittel bedient und alle Erkenntnisse, die aus dem Schrifttum ganz oder annähernd übernommen sind, als solche kenntlich gemacht und nach ihrer Herkunft unter Bezeichnung der Fundstelle einzeln nachgewiesen habe.

Ich erkläre des Weiteren, dass die hier vorgelegte Dissertation nicht in gleicher oder in ähnlicher Form bei einer anderen Stelle zur Erlangung eines akademischen Grades eingereicht wurde.

München, 09.02.2017

Ort, Datum

Unterschrift Doktorandin/Doktorand

Meinen Eltern für ihre bedingungslose Unterstützung gewidmet.

Inhaltsverzeichnis

1. Einleitung
2. Eigene Arbeiten
   2.1 Originalarbeit: Schatz, C.; Strickstrock, M.; Roos, M.; Edelhoff, D.; Eichberger, M.; Zylla, I.-M.; Stawarczyk, B. Influence of Specimen Preparation and Test Methods on the Flexural Strength Results of Monolithic Zirconia Materials. Materials 2016, *9*, 180. (Impact Factor: 2.728)
   2.2 Originalarbeit: Roos, M.; Schatz, C.; Stawarczyk, B. Two Independent Prospectively Planned Blinded Weibull Statistical Analyses of Flexural Strength Data of Zirconia Materials. Materials 2016, 9, 512. (Impact Factor: 2.278)
3. Diskussion
   3.1 Einfluss der Prüfkörperherstellung und der Prüfmethode auf die Biegefestigkeit von monolithischen Zirkonoxidmaterialien
   3.2 Zwei unabhängige, prospektiv geplante und verblindete statistische Weibullanalysen der Biegefestigkeitsdaten von Zirkonoxidmaterialien.
4. Zusammenfassung und Ausblick
5. Literaturverzeichnis
6. Danksagung
7. Lebenslauf

1. Einleitung

Monolithische Y-TZP (mit Yttriumoxid tetragonal teilstabilisiertes Zirkonoxid) Keramiken bekommen innerhalb der zeitgemäßen Hochleistungskeramiken immer mehr Beachtung. Die werkstoffkundliche Weiterentwicklung dieser Zirkonoxidkeramiken zielte auf die Modifikationen der Mikrostruktur zur Verbesserung des ästhetischen Erscheinungsbildes unter Beibehaltung der guten Materialeigenschaften ab. Mit der Verwendung von monolithischen Keramiken wird zudem das Problem des Chippings der Verblendkeramik umgangen, da die potentielle Schwachstelle des Verbundes Zirkonoxid zu Verblendkeramik entfällt.

Es ist wichtig ein Material in Bezug auf seine werkstoffkundlichen Eigenschaften und auch in Bezug auf seine Zuverlässigkeit zu charakterisieren. Um die Biegefestigkeit von monolithischen Zirkonoxidmaterialien zu testen, hat sich in der Praxis der Biaxiale-, Drei-Punkt-, und der Vier-Punkt-Biegeversuch etabliert. Auch die Art der Prüfkörperherstellung kann durch Alteration und Manipulation der Prüfkörperoberfläche die Biegefestigkeit beeinflussen. Um Aussagen über die Zuverlässigkeit eines Werkstoffes machen zu können, wird die Weibullstatistik angewandt. Die zwei-parametrische Weibullverteilung besteht aus dem Parameter des Weibullmoduls m und der charakteristischen Festigkeit s. Der Weibullmodul ermöglicht eine Interpretation der Zuverlässigkeit und der Strukturhomogenität eines getesteten Materials. Die charakteristische Festigkeit entspricht hierbei der 63.2% - Versagenswahrscheinlichkeit eines Werkstoffes.

Der statistischen Analyse mit verschiedenen Möglichkeiten der Herangehensweise an einen Datensatz kommt eine wichtige Funktion zu, um die vorliegenden Ergebnisse interpretieren zu können. So können die verfügbaren statistischen Hilfsmittel und Techniken, und nicht zuletzt die persönliche Erfahrung eines Statistikers einen Einfluss auf die Analyse der Daten haben.

Das Ziel der vorliegenden Arbeiten war es sowohl Erkenntnisse über die Biegefestigkeit von monolithischen Zirkonoxidmaterialien in verschiedenen Biegeversuchsaufbauten zu gewinnen und zu überprüfen, welchen Einfluss die Prüfkörperherstellung auf die Biegefestigkeit hat. Ein Teil der Daten sollte in einem verblindeten Design von zwei unabhängigen Statistikern analysiert werden. Es galt festzustellen, ob die statistische Herangehensweise einen Einfluss auf die Interpretation der Daten hat und inwiefern die Schlussfolgerungen möglicherweise differieren.

2. Eigene Arbeit

Im Folgenden werden zwei Originalarbeiten in englischer Sprache präsentiert und diskutiert.

2.1. Originalarbeit:

Schatz, C.; Strickstrock, M.; Roos, M.; Edelhoff, D.; Eichberger, M.; Zylla, I.-M.; Stawarczyk, B. Influence of Specimen Preparation and Test Methods on the Flexural Strength Results of Monolithic Zirconia Materials. Materials 2016, *9*, 180.

Zusammenfassung

Ziel: Das Ziel dieser Arbeit war es, den Einfluss der Prüfkörperherstellung und der Prüfmethode auf die Biegefestigkeit von monolithischen Zirkonoxidmaterialen zu überprüfen.

Material und Methode: Es wurden drei vorgesinterte monolithische Zirkonoxidmaterialen Ceramill Zolid (C), Zenostar ZrTranslucent (Z), und DD Bio zx^2 (D) mit drei unterschiedlichen Methoden der Biegeprüfung getestet: Drei-Punkt-, Vier-Punkt-, und Biaxiale-Biegeprüfung. Zusätzlich wurden zwei unterschiedliche Methoden der Prüfkörperherstellung angewandt. Zum einen wurden die Prüfkörper manuell und im trockenen Zustand vor dem finalen Sintern mit SiC-Papier poliert. Bei der zweiten Methode wurden diese maschinell mit Polierscheiben und unter Wasserkühlung nach dem Sintern poliert. Jede Untergruppe umfasste 40 Prüfkörper. Insgesamt wurden 720 Prüfkörper hergestellt. Die Oberflächenbeschaffenheit wurde mit Hilfe der Rasterelektronenmikroskopie und der Profilometrie untersucht und der Anteil der monoklinen Phasenumwandlung wurde unter Verwendung der Röntgendiffraktographie evaluiert. Die Daten wurden mit einer dreifaktoriellen Varianzanalyse (ANOVA) unter Berücksichtigung der drei Faktoren: Zirkonoxidmaterial (ZM: C, Z, D), Prüfkörperherstellung (SP: vorher, nachher) und Biegefestigkeitsprüfung (TM: Biaxiale-, Drei-Punkt-, Vier-Punkt-Biegefestigkeitsprüfung) analysiert. Eine einfaktorielle ANOVA wurde für die Biegefestigkeitsprüfung und Zirkonoxidmaterial-Faktoren mit der Kombination von zwei anderen Faktoren angewandt. Es wurde mit einer zwei-parametrischen Weibullverteilungsannahme gearbeitet, um die Zuverlässigkeit des Materials unter den vorliegenden Testbedingungen zu überprüfen.

Ergebnisse: Im Allgemeinen präsentierte die Vier-Punkt-Biegeprüfung die niedrigsten Biegefestigkeitswerte der Prüfkörper. Die Biegefestigkeitswerte können in folgender Rangordnung gruppiert werden: Vier-Punkt-Biegeprüfung < Drei-Punkt-Biegeprüfung < Biaxiale-Biegeprüfung. Die nach dem Sintern polierten Prüfkörper zeigten eine signifikant höhere Biegefestigkeit als die vor dem Sintern polierten Prüfkörper. Die Weibullmoduli reichten von 5.1-16.5. Es konnte eine erhöhte Oberflächenrauhigkeit bei den vor dem Sintern

polierten Prüfkörpern festgestellt werden. Die Poliermethoden führten nicht zu der Bildung einer starken monoklinen Oberflächenschicht. Ein maßgeblicher Einfluss des Zirkonoxidmaterials auf die Biegefestigkeitswerte konnte nicht beobachtet werden. Die Art der Biegeprüfung und die Art der Prüfkörperherstellung beeinflussten die vorliegenden Werte.

materials

MDPI

Article

Influence of Specimen Preparation and Test Methods on the Flexural Strength Results of Monolithic Zirconia Materials

Christine Schatz [1], Monika Strickstrock [2], Malgorzata Roos [3], Daniel Edelhoff [1], Marlis Eichberger [1], Isabella-Maria Zylla [2] and Bogna Stawarczyk [1,*]

1 Department of Prosthodontics, Dental School, Ludwig-Maximilians-University Munich, Goethestrasse 70, Munich 80336, Germany; c.sa.schatz@googlemail.com (C.S.); daniel.edelhoff@med.uni-muenchen.de (D.E.); marlis.eichberger@med.uni-muenchen.de (M.E.)

2 Faculty of Engineering and Computer Science, Material Science and Analysis, University of Applied Sciences, Hochschule Osnabrück, Albrechtstrasse 30, Osnabrück 49076, Germany; monika.strickstrock@hs-osnabrueck.de (M.S.); i.m.zylla@hs-osnabrueck.de (I.-M.Z.)

3 Department of Biostatistics, Epidemiology Biostatistics and Prevention Institute, University of Zurich, Hirschengraben 84, Zurich 8001, Switzerland; malgorzata.roos@uzh.ch

* Correspondence: bogna.stawarczyk@med.uni-muenchen.de; Tel.: +49-89-4400-59573

Academic Editor: Jérôme Chevalier

Received: 21 December 2015; Accepted: 2 March 2016; Published: 9 March 2016

Abstract: The aim of this work was to evaluate the influence of specimen preparation and test method on the flexural strength results of monolithic zirconia. Different monolithic zirconia materials (Ceramill Zolid (Amann Girrbach, Koblach, Austria), Zenostar ZrTranslucent (Wieland Dental, Pforzheim, Germany), and DD Bio zx^2 (Dental Direkt, Spenge, Germany)) were tested with three different methods: 3-point, 4-point, and biaxial flexural strength. Additionally, different specimen preparation methods were applied: either dry polishing before sintering or wet polishing after sintering. Each subgroup included 40 specimens. The surface roughness was assessed using scanning electron microscopy (SEM) and a profilometer whereas monoclinic phase transformation was investigated with X-ray diffraction. The data were analyzed using a three-way Analysis of Variance (ANOVA) with respect to the three factors: zirconia, specimen preparation, and test method. One-way ANOVA was conducted for the test method and zirconia factors within the combination of two other factors. A 2-parameter Weibull distribution assumption was applied to analyze the reliability under different testing conditions. In general, values measured using the 4-point test method presented the lowest flexural strength values. The flexural strength findings can be grouped in the following order: 4-point < 3-point < biaxial. Specimens prepared after sintering showed significantly higher flexural strength values than prepared before sintering. The Weibull moduli ranged from 5.1 to 16.5. Specimens polished before sintering showed higher surface roughness values than specimens polished after sintering. In contrast, no strong impact of the polishing procedures on the monoclinic surface layer was observed. No impact of zirconia material on flexural strength was found. The test method and the preparation method significantly influenced the flexural strength values.

Keywords: monolithic zirconia; flexural strength; biaxial strength; 3-point flexural strength; 4-point flexural strength; specimen preparation

1. Introduction

Y-TZP (Yttria partially stabilized tetragonal zirconia) Zirconia has become of interest in dentistry, because of its high flexural strength [1] and its well-known transformation toughening ability [2].

Conventional zirconia (3Y-TZP: Yttrium-cation-doped tetragonal zirconia polycrystals, typically 2–3 mol% Y_2O_3) shows translucency on human dentin level and is therefore not suitable for monolithic tooth restorations from the esthetic point of view [3].

Modifications of the microstructure and composition were conducted in recent years to accomplish an adequate esthetic appearance for full anatomical zirconia restorations. For example, the Al_2O_3 percentage was reduced from 0.25 to 0.05 wt % and customized with a smaller grain size. In addition, the Al_2O_3 grains were positioned on the grain boundaries of the ZrO_2 grains. The survival time of all-ceramic systems is strongly related to the flexural strengths of the restoration, which is the result of the flexural strength of core and veneering material, and the bond strength between both materials [4,5]. Although excellent strengths were reported for the zirconia core material, many studies indicate fractures in the veneering ceramic [6,7]. To avoid this problem and to facilitate the fabrication process, the demand for zirconia restorations in a monolithic design without veneering has increased [8]. Different studies reported on a higher fracture resistance of monolithic zirconia crowns compared to conventional veneered versions [8,9].

Many test methods have been introduced to measure the flexural strength of brittle ceramic materials [10,11]. Among them biaxial, 3-point testing and 4-point testing have evolved to be the most common. In 3-point testing a non-uniform stress field under the loading piston is created, whereas in 4-point testing the stress field between the support rolls is uniform, which can lead to different flexural strength findings. Variations in specimen sizes and testing-methods lead to different flaw size populations and thus different strength measurements [11,12]. The problem with brittle ceramic materials is that inherent material inhomogeneity can induce flaws such as micro-cracks or grain pullouts throughout the volume or on the surface of a material. This can lead to catastrophic failure. The Weibull distributional assumption with its parameters Weibull modulus m and characteristic strength s tries to take the largest flaw population into account. Whereas s corresponds to the 63.2% failure probability, a high Weibull modulus m is associated with a higher reliability of the material.

Regarding surface preparation methods, several studies showed that the preparation of the specimen has an important influence on the obtained flexural strength data [13–18]. Therefore, the aim of this study was to investigate the influence of specimen preparation and the test method on the flexural strength and Weibull statistics of different monolithic zirconia materials. The underlying null hypothesis was that zirconia materials, the specimen preparation methods, as well as the flexural strength test methods do not influence the flexural strength values and the reliability of monolithic zirconia materials.

2. Material and Methods

Three different pre-sintered zirconia materials for monolithic restorations were tested in this study: Ceramill Zolid (C), Zenostar Zr (Z), and DD Bio zx^2 (D) (Table 1).

Table 1. Summary of materials used in the present study, their manufacturer with LOT number, chemical components, and grain size.

Abbreviations	Zirconia Materials	Manufacturers	LOT Number	Chemical Components (%)	Grain Size (μm^2) Mean ± SD [19]
C	Ceramill Zolid	Amann Girrbach, Koblach, Austria	1111813	$ZrO_2 + HfO_2 + Y_2O_3 > 99$; Y_2O_3: 4.5–5.6; $HfO_2 < 5$; $Al_2O_3 < 0.5$	0.088 ± 0.004 [a]
Z	Zenostar Zr Translucent	Wieland+Dental, Pforzheim, Germany	20120306-27	$ZrO_2 + HfO_2 + Y_2O_3 > 99$; $4{,}5 < Y_2O_3 \leqslant 6$; $HfO_2 \leqslant 5$; Al_2O_3 + other oxides $\leqslant 1$	0.092 ± 0.003 [a]
D	DD Bio zx^2	Dental Direkt, Spenge, Germany	30712803	$ZrO_2 + HfO_2 + Y_2O_3 > 99$; $Al_2O_3 < 0.5$; other oxides $\leqslant 1$	0.124 ± 0.006 [b]

[a,b] Different letters present significant differences between materials.

Two different polishing methods were applied. Manually dry polishing before sintering and machine wet polishing after sintering. The zirconia groups (C, Z, D/dry and wet polished) were tested for biaxial, 3-point or 4-point flexural strength. Each group contained 40 specimens. In summary, 720 zirconia specimens were fabricated (Table 2).

Table 2. Division of specimens with abbreviation used. TM: test method, SP: specimen preparation, ZM: zirconia material.

Total	ZM	TM	SP	Subgroup
$N = 720$	C, D, Z $n = 240$	biaxial $n = 80$	dry-polished before sintering	$n = 40$
			wet-polished after sintering	$n = 40$
		3-point $n = 80$	dry-polished before sintering	$n = 40$
			wet-polished after sintering	$n = 40$
		4-point $n = 80$	dry-polished before sintering	$n = 40$
			wet-polished after sintering	$n = 40$

2.1. Specimen Preparation

All specimens were prepared in partially sintered state and in an enlarged size to compensate sintering shrinkage. The specimen for biaxial flexural strength measurement required a disc shape with a diameter of 16 mm. A CAM (computer aided manufacturing) machine (I-Mes 4030, Wieland Dental + Technik, Pforzheim, Germany) was used to mill cylinders out of the zirconia blanks (Software: Zenotec CAM 3.2 advanced V2.2017, Wieland Dental + Technik). A low speed (approximate 2.5 m/s) diamond saw (Well Diamantdrahtsägen, Mannheim, Germany; thread: type A 3–3) was used to cut the cylinders into slices. The saw uses a fine 0.3 mm diamond thread with embedded diamonds (diameter approximate 60 µm), the applied pressure was approximately 50 g. The specimens, for 3-point and 4-point flexural strength measurement were cut directly from the zirconia blanks with the diamond saw.

Manual polishing before sintering was done with SiC discs (Struers, Ballerup, Denmark). The sequence was SiC P400, SiC P500, and SiC P1000. The grinding time was 5 s per specimen-side and SiC disc. The load was applied by finger pressure and manual polishing was performed in small circles.

Machine polishing after sintering was performed with a water-cooled polishing machine (Struers Abramin, Struers, Ballerup, Denmark). Up to six specimens were polished during a polishing cycle at the same time. The polishing protocol consisted of coarse grinding with diamond pads of 40 µm and 20 µm for 6 min per each side (Pads: Code Granu 40 µm, Code Graku 20 µm, Struers; speed: 150 rpm; applied pressure: 20 N) and fine polishing with subsequent polishing solutions and a polishing plate for 6 min per side (Plate: MD-Largo (Struers, Ballerup, Denmark); Solutions (which are purely water-based diamond suspensions): Dia Pro Allegro/Largo (9 µm) and Dia Pro Largo (3 µm), Struers, Ballerup, Denmark). High polishing was conducted for 30 s per side, again with a polishing plate and polishing solution (Plate: MD-Chem (Struers, Ballerup, Denmark); Solution (colloidal silica suspension for final polishing): OP-S, Struers, Ballerup, Denmark; speed 150 rpm; applied pressure 200 N).

The zirconia materials were sintered (density from 3 g/cm^3 presintered to 6 g/cm^3 fully sintered) in a universal sintering oven (Nabertherm, Lilienthal/Bremen, Germany), according to each manufacturer's instructions (Table 3). The final dimensions DIN EN ISO 6872:2008 [20] were: for biaxial flexural strength measurement 16 mm × 1.2 mm (±0.05 mm), for 3-point flexural strength measurement 1.2 mm × 4 mm × 25 mm (±0.05 mm) and for 4-point flexural strength measurement 3 mm × 4 mm × 45 mm (±0.05 mm).

Table 3. Sintering parameters of all tested zirconia materials used in this study.

ZM	Heat Rate (°C/h)	Holding Temperature and Time (°C, h)	Final Temperature (°C)	Holding Time (h)	Cooling Rate (°C/h)
C	480	-	1450	2	300
Z	600	900; 0.5 h; further with 200 °C/h	1450	2	600
D	480	900; 0.5 h; further with 200 °C/h	1450	2	600

2.2. Flexural Strength Testing

Biaxial, 3-point, and 4-point flexural strength testing (Figure 1) are based upon DIN EN ISO 6872:2008 [20]. An appropriate sample holder was used for each specimen group to place it in a Universal Testing Machine (Zwick, Ulm, Germany) at a crosshead speed of 1 mm/min until failure. The specimens were tested dry at room temperature and dimensions were measured with a digital micrometer (Mitutoyo Deutschland, Neuss, Germany) to a precision of 0.01 mm.

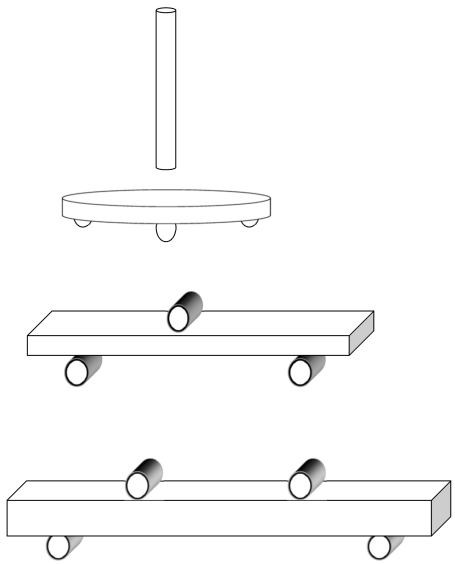

Figure 1. Schematic set up of biaxial; 3-point; 4-point test method.

The sample holder for the biaxial flexural strength test comprised three tempered steel balls with a diameter of 3.2 mm. The steel balls formed an equilateral triangle with an edge length of 10 mm and the ball support circle was 120°. The center of the specimens, which were put upon the steel balls and the center of the equilateral triangle were aligned coaxially. After the positioning the specimen's center was loaded from above with a plunger with a diameter of 1.4 mm until failure. The flexural strength was calculated according to the formula $\sigma = -0.25\ N(X-Y)/d^2$ (σ: flexural strength; N: fracture load; coefficients X and Y with:

$$X = (1 + \upsilon)\ln[(r2/r3)]^2 + [(1-\upsilon)/2](r2/r3)^2$$

$$Y = (1 + \upsilon)[1 + \ln(r1/r3)^2] + (1-\upsilon)(r1/r3)^2$$

where υ: Poisson's ratio (=0.25); r1: support (mean) contact diameter (mm); r2: (mean) loaded contact diameter (mm); r3: diameter of the specimen (mm); and d: thickness of the specimen (mm).

For the 3-point flexural strength measurement the specimen was placed on two tempered steel support rolls (diameter 1.6 mm). The distance between the two support rolls was 15 mm. Now a plunger (diameter 1.6 mm) loaded the specimen until failure. The formula $\sigma = 3Nl/(2bd^2)$ (σ: flexural strength; N: fracture load; l: distance between supports (mm); b: width of the specimen (mm); d: thickness of the specimen (mm)) was used to calculate 3-point flexural strength measurement.

The specimen for the 4-point flexural strength testing was placed on two tempered steel support rolls (diameter 4.0 mm) with a distance of 40 mm between the exterior supports. The plunger apparatus, consisting of two tempered steel rolls with a distance of 20 mm applied force until the specimens' failure. The calculation formula was $\sigma = 3Nl/(4bd^2)$ (σ: flexural strength; N: fracture load; l: mid to mid distance between exterior supports (mm); b: width of the specimen (mm); d: thickness of the specimen (mm)).

2.3. Profilometry

A profilometer (MarSurf 400 SD26, Mahr, Göttingen, Germany) was used to detect surface roughness (R_a) after preparation procedure. Six measurements for two specimens for each zirconia, test and polishing method were recorded (λ_c = 0.25 mm; L_t = 1.750 mm; pressure 0.7 mN) and a mean value was computed.

2.4. Scanning Electron Microscopy (SEM)

Scanning electron microscopy (EVO MA 10, Zeiss, Oberkochen, Germany) was used to assess the influence of specimen preparation on surface topography (15 kV, 150 mA, working distance of 10–12 mm). The specimens (n = 3 per group) were ultrasonically cleaned (Sonorex RK 100H; Bandelin, Berlin, Germany) and gold coated (Sputter Coater SC7640, Quorumtech, Newhaven, UK) before SEM examination.

2.5. X-ray Diffraction (XRD)

X-ray diffraction analysis (D8 Advance, Bruker AXS, Karlsruhe, Germany; settings: Bragg-Brentano Geometrie, Lynxeye-Dector in 1D-mode, nickel-kβ-filter) was employed to identify surface phase transformations. For each zirconia material, polishing and test method three specimens underwent XRD. Diffraction data were collected with Cu-Kα rays from the 2θ range between 25° and 40° with 40 kV, 40 mA and a step size of 0.02° and a scan time of 1.0 s/step, specimen rotation was 60 rpm. Under this condition of measurement the analyzed layer was 4.0–6.5 µm. The relative amount of monoclinic phase was calculated by means of the Garvie-Nicholson Equation modified through Toraya [21] (Equation (1)) using the maximum intensities of the reflexes $I_t(101)$, $I_m(\bar{1}11)$, und $I_m(111)$.

$$X_m = \frac{I_m(\bar{1}11) + I_m(111)}{I_m(\bar{1}11) + I_m(111) + I_t(101)} \quad (1)$$

The volume fraction (V_m) is:

$$V_m = \frac{1.311 \times X_m}{1 + 0.311 \times X_m} \quad (2)$$

2.6. Statistical Analysis and Methods

The power analysis was calculated in R (R, R Development Core Team, The R Foundation for Statistical Computing) using the data from our previous studies [22,23]. A sample size of 40 in each group will have 99% power to detect the difference of 200 MPa between two different flexural strength methods assuming that the common standard deviation is 178 MPa using the Students two sample

t-test with the 0.017 two-sided significance level obtained after the Bonferroni correction (three tests between three different flexural strength methods within one zirconia material).

Anderson-Darling goodness-of-fit estimates were computed and the best fitting distributional assumption (normal or Weibull) was indicated by its smaller value. In addition, the assumption of normality of the data was investigated by means of the Kolmogorov-Smirnov test. Descriptive statistics mean, standard deviation (SD), and coefficient of variation (COV = SD/mean) together with the 95% confidence intervals (95% CI) for the mean were computed. A three-way ANOVA for flexural strength with respect to three factors: test method (TM: biaxial, 3-point, 4-point), specimen preparation (SP: before, after sintering), and zirconia material (ZM: C, Z, D) was computed using SPSS Version 22.0 (IBM Deutschland GmbH, Ehningen, Germany). These results were amended by one-way ANOVAs for test method and zirconia material factors within the combination of two other factors, separately. The two-sample t-test with Welch adjustment for differing variances was used to identify the influence of treatment within the combination of two other factors. In addition, the two-parameter Weibull distributional assumption was used to compute Weibull parameters (modulus = m and characteristic strength = s) with Least Squares and median rank plotting positions assumption [24,25]. Equal m, s, SD, and mean Bartlett's modified likelihood ratio tests together with the appropriate Bonferroni post-hoc confidence interval were conducted with Minitab Version 14 (Minitab Ltd., Coventry, UK). Associations between two continuous variables were characterized by the non-parametric Spearman's rho correlation. Results of statistical analysis with p-value smaller than 0.05, were considered to be statistically significant.

3. Results

Table 4 depicts the Anderson-Darling goodness-of-fit estimates.

Table 4. Anderson-Darling goodness of fit estimates.

TM	SP	ZM	Weibull	Normal	Optimal Fit	Optimal Fit Distribution
Biaxial	before	C	0.637	0.702	0.611	3-par Weibull
		Z	0.793	0.821	0.726	Logistic
		D	1.102	1.735	0.947	SEV
	after	C	2.405	0.799	0.482	3-par Log normal
		Z	2.070	0.713	0.593	3-par Log normal
		D	1.111	1.589	0.940	SEV
3-point	before	C	0.916	0.455	0.434	3-par Log normal
		Z	0.613	0.594	0.536	3-par Weibull
		D	0.577	0.508	0.480	3-par Weibull
	after	C	1.101	0.568	0.512	Logistic
		Z	2.053	0.942	0.669	3-par Log logistic
		D	0.946	1.412	0.835	3-par Weibull
4-point	before	C	1.024	0.488	0.472	3-par Log normal
		Z	0.946	0.746	0.743	3-par Log normal
		D	0.755	0.496	0.496	normal
	after	C	0.693	1.033	0.691	SEV
		Z	0.685	0.708	0.708	normal
		D	0.625	0.698	0.625	Weibull

The Kolmogorov-Smirnov test showed that not all of the data were normally distributed (Table 5).

Table 5. Descriptive statistics for flexural strength values of all measured groups. TM: test method; SP: specimen preparation; ZM: zirconia material; SD: standard deviation; 95% CI: 95% confidence interval; Min: minimum; Max: maximum; COV: coefficient of variation (%). If not otherwise indicated all values are presented in MPa.

TM	SP	ZM	SD	95% CI (SD)	Mean	95% CI (Mean)	COV %
Biaxial	before	C $^{a/A}$	81 ^{a}	(64;102) ^{a}	757 ^{A}	(731;782) ^{B}	11
		Z $^{a/B}$	115 ^{a}	(91;144) ^{a}	891 ^{A}	(855;927) ^{C}	13
		D $^{a/B}$	101 ^{a}	(80;126) ^{a}	835 ^{*,A}	(803;866) ^{C}	12
	after	C $^{a/A}$	115 ^{a}	(91;145) ^{a}	1077 ^{B}	(1040;1112) ^{B}	11
		Z $^{ab/A}$	116 ^{a}	(92;146) ^{a}	1126 ^{B}	(1090;1162) ^{B}	10
		D $^{b/B}$	212 ^{b}	(170;265) ^{a}	1322 ^{*,B}	(1256;1388) ^{B}	16
3-point	before	C $^{a/A}$	110 ^{a}	(87;139) ^{a}	752 ^{A}	(718;787) ^{B}	15
		Z $^{a/A}$	124 ^{a}	(98;156) ^{a}	755 ^{A}	(716;793) ^{B}	16
		D $^{a/A}$	97 ^{a}	(76;122) ^{a}	743 ^{A}	(712;773) ^{B}	13
	after	C $^{a/AB}$	86 ^{a}	(68;108) ^{a}	1118 ^{B}	(1090;1144) ^{B}	8
		Z $^{ab/A}$	156 ^{a}	(123;196) ^{a}	1039 ^{B}	(990;1087) ^{AB}	15
		D $^{b/B}$	173 ^{a}	(138;216) ^{a}	1183 ^{*,B}	(1129;1237) ^{B}	15
4-point	before	C $^{a/A}$	78 ^{a}	(61;99) ^{a}	561 ^{*,A}	(536;585) ^{A}	14
		Z $^{a/B}$	88 ^{a}	(69;110) ^{a}	646 ^{A}	(618;673) ^{A}	14
		D $^{a/AB}$	112 ^{a}	(88;141)^{a}	637 ^{A}	(602;672) ^{A}	18
	after	C $^{a/A}$	122 ^{a}	(97;153)^{a}	873 ^{B}	(834;911) ^{A}	14
		Z $^{a/A}$	206 ^{b}	(163;260) ^{a}	922 ^{B}	(858;987) ^{A}	22
		D $^{a/A}$	182 ^{a}	(144;228) ^{a}	947 ^{B}	(890;1004) ^{A}	19

* not normally distributed; ^{abc} significant differences for SD values; column ZM indicates significant difference between ZM within TM and SP; column SD indicates significant difference between SP within TM and ZM; column 95% CI (SD) indicates significant difference between TM within SP and ZM; ^{ABC} significant differences for mean values; column ZM indicates significant difference between ZM within TM and SP; column mean indicates significant difference between SP within TM and ZM; column 95% CI (mean) indicates significant difference between TM within SP and ZM.

Additionally, the 2-parameter Weibull analysis was used to describe the flexural strength (Table 6). The data are presented in Figure 2. The 3-way ANOVA observed the impact of specimen preparation method on the Weibull modulus: Zirconia C specimen tested in 3-point flexural strength had a significant higher Weibull modulus when polished after sintering ($p < 0.001$).

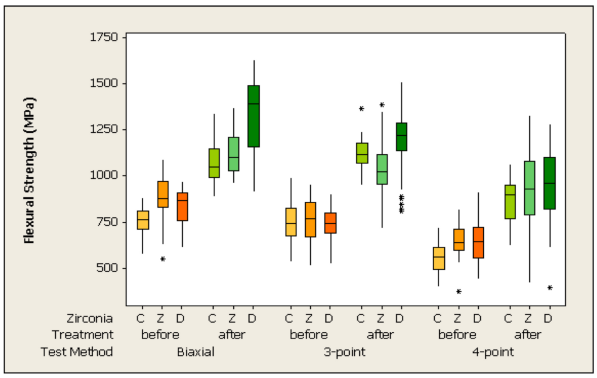

Figure 2. Influence of specimen preparation test method and zirconia material using different flexural strength testing methods.

Table 6. Weibull statistics for flexural strength values of all measured groups. TM: test method; SP: specimen preparation; ZM: zirconia material; m: Weibull modulus; 95% CI: 95% confidence interval; s: characteristical strength. All values are presented in MPa.

TM	SP	ZM	m	95%CI (m)	s	95%CI (s)
Biaxial	before	C $^{a/A}$	11.2 ^{a}	(8.7;14.4) ^{a}	791 ^{A}	(768;814) ^{A}
		Z $^{a/B}$	9.0 ^{a}	(6.9;11.6) ^{a}	940 ^{A}	(906;975) ^{C}
		D $^{a/B}$	9.6 ^{a}	(7.1;13.1) ^{a}	878 ^{A}	(848;908) ^{C}
	after	C $^{a/A}$	12.4 ^{a}	(10.5;14.6) ^{a,b}	1120 ^{B}	(1090;1152) ^{B}
		Z $^{a/A}$	12.8 ^{a}	(10.7;15.2) ^{b}	1170 ^{B}	(1139;1202) ^{B}
		D $^{a/B}$	7.3 ^{a}	(5.4;9.8) ^{a}	1408 ^{B}	(1346;1472) ^{B}
3-point	before	C $^{a/A}$	8.5 ^{a}	(7.0;10.4) ^{a}	795 ^{A}	(764;827) ^{A}
		Z $^{a/A}$	7.3 ^{a}	(5.7;9.5) ^{a}	804 ^{A}	(768;841) ^{B}
		D $^{a/A}$	9.3 ^{a}	(7.3;11.8) ^{a}	782 ^{A}	(755;811) ^{B}
	after	C $^{b/AB}$	16.5 ^{b}	(14.1;19.1) ^{b}	1153 ^{B}	(1129;1177) ^{B}
		Z $^{a/A}$	8.5 ^{a}	(7.1;10.2) ^{a,b}	1097 ^{B}	(1055;1142) ^{A,B}
		D $^{a/B}$	7.9 ^{a}	(6.0;10.4) ^{a}	1255 ^{B}	(1204:1308) ^{B}
4-point	before	C $^{a/A}$	9.1 ^{a}	(7.4;11.1) ^{a}	591 ^{A}	(570;614) ^{B}
		Z $^{a/B}$	8.8 ^{a}	(7.0;11.0) ^{a}	681 ^{A}	(656;708) ^{A}
		D $^{a/B}$	7.1 ^{a}	(5.8;8.6) ^{a}	679 ^{A}	(648;712) ^{A}
	after	C $^{a/A}$	8.4 ^{a}	(6.4;11.1) ^{a}	923 ^{B}	(888;959) ^{A}
		Z $^{a/A}$	5.1 ^{a}	(3.9;6.7) ^{a}	1002 ^{B}	(940;1069) ^{A}
		D $^{a/A}$	5.7 ^{a}	(4.2;7.7) ^{a}	1023 ^{B}	(965;1084) ^{A}

^{a,b,c} significant differences for m-values; column ZM indicates significant difference between ZM within TM and SP; column m indicates significant difference between SP within TM and ZM; column 95% CI(m) indicates significant difference between TM within SP and ZM; ^{A,B,C} significant differences for *s*-values; column ZM indicates significant difference between ZM within TM and SP; column s indicates significant difference between SP within TM and ZM; column 95% CI (s) indicates significant difference between TM within SP and ZM.

3.1. Influence of Test Method

The 4-point flexural strength test method indicated the lowest flexural strength values in all groups, regardless of which zirconia material was tested and which specimen preparation method was used ($p < 0.001$). An exception showed Z zirconia material polished after sintering. In this group 4-point and 3-point results were in the same range of values, but significantly lower than the biaxial values.

The biaxial strength test method provided the highest strength values. However, it was not significantly different from the 3-point test within the wet polished groups and within the dry polished zirconia C specimen ($p > 0.05$).

Regarding the Weibull modulus, 3-point testing showed a significantly higher Weibull modulus for the wet polished zirconia C group than for 4-point testing and biaxial testing ($p < 0.001$). Specimens tested in biaxial flexural strength presented a significantly higher Weibull modulus for wet polished Zirconia Z than those tested in 4-point strength test methods ($p < 0.001$). Other groups showed no influence of test method on the Weibull modulus ($p > 0.05$).

3.2. Influence of Specimen Preparation

The different polishing procedures greatly influenced the measured mean flexural strength independently of which zirconia was tested and which test method was applied. After sintering the wet polished specimen produced significantly higher flexural strength than specimens polished before sintering ($p < 0.001$). Only one group showed the impact of the specimen preparation method on the Weibull modulus. Zirconia C specimens tested in 3-point flexural strength ($p < 0.001$) provided a higher Weibull modulus when polished after sintering.

3.3. Influence of Tested Zirconia Materials

No impact of zirconia material on flexural strength values was found for 3-point tested dry polished specimen and 4-point tested wet polished specimen ($p = 0.671$).

The biaxial test method showed considerably lower values for dry polished zirconia C material than for material Z and D ($p < 0.001$). In contrast, wet polished Zirconia D presented noticeably higher flexural strength values compared to C and Z ($p = 0.034$).

Within specimens polished after sintering and tested in 3-point flexural strength, material Z showed significantly lower values than C and D ($p = 0.046$). Within specimens polished before sintering and tested in 4-point flexural strength, material C indicated significantly lower values than Z ($p < 0.001$). D was not different to Z. It is difficult to discern patterns in the ZM effect, as they seem to act differently for each TM and SP combination.

An impact of tested zirconia materials on Weibull modulus was observed only for specimens prepared after sintering and tested in 3-point test. In this group, a significantly higher Weibull modulus for C than for Z and D was observed ($p < 0.001$). All other groups were in the same Weibull modulus range.

3.4. Surface Roughness (Profilometer) and Surface Topography (SEM)

Dry polished specimens showed higher surface roughness compared to wet polished specimens. The surface roughness R_a values for the dry polishing method range from 0.31 to 0.41 µm and for the wet polishing method from 0.011 to 0.014 µm. An impact of zirconia materials (C, Z, D) was not observed. SEM pictures confirmed the results of the profilometric measurement (Figure 3). Roughness of the tested specimen was found to be associated with Weibull modulus (m) (Spearman's rho correlation = 0.199, $p = 0.428$).

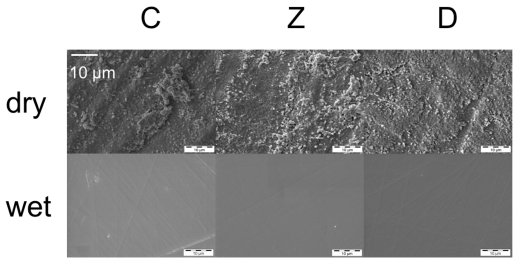

Figure 3. Scanning electron microscopy (SEM) pictures of dry and wet ground zirconia materials.

3.5. Characterization of Monoclinic Phase Transformation (X-ray Diffraction)

The volume fraction of the monoclinic phase at the surface was low (~2%) for all tested specimens (Figure 4), no material was noticeable different.

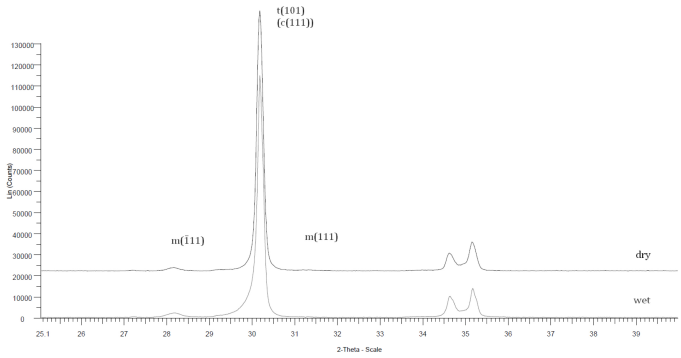

Figure 4. X-ray diffraction pattern example: Z, biaxial dry and wet.

4. Discussion

To the knowledge of the authors, no data for monolithic zirconia are existent comparing the different flexural strength test methods. For the discussion of our data we will also consider studies with other ceramic materials.

The tested null hypothesis was that the flexural strength test methods would show no impact on strength values. This has to be rejected. It could be clearly discriminated between biaxial and 4-point tested groups, which presented the lowest flexural strength findings. In 4-point flexural strength testing, a larger area of the material is involved in stress application. Thus, the probability of crack initiation, reflected by lower flexural strength, is higher than for the 3-point test method [10]. In our data, 4-point tested groups were in all but one group statistically different from 3-point tested groups. This is in accordance with the data of a previous study, where veneering ceramics for zirconia have been tested [26].

In another study with zirconia core materials, biaxial strength testing resulted in higher and statistically different values compared to those, measured by 3-point flexural strength testing [27]. Except for one group, our data showed the highest mean biaxial flexural strength values, but we could only observe a clear discrimination between 3-point and biaxial testing in two groups. The high values in the biaxial test can be related to the negligible effect of undesirable edge failures [28,29]. Jin *et al.* [30] used all of the three test methods on different ceramic types. Although no zirconia material was included in this study, the authors also identified the lowest values for 4-point flexural strength measurement and could not clearly differentiate between biaxial and 3-point testing. A ranking order of the different flexural strength test methods can be described as the following: biaxial > 3-point > 4-point flexural strength measurement. Regarding the influence of the test method on the Weibull modulus it was observed that only wet polished, 3-point tested zirconia C and wet polished, 4-point tested zirconia Z had a higher Weibull modulus compared to other groups. Neither surface roughness nor SEM pictures can explain these findings.

The second tested hypothesis was, that the specimen preparation methods do not influence the flexural strength and the reliability of monolithic zirconia. The inherent reliability of the material can be described with the Weibull modulus. A high Weibull modulus is associated with a statistically higher reliability. It was found that specimen preparation after sintering showed significantly higher flexural strength values whereas no general trend for the Weibull moduli was observed. The Weibull modulus for 3-point tested Zirconia C polished after sintering was even higher, than for those polished before sintering. Thus the first part of the second hypothesis could not be confirmed. Regarding the reliability, the other part of the hypothesis is accepted within flexural strength test-groups and zirconia material. In contrast to that, a statistically lower Weibull modulus was noticed in wet polished 4-point tested groups for Zirconia C and Z. This indicates that polishing may induce flaws, but the probability of critical flaw detection increases, when a larger volume of the material is involved in the testing method. As failure is initiated by the largest flaw or element, this effect is also referred to as the weakest link hypothesis [31]. The authors believe that surface scratches shown in the SEM pictures, functioned as crack origin and led to significantly lower strength values of the 4-point tested specimen. The values for Weibull moduli ranged between (5.1 and 16.5), which is in accordance with the values reported for Y-TZP core materials tested in other studies [25,32]. In the present literature, contradictory findings are reported for flexural strength of zirconia materials, dependent on different surface alteration methods after the sintering process. Several authors described a decrease in flexural strength and Weibull modulus of zirconia, when the surface underwent a grinding process [14,18,33,34]. On the other hand, an increase in mean flexural strength and a decreased reliability was observed with the use of fine-grained diamond burs [35] and with wet hand grinding [36] although in this study the Weibull distribution was not discussed. When corund-blasting as a surface method is conducted, an increase of strength is also reported associated with a reduction of the Weibull modulus [18]. The increase in flexural strength can be explained with the phase transformation in zirconia materials. The amount of the monoclinic phase was found to rise, when the surface was treated after sintering [18,37].

This well-known transformation toughening mechanism of zirconia materials can lead to compressive surface stresses and thus elevate the flexural strength [38,39]. In contrast to that, in the present study no higher fraction of monoclinic zirconia on the surface of wet polished specimen was found. Thus the polishing procedure must have removed the layer of monoclinic zirconia. This effect was also observed by Guazatto *et al.* [17]. The above discussion of possible origins of the flaws is supported by Anderson-Darling (AD) goodness of fit estimates in Table 4. Smaller AD values indicate a better fit of the data by the assumed distribution. The authors were unable to find a single distributional assumption of flexural strength values fitting optimally for all test configurations. However, both Weibull and normal assumptions provided reasonable and useful approximations to compare all test groups. This conclusion is supported by a strong negative association between Weibull moduli (m) and COV (rho = -0.972, $p < 0.001$). Weibull modulus estimates ranged from 5.1 to 16.5 whereas those of COV from 0.077 to 0.223 indicating that both Weibull and normal assumptions were applicable to all test groups.

The third hypothesis was that different zirconia materials do not influence flexural strength. This has to be accepted, as a general trend as the zirconia materials could not be identified. It was observed that the zirconia materials used in this study act differently dependent on flexural strength test and treatment methods. Literature shows [19] that D contains a significantly larger grain size compared to zirconia C and Z. However, no association between grain size and all tested parameters was found in our study.

5. Conclusions

Within the limitations of the laboratory investigation, the following conclusions can be drawn:

(1) The 4-point flexural strength testing shows the lowest flexural strength data; biaxial test method the highest.

(2) The specimen preparation method significantly impacts the flexural strength findings; surface roughness was higher with dry polished specimens.

(3) Flexural strength values of tested zirconia materials range within the same values and no clear effect of the zirconia material could be observed.

Acknowledgments: The authors would like to thank Amann Girrbach, Wieland+Dental and Dental Direkt for supporting this study with materials.

Author Contributions: Christine Schatz: experimental design, wrote the manuscript, Monika Strickstrock: performed the material analysis, Malgorzata Roos: performed the statistical analyses, proofread the manuscript, Daniel Edelhoff: proofread the manuscript, Marlis Eichberger: supported by specimen preparations, Isabella-Maria Zylla: proofread the manuscript, Bogna Stawarczyk: had the idea, co-wrote the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Christel, P.; Meunier, A.; Heller, M.; Torre, J.P.; Peille, C.N. Mechanical properties and short-term *in vivo* evaluation of yttrium-oxide-partially-stabilized zirconia. *J. Biomed. Mater. Res. Part A* **1989**, *23*, 45–61. [CrossRef] [PubMed]
2. Garvie, R.C.; Hannink, R.H.; Pascoe, R.T. Ceramic steel? *Nature* **1975**, *258*, 703–704. [CrossRef]
3. Pecho, O.E.; Ghinea, R.; Ionescu, A.M.; Cardona, J.D.L.C.; Paravina, R.D.; Pérez, M.D.M. Color and translucency of zirconia ceramics, human dentine and bovine dentine. *J. Dent.* **2012**, *40*, e34–e40. [CrossRef] [PubMed]
4. Aboushelib, M.N.; de Jager, N.; Kleverlaan, C.J.; Feilzer, A.J. Microtensile bond strength of different components of core veneered all-ceramic restorations. *Dent. Mater.* **2005**, *21*, 984–991. [CrossRef] [PubMed]
5. Conrad, H.J.; Seong, W.-J.; Pesun, I.J. Current ceramic materials and systems with clinical recommendations: A systematic review. *J. Prosthet. Dent.* **2007**, *98*, 389–404. [CrossRef]
6. Al-Amleh, B.; Lyons, K.; Swain, M. Clinical trials in zirconia: A systematic review. *J. Oral Rehabil.* **2010**, *37*, 641–652. [CrossRef] [PubMed]

7. Sax, C.; Hammerle, C.H.; Sailer, I. 10-year clinical outcomes of fixed dental prostheses with zirconia frameworks. *Int. J. Comput. Dent.* **2011**, *14*, 183–202. [PubMed]
8. Johansson, C.; Kmet, G.; Rivera, J.; Larsson, C.; Vult Von Steyern, P. Fracture strength of monolithic all-ceramic crowns made of high translucent yttrium oxide-stabilized zirconium dioxide compared to porcelain-veneered crowns and lithium disilicate crowns. *Acta Odontol. Scand.* **2013**, *72*, 145–153. [CrossRef] [PubMed]
9. Beuer, F.; Stimmelmayr, M.; Gueth, J.-F.; Edelhoff, D.; Naumann, M. *In vitro* performance of full-contour zirconia single crowns. *Dent. Mater.* **2012**, *28*, 449–456. [CrossRef] [PubMed]
10. Quinn, G.D.; Morrell, R. Design data for engineering ceramics: A review of the flexure test. *J. Am. Ceram. Soc.* **1991**, *74*, 2037–2066. [CrossRef]
11. Shetty, D.K.; Rosenfield, A.R.; Duckworth, W.H.; Held, P.R. A biaxial-flexure test for evaluating ceramic strengths. *J. Am. Ceram. Soc.* **1983**, *66*, 36–42. [CrossRef]
12. Quinn, J.B.; Quinn, G.D. A practical and systematic review of Weibull statistics for reporting strengths of dental materials. *Dent. Mater.* **2010**, *26*, 135–147. [CrossRef] [PubMed]
13. Bhamra, G.; Palin, W.M.; Fleming, G.J.P. The effect of surface roughness on the flexure strength of an alumina reinforced all-ceramic crown material. *J. Dent.* **2002**, *30*, 153–160. [CrossRef]
14. Curtis, A.R.; Wright, A.J.; Fleming, G.J.P. The influence of surface modification techniques on the performance of a Y-TZP dental ceramic. *J. Dent.* **2006**, *34*, 195–206. [CrossRef] [PubMed]
15. de Castro, H.L.; Corazza, P.H.; Paes-Junior, T.D.A.; Della Bona, A. Influence of Y-TZP ceramic treatment and different resin cements on bond strength to dentin. *Dent. Mater.* **2012**, *28*, 1191–1197. [CrossRef] [PubMed]
16. Della Bona, A.; Donassollo, T.A.; Demarco, F.F.; Barrett, A.A.; Mecholsky, J.J., Jr. Characterization and surface treatment effects on topography of a glass-infiltrated alumina/zirconia-reinforced ceramic. *Dent. Mater.* **2007**, *23*, 769–775. [CrossRef] [PubMed]
17. Guazzato, M.; Albakry, M.; Quach, L.; Swain, M.V. Influence of surface and heat treatments on the flexural strength of a glass-infiltrated alumina/zirconia-reinforced dental ceramic. *Dent. Mater.* **2005**, *21*, 454–463. [CrossRef] [PubMed]
18. Kosmač, T.; Oblak, C.; Jevnikar, P.; Funduk, N.; Marion, L. The effect of surface grinding and sandblasting on flexural strength and reliability of Y-TZP zirconia ceramic. *Dent. Mater.* **1999**, *15*, 426–433. [CrossRef]
19. Stawarczyk, B.; Frevert, K.; Ender, A.; Roos, M.; Sener, B.; Wimmer, T. Comparison of four monolithic zirconia materials with conventional ones: Contrast ratio, grain size, four-point flexural strength and two-body wear. *J. Mech. Behav. Biomed. Mater.* **2016**, *59*, 128–138. [CrossRef] [PubMed]
20. *Dentistry—Ceramic Materials*; ISO 6872; International Organizatino for Standardization: Geneva, Switzerland, 2008.
21. Toraya, H.; Yoshimura, M.; Somiya, S. Calibration curve for quantitative analysis of the monoclinic-tetragonal ZrO2 system by X-ray diffraction. *J. Am. Ceram. Soc.* **1984**, *67*. [CrossRef]
22. Roos, M.; Stawarczyk, B. Evaluation of bond strength of resin cements using different general-purpose statistical software packages for two-parameter Weibull statistics. *Dent. Mater.* **2012**, *28*, e76–e88. [CrossRef] [PubMed]
23. Stawarczyk, B.; Özcan, M.; Hämmerle, C.H.F.; Roos, M. The fracture load and failure types of veneered anterior zirconia crowns: An analysis of normal and Weibull distribution of complete and censored data. *Dent. Mater.* **2012**, *28*, 478–487. [CrossRef] [PubMed]
24. Bütikofer, L.; Stawarczyk, B.; Roos, M. Two regression methods for estimation of a two-parameter Weibull distribution for reliability of dental materials. *Dent. Mater.* **2015**, *31*, e33–e50. [CrossRef] [PubMed]
25. Stawarczyk, B.; Özcan, M.; Trottmann, A.; Hämmerle, C.H.F.; Roos, M. Evaluation of flexural strength of hipped and presintered zirconia using different estimation methods of Weibull statistics. *J. Mech. Behav. Biomed. Mater.* **2012**, *10*, 227–234. [CrossRef] [PubMed]
26. Fischer, J.; Stawarczyk, B.; Hämmerle, C.H.F. Flexural strength of veneering ceramics for zirconia. *J. Dent.* **2008**, *36*, 316–321. [CrossRef] [PubMed]
27. Chai, J.; Chu, F.C.; Chow, T.W.; Liang, B.M. Chemical solubility and flexural strength of zirconia-based ceramics. *Int. J. Prosthodont.* **2007**, *20*, 587–595. [PubMed]
28. Anusavice, K.J.; Kakar, K.; Ferree, N. Which mechanical and physical testing methods are relevant for predicting the clinical performance of ceramic-based dental prostheses? *Clin. Oral Implants Res.* **2007**, *18*, 218–231. [CrossRef] [PubMed]
29. Ban, S.; Anusavice, K.J. Influence of test method on failure stress of brittle dental materials. *J. Dent. Res.* **1990**, *69*, 1791–1799. [CrossRef] [PubMed]

30. Jin, J.; Takahashi, H.; Iwasaki, N. Effect of test method on flexural strength of recent dental ceramics. *Dent. Mater. J.* **2004**, *23*, 490–496. [CrossRef] [PubMed]
31. Danzer, R.; Lube, T.; Supancic, P.; Damani, R. Fracture of ceramics. *Adv. Eng. Mater.* **2008**, *10*, 275–298. [CrossRef]
32. Papanagiotou, H.P.; Morgano, S.M.; Giordano, R.A.; Pober, R. *In vitro* evaluation of low-temperature aging effects and finishing procedures on the flexural strength and structural stability of Y-TZP dental ceramics. *J. Prosthet. Dent.* **2006**, *96*, 154–164. [CrossRef] [PubMed]
33. Kosmač, T.; Oblak, Č.; Jevnikar, P.; Funduk, N.; Marion, L. Strength and reliability of surface treated Y-TZP dental ceramics. *J. Biomed. Mater. Res.* **2000**, *53*, 304–313. [CrossRef]
34. Luthardt, R.G.; Holzhüter, M.; Sandkuhl, O.; Herold, V.; Schnapp, J.D.; Kuhlisch, E.; Walter, M. Reliability and properties of ground Y-TZP-zirconia ceramics. *J. Dent. Res.* **2002**, *81*, 487–491. [CrossRef] [PubMed]
35. Denry, I.L.; Holloway, J.A. Microstructural and crystallographic surface changes after grinding zirconia-based dental ceramics. *J. Biomed. Mater. Res. Part B Appl. Biomater.* **2006**, *76*, 440–448. [CrossRef] [PubMed]
36. Qeblawi, D.M.; Muñoz, C.A.; Brewer, J.D.; Monaco, E.A., Jr. The effect of zirconia surface treatment on flexural strength and shear bond strength to a resin cement. *J. Prosthet. Dent.* **2010**, *103*, 210–220. [CrossRef]
37. Karakoca, S.; Yilmaz, H. Influence of surface treatments on surface roughness, phase transformation, and biaxial flexural strength of Y-TZP ceramics. *J. Biomed. Mater. Res. Part B Appl. Biomater.* **2009**, *91*, 930–937. [CrossRef] [PubMed]
38. Gupta, T.K. Strengthening by surface damage in metastable tetragonal zirconia. *J. Am. Ceram. Soc.* **1980**, *63*, 117–117. [CrossRef]
39. Hannink, R.H.J.; Kelly, P.M.; Muddle, B.C. Transformation toughening in zirconia-containing ceramics. *J. Am. Ceram. Soc.* **2000**, *83*, 461–487. [CrossRef]

© 2016 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons by Attribution (CC-BY) license (http://creativecommons.org/licenses/by/4.0/).

2.2 Originalarbeit:

Roos, M.; Schatz, C.; Stawarczyk, B. Two Independent Prospectively Planned Blinded Weibull Statistical Analyses of Flexural Strength Data of Zirconia Materials. Materials 2016, 9, 512.

Zusammenfassung

Ziel: Es wurde ein verblindeter Datensatz zwei unabhängigen Statistikern präsentiert. Das Ziel war es, in zwei parallel geführten statistischen Analysen den optimalen statistischen Ansatz der Analyse für in-vitro gemessene Biegefestigkeitsdaten von Zirkonoxidmaterialen zu finden.

Material und Methode:

Der Datensatz bestand aus 240 Biegefestigkeitswerten, die mittels der Biaxialen-Biegefestigkeitsprüfung ermittelt wurden. Sie wurden in sechs Gruppen G1-G6 aufgeteilt, wobei jede Gruppe durch den Faktor des Zirkonoxidmaterials (ZM mit drei Ebenen C = Ceramill Zolid, Z= Zenostar ZrTranslucent und D = DD Bio zx^2) und durch die Methode der Prüfkörperherstellung (SP mit zwei Ebenen vorher = Trockenpolitur der Prüfkörper vor dem Sintern, nachher = Nasspolitur der Prüfkörper nach dem Sintern) definiert wurde. Jede Gruppe bestand aus 40 Prüfkörpern. Der verblindete Datensatz wurde den Statistikern zur Analyse übergeben. Sie wurden aufgefordert zwei unabhängige Entwürfe der Manuskriptabschnitte „Ergebnisse“, „Zusammenfassung“, „Schlussfolgerung“, sowie „Begründung der Wahl der statistischen Methoden“ vorzulegen. Nach Vergleich und Diskussion der vorgelegten Manuskripte wurden die Daten entschlüsselt. In der anschließenden Zusammenfassung der beiden unabhängigen Entwürfe sollten sowohl identische Ergebnisse sowie Gegensätze und Vervollständigungen beschrieben werden.

Ergebnisse: Beide Statistiker beschritten unterschiedliche Wege zur Analyse der Daten. Sie kamen zu komplementären Ergebnissen. Der Hauptunterschied bestand aus der unterschiedlichen Verteilungsannahme (Weibull/Normal) und alternativen Fitting-Methoden (LS/ML).

materials

MDPI

Article

Two Independent Prospectively Planned Blinded Weibull Statistical Analyses of Flexural Strength Data of Zirconia Materials

Malgorzata Roos [1,*], Christine Schatz [2] and Bogna Stawarczyk [2]

[1] Department of Biostatistics, Epidemiology Biostatistics and Prevention Institute, University of Zurich, Hirschengraben 84, 8001 Zurich, Switzerland

[2] Department of Prosthodontics, Dental School, Ludwig-Maximilians-University Munich, Goethestrasse 70, 80336 Munich, Germany; c.sa.schatz@googlemail.com (C.S.); bogna.stawarczyk@med.uni-muenchen.de (B.S.)

* Correspondence: malgorzata.roos@uzh.ch; Tel.: +41-44-63-44-648

Academic Editors: Nicola Pugno and Dirk Lehmhus
Received: 13 April 2016; Accepted: 20 June 2016; Published: 24 June 2016

Abstract: Zirconia as a restoration dental material are gaining attention because of their high mechanical properties and good biocompatibility. Therefore, investigation of the flexural strength of zirconia is of great interest. For this purpose, Weibull statistics for description of the material reliability are frequently used. The aim of this work was to present a blinded data set to two independent statisticians for two parallel statistical analyses in order to find an optimal statistical approach for analysis of in-vitro measured flexural strength data of zirconia materials. A prospectively planned independent blinded statistical analysis implementing three quality control actions "blinded data set", "independent statistical analyses" and "parallel manuscript writing" was designed. Statistical analysis paths taken by both biostatisticians differed. They arrived at complementary results. The major difference was caused by two alternative distributional assumptions (Weibull/Normal) and alternative fitting methods (LS/ML). The parallel statistical analysis and manuscript writing approach on a blinded data set greatly supported our choice of statistical methods for analysis of flexural strength results of zirconia materials.

Keywords: flexural strength data; Weibull analysis; blinded statistical analysis; permutation tests

1. Introduction

Monolithic zirconia was introduced to avoid the risk of dental restoration failure due to veneer-chipping fractures. To improve the esthetic appearance of zirconia restorations, attempts were made to improve translucency. One way was to increase the sintering temperature leading to microstructural changes in the material. As literature shows this method has a significant drawback. Grain growth and a non-homogenous structure influences the flexural-strength negatively [1]. Hence the Al_2O_3 percentage was reduced from 0.25 to 0.05 wt % and the smaller Al_2O_3 grains were positioned on the boundaries of the zirconia grains, which also results in a higher translucency of the material. At this moment, insufficient information is available about the flexural strength of second generation zirconia. The reliability of these flexural strength findings gives further insight to the performance and is an indicator of a zirconia material's quality. For this purpose, Weibull statistics are used. However, Weibull statistics can lead to minimal deviations depending on the estimation method [2]. The aim of this work was to present the blinded data to two statisticians in order to see how and if the further statistical investigation leads to a consensus view of the quality of the zirconia materials. The blinded data set provided for the parallel statistical analysis contained 240 measurements from a previous study [3] on biaxial flexural strength for three monolithic zirconia materials (ZM) Ceramill Zolid,

Zenostar ZrTranslucent and DD Bio zx2 for which two different specimen preparation (SP) methods: either dry polishing before sintering or wet polishing after sintering were applied.

1.1. Motivation for a Weibull Analysis

In dentistry, brittle materials are well characterized using Weibull statistics [4,5]. Ceramics, especially the high-performance ceramics such as zirconia or alumina are standardly tested for the reliability and homogeneity of the structure of the material by Weibull modulus (m).

For statistical analysis of flexural strength data [2,6–8] both the two-parameter Weibull (s, m) and the two-parameter Normal (mean, sd^2) distributional assumptions are frequently used. Weibull modulus m describes the reliability of the measurements with higher values corresponding to better reliability of the material. Normal mean corresponds to the characteristic strength (scale, s) of the Weibull distribution. Weibull modulus (m) is approximately inverse proportional to the Normal standard deviation (sd) ($m = 1/sd$). A reliable material with a high Weibull modulus (m) has low values of standard deviation (sd) for Normal distributional assumption as $sd = 1/m$. When analyzing data both differences of s (mean) and m (sd) estimates between the different groups for Weibull (Normal) distributional assumption are of interest.

Application of Normal and Weibull distributional assumptions implies a different perception of the underlying truth. First, the Normal distribution emerges when each observation is a sum of a possibly large number of independent random fluctuations [9]. In contrast, the stochastic process governing the Weibull distribution follows the "weakest link" concept [10–12]. Internal independent flaws propagate under external stress and eventually a specimen breaks at the weakest place. This way a single flaw dictates the strength of the whole system. Second, the support of the Normal distribution comprises the whole real line. Consequently, it is not perfectly suitable for modelling positive flexural strength observations. In this respect the Weibull distribution is more appropriate as it is defined on the positive real line. The third argument in favour of the Weibull distribution is its shape flexibility. Weibull distribution is capable of modelling symmetric as well as negatively and positively skewed data. In contrast, the Normal distribution is only acceptable for symmetric data.

Although the probabilistic basis for application of the Weibull distribution for brittle materials was quite strong its utilization in practice was hampered by a fairly cumbersome parameter estimation [4,10]. Recently, much progress has been made with respect to the availability of the Weibull distribution. There are flexible general-purpose statistical programs for estimation of the two-parameter Weibull distribution applying either the maximum likelihood (ML) or the least squares (LS) methodologies or both [6]. Many of them provide 95% confidence intervals for the Weibull parameters, convenient probability plots and conduct tests for parameter differences between factor levels. There is also a free available open source Excel-calculator facilitating an automatic LS estimation of Weibull parameters together with the corresponding 95% CI (Appendix C in [13]). Aside of this eminent progress there is still an uncertainty caused by a data analyst [14].

1.2. Motivation for a Prospectively Planned Independent/Parallel Blinded Statistical Analysis

A typical study passes through three work phases consisting of data generation, statistical analysis and manuscript writing (Table 1, study phases a–c). Each phase can be afflicted by several sources of uncertainty [14], which can be intentional or unconscious. Various problems in design, conduct, analysis and reporting of research may lead to biases toward findings that can distort the perception of the research progress [15].

Table 1. Sources of Uncertainty.

Study Phase	Uncertainty Source
Data generation (a)	Specimens/Subjects Investigators Data collectors and managers Precision of measuring devices Outcome assessors
Statistical data analysis (b)	Sample size Data analyst Descriptive statistics Assumption on the sampling distribution (model uncertainty) Outliers (data uncertainty) Choice of the statistical approach (frequentist or Bayesian) If Bayesian, then prior elicitation (prior uncertainty) Transformation of variables Parametric or non-parametric analysis Tests/Confidence intervals Choice of the estimation technique within the approach chosen Interpretation of the results Missing data handling Subgroup analysis Covariates selection
Writing of the manuscript (c)	Choice of the findings to report on Choice of the graphs to be shown Manuscript writer

Several authors [14,16,17] warn that an objectivity of statistical data analysis (Table 1, study phase b) is a clear misperception. The pretended objectivity of data analysts can be affected adversely [17] by many factors. For example, a statistician's bias is caused by differences in available statistical tools, techniques, programs and personal experience. In the course of statistical data analysis numerous semi-subjective decisions (Table 1, study phase b) have to be made that can have major effects on the results of the study [14,16]. Such semi-subjective decisions include assumptions to be made and models to be applicable [14], finally leading to the choice of a particular statistical method. Subjectivity is also involved in performing the analysis and interpreting the results. Perception of the data and the choice of the analysis can considerably vary across statisticians. A bias in direction of preferred methods, prior theoretical expectations or some other preconceptions is possible and very likely [14].

Usually, the size of the statistician's bias cannot be estimated. Therefore, it is beneficial to use methodology that suppresses, prevents or indicates it clearly. One possible resort is the use of the blinded design [18–20]. Blinding is a research strategy that involves the deliberate withholding of information from people who play a role in a study [16,21,22]. In principle, blinding of subjects, investigators, outcome assessors, data managers, biostatisticians and manuscript writers or any combination of them is possible [16,21–23]. The idea of blinding is a well-known scientific method frequently used in many fields of research and notably in the context of clinical trials [18,19,23,24]. The wide ranging applicability of the blinding technique guarantees its validity not only in in-vivo but also in in-vitro studies. For example, it is extensively used in modern statistical analysis of particle physics experiments [25], where measurements and conducted experiment are completely unknown to the involved analysts.

Whereas single- and double-blinded study designs concentrate mainly on reduction of uncertainties at the data generation step (Table 1, study phase a), by blinding the subjects only and both subjects and investigators, respectively, the statistical data analysis phase (Table 1, study phase b) has been perceived as an objective one, especially if the statistician is not involved in the experimental process [18]. This belief, however, clearly disagrees with observation that both statistical analysis and writing of the manuscript are potential sources of bias that persist even in double-blinded trials [17].

To counteract these problems a triple-blinded design has been suggested [23,24]. The triple-blinded design—a double-blind trial that also maintains a blind data analysis—aims for reduction of a potential bias introduced by the statistical data analysis [23,24] by keeping the involved analysts in dark about the meaning of the treatment groups specification during data analysis. In such a case the analyst is blinded to the meaning of the data. Ideally, data entry is done independently of the analyst and codes for treatment group assignments are allocated randomly [16]. The "blinded data set" approach aims for objective and straight decisions during the conduct of statistical analysis.

Unfortunately, an analysis of a blinded data set can be still insufficient for the bias reduction induced by the statistician. Therefore, Polit [16] and Miller & Stewart [23] suggest different strategies for broadening of the blinding technique within the statistical analysis phase (Table 1, study phase b). They claim that a blinded data analysis is most efficiently achieved by an "independent statistical analysis" when two independent data analysts are involved and both are blinded to the treatment group status. In a scenario when two statisticians are analyzing the same data set, free communication between them could promote convergence towards a consensus and important clues might be missed. The respective final analyses might be distorted and the results biased. Hence, in order to fully profit from a parallel data analysis, statisticians should analyze the data independently and provide their stand-alone description of the outcomes. Keeping this in mind Gøtzsche [17] suggests application of independent investigation not only within statistical data analysis phase (Table 1, study phase b) but also during the process of manuscript writing (Table 1, study phase c). In particular, "Results", "Summary", "Conclusion" and "Rationale for the choice of the statistical methods" sections should be drafted in two independent versions [17] leading to a "parallel manuscript writing" approach. Moreover, any additional analyses, performed after code breaking, should be identifiable as such. Hence, any independent blinded statistical analysis which joins both "independent statistical analyses" and "parallel manuscript writing" techniques and applies them to a "blinded data set" aims for diminishing not only the outcome reporting bias but also the risk that the conclusion and treatment recommendation are influenced by irrelevant factors [17].

Blinding of the data analyst has been identified as an easy way to minimize bias and to enhance credibility of the results [16]. The blinded data strategy initiated by Gøtzsche [17] has been applied in an increasing amount of studies. Depending on the field of research data analysts are blinded in as much as 2.5%–15% of research projects [16]. In each case, a blinded design seeks to achieve a higher standard of scientific rigor than a conventional non-blinded one.

Although over the years blinded analysis strategy has been steadily gaining in importance in numerous fields of research, its application in the context of dental materials research is still lacking. Therefore, we prospectively designed an independent blinded statistical analysis which could be conveniently applied for dental material projects. The aim of this study was to investigate by means of two independent prospectively planned blinded analyses if there is an agreement in Weibull statistics of flexural strength data of different zirconia ceramics analyzed independently by two statisticians using differing statistical approaches. We applied a random allocation of coding to tested groups in the data set and kept the analysts in dark about their meaning. What is more, any communication between statisticians during this time period was forbidden. Two independent drafts describing findings and statistical methods were written before code breaking.

The first hypothesis for the dental material research states that monolithic zirconia specimen preparation (SP) has no impact on the flexural strength. The second hypothesis states that all three tested zirconia materials (ZM) show similar flexural strength results. Additionally, two hypotheses with respect to two independent prospectively planned blinded Weibull statistical analyses were formulated. First, there is an agreement in statistical analysis paths chosen by both biostatisticians for Weibull analysis. Second, the three quality control actions: "blinded data set", "independent statistical analyses" and "parallel manuscript writing" conducted by two independent biostatisticians (implemented within our prospectively planned parallel blinded statistical analysis) have no influence on findings for flexural strength data.

2. Material and Methods

2.1. Experimental Data Description

Data for blinded analysis were a subset of a larger data collection in [3]. Data provided to the statisticians by the study supervisor A consisted of 240 biaxial flexural strength measurements divided randomly into six groups G1–G6. Each group was defined by a factor zirconia material (ZM with three levels C = Ceramill Zolid, Z = Zenostar ZrTranslucent and D = DD Bio zx2) and specimen preparation (SP with two levels before = dry polishing before sintering, after = wet polishing after sintering) and contained 40 specimens (Table 2). The specimen preparation process consisted of the following steps. The specimens were cut out of zirconia cylinders with a low speed diamond saw (Well, Diamantdrahtsägen, Mannheim, Germany). Manual dry polishing was conducted with SiC discs (Struers, Ballerup, Denmark) and machine wet polishing after sintering was executed with a water-cooled polishing machine (Struers Abramin, Struers, Ballerup, Denmark). For the sintering process a universal sintering oven (Nabertherm, Lilienthal/Bremen, Germany) was used. Only one operator handled the specimens. The final dimension DIN EN ISO 6872:2008 [26] of all disc shaped specimens was 16 mm × 1.2 mm (±0.05 mm). For biaxial flexural strength measurement, the specimen were put on to a sample holder, which consisted of three tempered steel balls (diameter 3.2 mm) forming an equilateral triangle (edge length 10 mm and ball support circle 120°). The plunger (diameter 1.4 mm) of the Universal Testing Machine (Zwick, Ulm, Germany) loaded the specimens with a crosshead speed of 1 mm/min until failure.

Table 2. Data summary for Normal assumption. n: number of observations; q1: first quartile; q2: third quartile.

Tested Groups	ZM	SP	n	Min	q1	Mean	Median	q2	Max	sd
G1	C	before	40	575	719	757	765	804	884	79
G2	C	after	40	890	997	1077	1050	1143	1340	113
G3	Z	before	40	551	842	891	878	966	1090	115
G4	Z	after	40	962	1030	1126	1100	1203	1370	114
G5	D	before	40	615	764	835	869	908	969	102
G6	D	after	40	915	1180	1322	1390	1490	1630	214

2.2. A prospectively Planned Independent Blinded Statistical Analysis

The independent blinded statistical analysis technique was devised prior to the data analysis. Figure 1 depicts the details of our plan, which consisted of five stages. For sake of clarity of presentation, the contributors in the sequel are identified by capital letters A, B, C and D. In the planning stage of our study the contributor A was designated to be the supervisor of the project and had an unlimited insight in all phases of the study.

In the first stage contributors A and B were responsible for conduction and execution of experiments and recording of the measurements in an Excel file. In the second stage A randomly assigned coding numbers to the tested groups, kept the key identifying the meaning of the factors in the data set and was not allowed to reveal it to both statisticians' C and D until the statistical analysis of the measurements was accomplished. In addition, contributor A checked graphically the plausibility of recorded measurements and correctness of factors coding.

Beginning of the third stage was marked by the release of the blinded data set to statisticians by contributor A. Both statisticians' C and D were not allowed to communicate and interact with each other during this stage. In case of any questions they might, however, address the supervisor A directly. They were given 3 weeks to accomplish statistical analysis of the data and to describe their findings and statistical methods applied with all tables, graphs and references necessary to back up their conclusions. During this time, they wrote two independent drafts of "Results", "Summary", "Conclusion" and "Rationale for the choice of the statistical methods" sections. Apart of that the time amount needed for accomplishing every step of the statistical analysis and draft writing was recorded.

The statisticians were blinded to the meaning of data coding and their knowledge about the goal of statistical analysis was limited to a short note provided by A accompanying the data set.

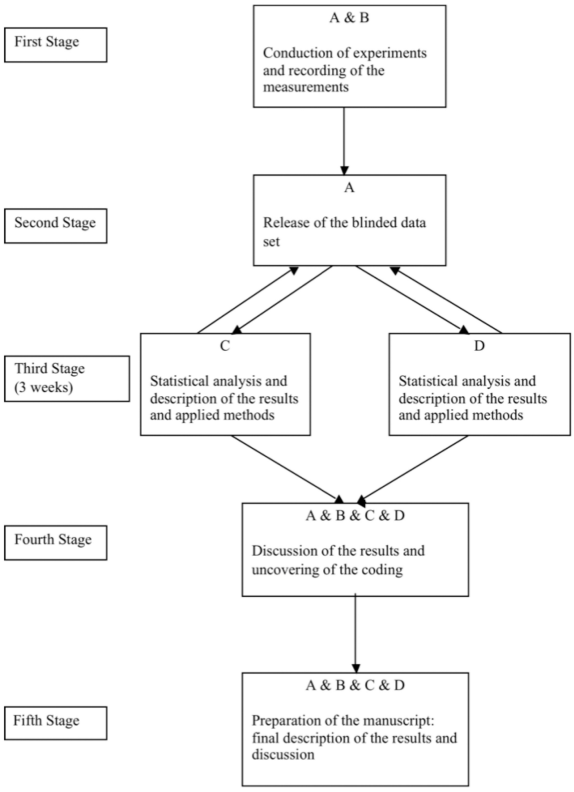

Figure 1. Plan of the independent blinded statistical analysis.

Only after all data had been analyzed and two independent drafts had been written the fourth stage consisting of a meeting attended by all involved contributors (A, B, C, D) was scheduled. During this meeting not only written exposition of the results provided independently by C and D and applied statistical methods were compared with each other but also possible reasons for the discrepancies in findings were discussed. The amount of time invested in analysis and draft writing was compared. Eventually, A resolved the key for the meaning of the factor levels and the primary outcome in the data making the results interpretable. In case of major differences in results the following procedure would have been applied: statistical analysis paths would be compared and the reason for the differences would be sought. In case of an error in one of the analyses the more accurate analysis would have been reported.

In the fifth stage all authors (A, B, C, D) were involved in the process of report writing. It was intended that in the final report the identical findings provided by both statisticians (C and D) would be presented only once. In contrast the findings, which disagreed or complemented each other would be described separately. No additional statistical analyses were permitted after the fourth stage. However, an appropriate reformatting (adjustment) of graphs and tables was allowed. The data and the source code used can be obtained under request from the authors.

In the original final study report both statistical contributors wrote their own statistical methods sections, two results sections and two separate discussions concluding by a conjoint discussion. For the sake of compactness only a concise summary of the original final report is provided below.

2.3. Statistical Methods

Data were visualized in two different ways. Statistician C preferred boxplots whereas statistician D favored histograms with superimposed density functions. As far as the distributional assumption is concerned statistician D considered only the Weibull distribution. In contrast, statistician C computed adjusted Anderson–Darling (AD) goodness-of-fit estimates and probability plots in order to clarify the true sampling distribution [11]. Finally, statistician C concentrated on the two-parameter Weibull and the Normal distributions.

Estimation techniques applied by both statisticians differed as well. Statistician C fitted the parameters of the Weibull and Normal distributions by both Least Squares (LS) and Maximum Likelihood (ML) techniques [6,13]. For LS fit the median rank (Benard) default assumption in Minitab was used [6]. The corresponding 95% confidence intervals (95% CI) were computed. In contrast, statistician D provided Weibull parameters LS estimates according to the suggestions in [13] for $n = 40$. Statistician D applied two competing methods: regression of XonY together with median ranks (XonY/median) and regression of YonX with hazen ranks (YonX/hazen). 95% CI were calculated for all estimates according to the procedure suggested in [13] (termed Menon 95% CI).

Statistician C tested Weibull parameters for differences with permutation tests programmed in *R* [27]. A permutation test generates a reference distribution under the null hypothesis (H_0, no difference between groups) by randomly rearranging group labels and computing the value of the test statistic for a lot of such rearrangements. The reference distribution represents values that are plausible under H_0. The value of the test statistic actually observed is compared to the reference distribution and *p*-values are calculated as the fraction of cases in the reference distribution that show a value at least as extreme as the one actually observed. In this study $R = 10{,}000$ permutations were performed and the test statistic was the difference of estimated Weibull parameters between different groups ($\underline{m}_i - \underline{m}_j$ and $\underline{s}_i - \underline{s}_j$). It is zero under H_0 (parameters are the same) and larger differences provide more evidence against H_0. The test for m was always done first since the test for s requires homogeneity in m and was only performed if no significant difference in m was found. In order to compare Weibull parameters among more than two groups a global test was developed. The mean absolute differences between Weibull parameters estimated in all groups under comparison were used as test statistic (mean $|\underline{m}_i - \underline{m}_j|$ and mean $|\underline{s}_i - \underline{s}_j|$). If significant differences were found in the global test, pairwise comparisons between all groups were performed, including correction for multiple comparisons. Under H_0 the probability of at least one significant result in multiple comparisons (family-wise error rate, FWER) will be larger than the nominal type I error rate α (typically $\alpha = 5\%$). A simple method to control FWER is the Bonferroni-Holm [28] method in which the smallest *p*-value is multiplied by the number of individual tests (k), the second smallest by k-1 and so on (with the restrictions that the initial order of *p*-values is kept and all *p*-values < 1), and evaluated at the initial level α:

ordered *p*-values: $p_1 \leqslant p_2 \leqslant \ldots \leqslant p_k$
adjusted *p*-values: $p_i^{adj} = \max_{j \leqslant i} [(k-j+1)\, p_j]_1$
where $[x]_1 = \min(x,1)$ and $i = 1, 2, \ldots, k$

Such Bonferroni-Holm adjusted *p*-values are reported when multiple comparisons are performed.

In contrast, statistician C conducted equal shape (standard deviation) and equal scale (mean) Bartlett's modified likelihood ratio tests together with the appropriate Bonferroni post-hoc confidence interval [6,8] with Minitab Version 14 [29]. The impact of SP and ZM factors was evaluated by a general test looking for differences in parameters between all six tested G groups. Results of statistical analyses with $p < 0.05$ were interpreted as statistically significant.

3. Results

Figure 2 shows the distributions of the biaxial flexural strength observations in each tested group. Probability plots for Weibull and Normal assumptions are depicted in Figure 3.

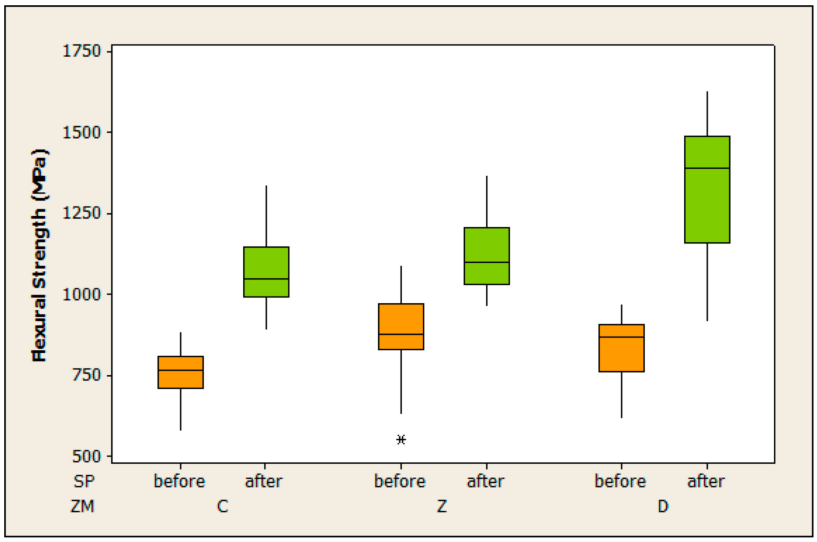

Figure 2. Boxplots for the biaxial flexural strength in each tested group G and ZM/SP levels.

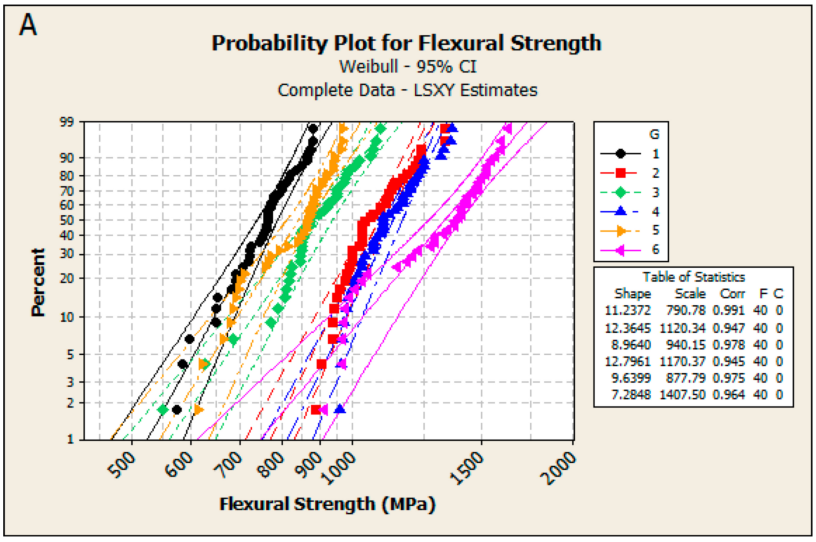

Figure 3. *Cont.*

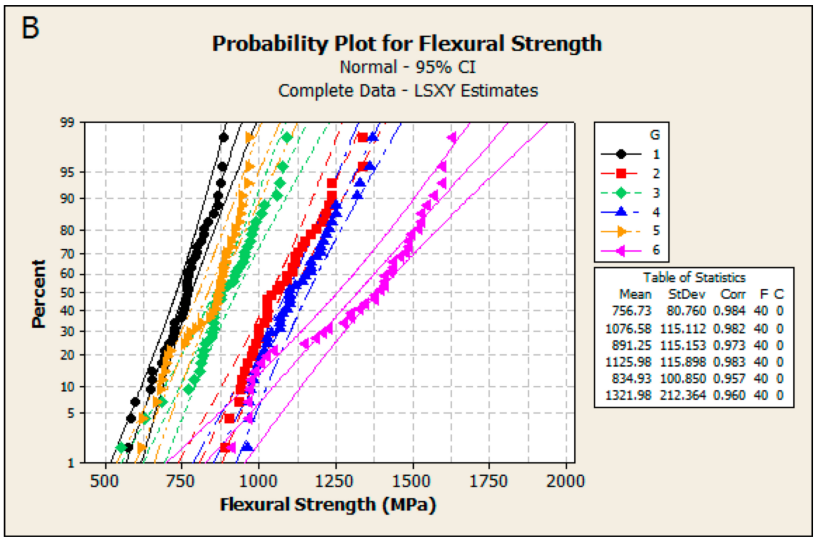

Figure 3. Probability plots for biaxial flexural strength in each tested group G for Weibull LS (**A**) and Normal LS (**B**).

The Anderson-Darling goodness-of-fit estimates for the two-parameter Weibull and Normal distributions showed that the Weibull assumption was better than the Normal one for all groups with exception of C/after and Z/after (see [3] Table 4). Although Anderson-Darling estimates for Weibull and Normal distributions differed between LS and ML fitting methodologies, their suggestions for the better fit were consistent. While probability plots in Figure 3 indicated discrepancies from straight lines it was impossible to find any alternative uniformly optimal fitting sampling distribution in all tested groups.

Generally, biaxial flexural strength observations appeared to be larger in groups G2, G4 and G6 within SP = "after" condition. Variance of strength values was small in G1, at a medium level in G2 to G5, and large in G6 (Table 2, Figure 2).

The descriptive statistics for the Weibull distribution are shown in Table 3. As the Weibull parameters estimates obtained by the XonY/median and Benard LS approaches were comparable to the YonX/hazen one, only the later results are reported. For completeness the ML estimates are provided.

Results obtained by Weibull and Normal estimation methods were comparable. Generally, larger mean biaxial flexural strength values resulted in larger estimates for s and lower variance of the biaxial flexural strength in larger estimates for m. Estimates for s were e.g., higher in the groups with rather large outcomes (G2, G4, G6). The large variance in G6 resulted in small m, the low variance in G1 in a rather large m. Generally, G2 and G4 appeared to be very similar, characterized by large s and m. G3 and G5, in contrast, were characterized by both small s and m. Taking 95% CI into account, differences in s compared to m were more prominent. In particular, s appeared to be much higher in groups G2, G4 and G6 (with after SP).

There were differences in Weibull estimates obtained by the ML and LS (YonX/hazen) approaches (Table 3). Their relevance is visualized in histograms with superimposed Weibull and Normal density functions in Figure 4.

Table 3. Point and interval estimates of the Weibull parameters modulus (**m**) and scale (**s**) based on ML or YonX/hazen, respectively.

Tested Groups	ZM	SP	Method	m	95% CI (m)	s	95% CI (s)
G1	C	before	ML	11.4	[8.9, 14.6]	791	[768, 814]
			YonX/hazen	11.4	[8.2, 15.9]	791	[768, 814]
G2	C	after	ML	9.6	[7.6, 12.0]	1129	[1090, 1168]
			YonX/hazen	11.4	[8.2, 15.9]	1126	[1093, 1159]
G3	Z	before	ML	9.4	[7.3, 11.9]	939	[906, 972]
			YonX/hazen	8.9	[6.4, 12.4]	942	[907, 977]
G4	Z	after	ML	10.3	[8.1, 13.0]	1178	[1141, 1217]
			YonX/hazen	11.8	[8.5, 16.3]	1176	[1143, 1210]
G5	D	before	ML	10.9	[8.3, 14.1]	877	[851, 904]
			YonX/hazen	9.5	[6.8, 13.2]	880	[849, 911]
G6	D	after	ML	7.9	[6.1, 10.3]	1409	[1352, 1468]
			YonX/hazen	7.0	[5.0, 9.7]	1414	[1348, 1484]

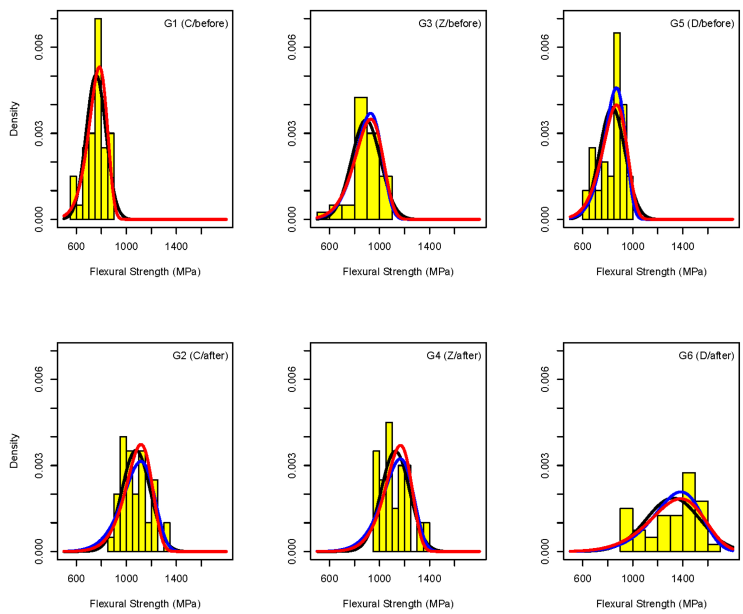

Figure 4. Histogram of biaxial flexural strength with superimposed densities: Weibull estimated by ML (**blue**) and YonX/hazen (**red**) and Normal (**black**). Factor G and levels of ZM/SP are indicated on top of each plot.

Permutation tests were applied to test for H_0 that there are no differences between Weibull parameters in the groups of interest. Tests for differences in s were only made if no differences in m were found, since they rely on homogeneity in m. As the results of the permutation tests applied to XonY/median and YonX/hazen estimates were comparable, only the latter are reported in Table 4.

Table 4. Permutation tests for differences in Weibull parameters estimated by YonX/hazen. (**a**) Differences between two levels of SP within each level of ZM. Test statistic: $\mathbf{m_i} - \mathbf{m_j}$ and $\mathbf{s_i} - \mathbf{s_j}$; (**b**) Global test for differences between the three levels of ZM within each level of SP. Test statistic: mean$|\mathbf{m_i} - \mathbf{m_j}|$ and mean$|\mathbf{s_i} - \mathbf{s_j}|$; (**c**) Pairwise tests to B for differences between three levels of ZM within each level of SP. Test statistic: $\mathbf{m_i} - \mathbf{m_j}$ and $\mathbf{s_i} - \mathbf{s_j}$. *p*-values were adjusted by the Bonferroni-Holm method.

	Comparison	Condition	Test Statistic		*p*-Values	
			m	**s**	**m**	**s**
(a)	before-after (G1-G2)	C	0.014	−334.9	0.9830	<0.0001
	before-after (G3-G4)	Z	−2.831	−234.4	0.0370	NA
	before-after (G5-G6)	D	2.494	−534.7	<0.0001	NA
(b)	C-Z-D (G1-G3-G5)	before	1.664	100.4	0.2010	<0.0001
	C-Z-D (G2-G4-G6)	after	3.186	192.4	<0.0001	NA
(c)	C-Z (G1-G3)	before	2.496	−150.6	NA	<0.0001
	C-D (G1-G5)	before	1.948	−88.8	NA	<0.0001
	Z-D (G3-G5)	before	−0.547	61.8	NA	0.0080
	C-Z (G2-G4)	after	−0.350	−50.2	0.8210	0.0840
	C-D (G2-G6)	after	4.428	−288.7	<0.0001	NA
	Z-D (G4-G6)	after	4.778	−238.5	<0.0001	NA

With respect to the impact of SP significant differences were found for zirconia Z between G3 and G4 ($p = 0.037$) as well as for zirconia D between G5 and G6 ($p < 0.0001$) (Table 4a). As there was no evidence against homogeneity in m for zirconia C, s was analyzed and found to differ significantly ($p < 0.0001$).

In order to analyze the effect of ZM comparisons of groups with constant SP were made (Table 4b). Significant differences in m were found between groups G2-G4-G6 (C-Z-D/after) ($p < 0.0001$). Accordingly, s was only tested for G1-G3-G5 (C-Z-D/before) and found to differ significantly ($p < 0.0001$).

As the global test for m was significant for G2-G4-G6 (C-Z-D/after), all pairwise comparisons G2-G4, G2-G6, G4-G6 were computed (Table 4c). Significant differences after Bonferroni-Holm correction for multiple testing were found for G2-G6 and G4-G6. The non-significant G2-G4 comparison for m ($p = 0.821$) was tested for differences in s and did also not reach significance ($p = 0.084$). Pairwise comparisons for s were also made in the groups G1-G3-G5 (C-Z-D/before), as the global test for s was significant. All of the possible pairwise tests were significant after Bonferroni-Holm correction (Table 4c).

Alternatively, the general analysis comparing all six tested G groups found that for Weibull (LS) modulus m6 < (m2, m4) but (m1, m3, m5, m6) and (m1, m3, m5, m2, m4) ($p = 0.006$) and characteristic strength s1 < (s3, s5) < (s2, s4) < s6 ($p < 0.001$). In contrast, for Weibull (ML) modulus no differences between moduli in all six factor levels were found (m1, m3, m5, m2, m4, m6) ($p = 0.409$) and the findings for the characteristic strength agreed with those for LS fitting technique leading to s1 < (s3, s5) < (s2, s4) < s6 with $p < 0.001$.

The results of the tests for LS and ML fitting techniques conducted under Normal sampling distribution assumption agreed perfectly well leading to sd6 > (sd1, sd3, sd5, sd2, sd4) ($p < 0.001$) and mean1 < (mean3, mean5) < (mean2, mean4) < mean6 ($p < 0.001$).

4. Discussion

Strictly speaking the Weibull distributional assumption is only a special case of a more general approach to strength distributions [5]. Weibull assumption is preferred in practice due to its flexibility. It provides nice fits to strength data [5,6]. It has also strong probabilistic foundations for strength measurements. Although the Weibull distribution is more appropriate to analyze strength data of

brittle materials than the Normal one [5], the Normal one is frequently used due to its convenience and availability in general-purpose statistical programs. It is difficult to discern the appropriateness and implications of Weibull and Normal assumptions in practice, notably, for small sample sizes [5].

As a rule of thumb at least 30 measurements in each tested group are necessary to be able to recognize the true sampling distribution at all [11]. Estimation of the scale (s) of the two-parameter Weibull distribution and mean of the Normal one requires fewer observations than the estimation of modulus (m) and sd, respectively [9]. Although Nohut [7] shows that for sample sizes below 150 per group no clear discrimination between Weibull and Normal sampling distributions is possible.

Abernethy [11] suggested a conservative approach to Weibull analysis. He recommended application of a two-parameter Weibull distribution as a working assumption irrespective of curvature in the probability plots. Given the respectable number of 40 observations in each group statistician C investigated the true underlying sampling distribution and found that no uniformly optimal fitting distribution could be suggested. The two-parameter Weibull distributional assumption was in 4 out of 6 tested groups better than the Normal one for the data at hand. Therefore, we considered both Weibull and Normal distributional sampling distributions for data analysis. We think that the dilemma of discerning Weibull and Normal distributions will stay an unsolved problem. As neither Weibull nor Normal distributional assumptions fit the data in all tested groups perfectly well, the truth seems to lie somewhere in between. In this respect our approach is an extension of the Abernethy's [11] conservative approach to two different distributional assumptions. The use of two different distributional assumptions protects us from being over-optimistic.

Our analysis indicated a strong evidence for the relevance of both ZM and SP on the biaxial flexural strength values. Generally, SP compared to ZM had more impact on mean biaxial flexural strength and estimated scale s than ZM. Wet polishing after sintering appears to generally increase s. Within this group, C and Z zirconia materials were not distinguishable and D reduced modulus m. However, a clear pattern for m was not observed and it seems to be influenced by the combination of ZM and SP. In particular, the combination in the group G6 (D/after) results in small m and a high variance in biaxial flexural strength. Assuming large values for both s and m would be beneficial, the combination in groups G2 (C/after) and G4 (Z/after) performed best.

After unblinding the project, we realized that the treatment of the specimens preparation in clusters might actually led to problems for group G6 (D/after). We were unable to identify the real cause. However, probability plots (Figure 3) and histograms (Figure 4) clearly indicate that the distribution in G6 (D/after) is bi- rather than uni-modal suggesting that possibly a cluster of specimens in this group had different properties. Interestingly, this cluster led to higher sd (lower m) estimates in G6 (D/after) provoking our recommendation not to use this technique in practice.

Moreover, the following additional potential uncertainties for Table 1a have been identified in the main experiment [3]: material (manufacturer, lot number), sintering temperature inaccuracy (unequal temperature inside the oven, too early opening of the oven's door), measurement error (micrometer screw precision), positioning within the testing device (splinters), measurement error of the testing machine for biaxial flexural strength measurements. Due to the oven or the polishing machine capacity specimens were frequently handled in clusters. If an inaccuracy occurred then all of the specimens in a cluster were possibly affected by it.

We obtained discordant decisions from the global Weibull test depending on the estimation method. LS indicated that there is a difference in moduli between the six tested groups. In contrast, ML found no evidence for any differences. One possible approach to dissolve this discrepancy would be to look closer at the relevance of the differences. According to Nelson [9] the sample size of 40 observations in each tested group implies that if the true modulus equals 10, then with probability 99.5% the modulus estimates should be found in interval (6.2, 16.0). Interestingly, modulus estimates in Table 3 can actually be found in this interval. This argument suggests that there are no genuine differences in moduli between the six tested G groups supporting the Weibull/ML finding. A second argument trying to dissolve the LS/ML discrepancy would be to apply a more stringent significance

level 0.005 instead of the inappropriately high but common 0.05 one as suggested by Johnson [30]. Johnson's suggestion aims for increasing reproducibility of the reported findings in the scientific literature. Application of a lower $\alpha = 0.005$ level for testing would imply an agreement of the global conclusions from LS/ML estimation techniques for the Weibull distributional assumption: there would be no evidence for differences in moduli between the six groups and the characteristic strengths would be ordered G1 < (G3, G5) < (G2, G4) < G6. For the Normal assumption the ordering of the means would agree with that for Weibull characteristic strengths. However, the standard deviation in G6 would be larger than standard deviations in all other tested groups. In this respect Weibull and Normal findings would still disagree.

Perhaps it is a misperception that the results of a statistical analysis have to be unique. We suggested two independent blinded statistical analyses by two statisticians who were unaware of the factor level's assignment and the meaning of the primary outcome. Our primary goal was not to obtain identical results but rather to get a better picture of the underlying truth. Therefore, we deliberately decided not to apply any rigid pre-specified guidelines for statistical analysis, any detailed analysis plan including rigorously predefined objectives and any inflexible specifications of the statistical methods for the primary endpoint prior to the study.

Gøtzsche [17] apprehends data analysis as a highly subjective process vulnerable to bias and suggests that actually two manuscripts should be written and both manuscripts must be completed and approved by the authors before the code is broken. We rather aimed for a single final report containing the identical results with additional sections describing explicitly possible discrepancies. We found that it is easier for a reader to concentrate on the main findings and grasp the differences.

It was the first time we applied "independent statistical analyses" and "parallel manuscript writing" to a "blinded data set". The "blinding of the data set" was very unfamiliar to both study supervisor A, as well as to the statisticians C, D. Usually, A would supply a statistician with unlimited information about the measurements and the experiment properties prior to the statistical analysis. On the other hand, both statisticians would ask for as much background information as possible about the meaning of the variables and the expected effects prior to the data analysis in order to correctly understand the data and to provide optimal statistical analysis for the project at hand. "Blinding of the data set" technique was clearly in conflict with our daily statistical consultation experience. Our experiment forced us, however, into an unusual situation and possibly inspired a more impartial statistical analysis.

The "independent statistical analyses" had definitely a considerable influence on both statisticians. They provided a more watchful analysis, tended to spontaneous self-verification and an increased alertness by the simple fact that the results would be verified and critically discussed by an independent statistician. We experienced that "independent statistical analysis" is a very powerful tool to ensure good statistical analysis. It is even stronger than a simple correctness check of the reasoning path provided by a statistician. It was a very inspiring experience. It led to a deeper and more thorough analysis. To our surprise both independently working statisticians went through two differing analysis paths for data visualization, estimation and testing and arrived at complementary results.

The "parallel manuscript writing" technique was also new for us. It forced both statisticians to be clear about their final conclusions, to tie themselves down to one explicit ultimate interpretation and to present concisely their findings in tables and figures. Both independently written manuscripts could be conveniently merged into one final report at the fifth stage of our study (Figure 1).

Given two parallel drafts the final writing of the report concentrated on collation of both abstract, results and discussion sections, notation unification and consolidation of tables and figures. We had to shorten the exposition and to counteract repetitions. For a similar future study, it would be beneficial to agree on the notation beforehand. Furthermore, it would be advantageous to put a limit on the length of paragraphs written independently by biostatisticians.

One possible limitation of a blinded statistical analysis is its inefficiency [17]. There is extra work needed to conduct statistical analysis twice and to produce two independent drafts. Miller

& Stewart [23] point out that a requirement for the biostatistician to remain blinded adds a level of complexity to the study implementation. When planning the amount of time for data analysis we followed suggestions by Pocock [18,31] who stressed that one should allow enough time for analyzing the measurements as good-quality statistical analysis cannot be achieved overnight so that an adequate provision of time for the analysis and interpretation of trial data should be recognized when planning a study. Therefore, we warranted both data analysts 3 weeks for the data analysis and their independent writing of "Results", "Summary", "Conclusions" and "Rationale for the choice of statistical methods" sections. In fact, each of the statisticians spent at least 25 h for analysis and draft writing. We feel that the time interval of 3 weeks provided an appropriate amount of time for a deepened statistical analysis. We admit that the expenditure of time was much increased as compared with a common non-parallel statistical analysis. In our case, however, such an effort was justified.

The statistical analysis paths took by the statisticians differed. Statistician D preferred the permutation tests approach for Weibull statistics. This innovative idea is motivated by the fact that permutation tests are non-parametric, allow for any arbitrary test statistic and do not make any assumptions about the distribution of this test statistic under the null hypothesis (H_0, no difference between groups). It can be programmed in R and is independent from sophisticated statistical software [6,11,12]. One deficiency of the approach was that the permutation tests for post-hoc tests in Table 4c had to be programmed separately and the application of the Bonferroni-Holm correction for the pairwise tests (Table 4c) was necessary. Interestingly, statistician D arrived at different conclusions mainly by not applying any global analysis. Separate comparisons in each factor level suffered from several drawbacks such as multiple comparisons, difficult interpretation and complicated design. An ANOVA framework including both factors would be better suited for the specific problem. However, normally distributed variables with equal variances would be required. Methods form transformation of Weibull distributed into normally distributed variables are available and could be a promising alternative [32].

The analysis path suggested by statistician C considered both Weibull and Normal distributions using the graphical and statistical facilities available in Minitab. It applied global likelihood ratio tests together with integrated post-hoc confidence intervals. No additional programming was required. Finally, this approach was considered for the analysis of the larger data set with a greater number of tested groups in [3].

Table 5 contains the final workflow for a pragmatic statistical analysis of flexural strength data developed in our study and successfully applied in [3]. It consists of seven steps starting with visualization of the measurements (Step 1), choice of at least one reasonable distributional assumption (Step 2), estimation of the parameters of a distribution (Step 3), check of the goodness-of-fit (Step 4) and estimation of 95%CI (Step 5). In order to understand which tested groups are better than others in terms of location and spread we have to apply appropriate statistical tests (Step 6). These tests require a specification of a single distribution applicable to all tested groups. In practice, however, it is quite impossible to find a single distributional assumption fitting all groups equally well. Therefore, we extended Abernethy's [11] conservative approach to two different reasonable distributional assumptions (Weibull and Normal in our case) as a working hypothesis (Step 2). As a consequence, we have to compare the results obtained under differing distributional assumptions (Step 7).

Table 5. A recommended statistical analysis workflow for flexural strength measurements. If not otherwise indicated [29] can be used (path: Stat/Reliability-Survival/Distribution Analysis (Right Censoring)/Parametric Distribution Analysis/).

Step	Decision/Action
Step 1	**Data visualization in each tested group** (see Figures 2 and 4): Check data visually by histograms, scatterplots and/or boxplots. Are there any outlying observations? If yes, check if they are possibly typing errors and correct them. Are histograms approximately symmetric in each tested group? If no, you may try to transform the measurements.
Step 2	**Distributional assumption for measurements** (see Section 1.1): Do you think that each measurement consists of a possibly large number of independent random fluctuations? If yes, go for a Normality assumption directly (for approximately symmetrical histograms) or after a (logarithmic) transformation of measurements. Do you believe in the "weakest link" process generating your data? If yes, go for a Weibull assumption. If both assumptions seem to be reasonable use both Normal and Weibull distributional assumptions for your working hypothesis.
Step 3	**Descriptive statistics: Estimation of parameters in each tested group:** Under Normality assumption: compute mean and standard deviation (sd). Under Weibull assumption: compute the characteristic strength (s) and modulus (m). See the open source Excel-calculator (Appendix C in [13]). Remember: (s "=" mean) and (m "=" 1/sd) (see Section 1.1)
Step 4	**Check the goodness-of-fit in each tested group:** Compute the goodness-of-fit estimates. Generate probability plots (see Figure 3) and check if they are linear. In case of approximate linearity the assumed distribution fits the data well. In case of clear non-linearity interpret the results with caution.
Step 5	**Estimation of 95%CI for parameters in each tested group:** Under Normality assumption: 95% CI (mean) and 95% CI (sd). Under Weibull assumption: 95% CI (s) and 95% CI (m). See the open source Excel-calculator (Appendix C in [13]).
Step 6	**Are there any differences between tested groups?** Normal mean: Apply an Analysis of Variance (ANOVA). Normal sd: Apply a Levene-Test. Weibull s and m: Apply the Bartlett's modified likelihood ratio tests.
Step 7	**Check the results:** Critically check if the results obtained in Step 6 agree with the graphs generated in Step 1. If you applied both the Normal and the Weibull assumptions critically check if the results obtained in Step 6 are comparable. Remember: (s "=" mean) and (m "=" 1/sd) (See Section 1.1).

We did not intend to cover the whole range of possible statistical approaches for the analysis of the strength data. We admit that the true range of the varying inter-personal statistical approaches is not well reflected by just two participating statisticians. For example, some researchers might opt for an equivalent volume approach within the Weibull analysis [5,10]. Others might consider a log-normal distributional assumption for the data analysis instead of Weibull or Normal ones. They would argue that the log-normal distribution may be a useful alternative for stabilizing and reducing variance as well as giving a sensible linearization of probability plots. An ordinary ANOVA applied to lognormal data may be robust enough to answer the questions in a very rapid and straightforward manner.

Despite these possible shortcomings, our blinded approach was a very helpful tool at the stage of finding a practicable statistical method for the final data analysis in [3]. It complements the recommendations by Hannigan and Lynch [33] and should definitely be taken into consideration.

5. Conclusions

Within the limitations of this investigation it can be concluded that:

- Zirconia specimen preparation method has an impact on characteristic strength (s) and mean of the biaxial flexural strength but in majority of tested groups practically no relevant impact on modulus (m) and standard deviation (sd) of the results.
- All three tested zirconia materials showed different characteristics strengths and mean flexural strength results. Group G6 (D/after) showed higher spread leading to smaller modulus and increased sd estimates.
- The "blinded data set", "independent statistical analyses" and "parallel manuscript writing" techniques had an influence on the findings for strength data. The impact of "independent statistical analyses" was most pronounced.
- Statistical analysis paths taken by both independently working biostatisticians differed.
- The major difference in the findings was caused by two alternative distributional assumptions (Weibull/Normal) and alternative estimation methods (LS/ML).

Acknowledgments: The authors would like to thank Amann Girrbach, Wieland + Dental and Dental Direkt for supporting this study with materials. We thank Lukas Bütikofer for his contributions to the project.

Author Contributions: Malgorzata Roos: idea, designed the experiments, performed the statistical analyses, wrote the manuscript, Christine Schatz: specimen fabrication, specimen measurements and proofread the manuscript, Bogna Stawarczyk: designed the experiments, co-wrote the manuscript.

Conflicts of Interest: The authors report no conflict of interest. The authors alone are responsible for the content and writing of the manuscript.

References

1. Stawarczyk, B.; Özcan, M.; Hallmann, L.; Ender, A.; Mehl, A.; Hämmerle, C.H.F. The effect of zirconia sintering temperature on flexural strength, grain size, and contrast ratio. *Clin. Oral Investig.* **2012**, *17*, 269–274. [CrossRef] [PubMed]
2. Stawarczyk, B.; Özcan, M.; Trottmann, A.; Hämmerle, C.H.F.; Roos, M. Evaluation of flexural strength of hipped and presintered zirconia using different estimation methods of Weibull statistics. *J. Mech. Behav. Biomed. Mater.* **2012**, *10*, 227–234. [CrossRef] [PubMed]
3. Schatz, C.; Strickstrock, M.; Roos, M.; Edelhoff, D.; Eichberger, M.; Zylla, I.M.; Stawarczyk, B. Influence of specimen preparation and test methods on the flexural strength results of monolithic zirconia materials. *Materials* **2016**. [CrossRef]
4. McCabe, J.F.; Carrick, T.E. A statistical approach to the mechanical testing of dental materials. *Dent. Mater.* **1986**, *2*, 139–142. [CrossRef]
5. Danzer, R.; Lube, T.; Supancic, P.; Damani, R. Fracture of ceramics. *Adv. Eng. Mater.* **2008**, *10*, 275–298. [CrossRef]
6. Roos, M.; Stawarczyk, B. Evaluation of bond strength of resin cements using different general-purpose statistical software packages for two-parameter Weibull statistics. *Dent. Mater.* **2012**, *28*, e76–e88. [CrossRef] [PubMed]
7. Nohut, S. Influence of sample size on strength distribution of advanced ceramics. *Ceram. Int.* **2014**, *40*, 4285–4295. [CrossRef]
8. Stawarczyk, B.; Özcan, M.; Hämmerle, C.H.F.; Roos, M. The fracture load and failure types of veneered anterior zirconia crowns: An analysis of normal and Weibull distribution of complete and censored data. *Dent. Mater.* **2012**, *28*, 478–487. [CrossRef] [PubMed]
9. Nelson, W. *Applied Life Data Analysis*; John Wiley & Sons, Inc.: New York, NY, USA, 1982.
10. Quinn, J.B.; Quinn, G.D. A practical and systematic review of Weibull statistics for reporting strengths of dental materials. *Dent. Mater.* **2010**, *26*, 135–147. [CrossRef] [PubMed]
11. Abernethy, R. *The New Weibull Handbook*, 5th ed.; Robert, B.A., Ed.; Robert. Abernethy: North Palm Beach, FL, USA, 2009.

12. Rinne, H. *The Weibull Distribution: A Handbook*; Chapman & Hall/CRC Press: Giessen, Germany, 2009.
13. Bütikofer, L.; Stawarczyk, B.; Roos, M. Two regression methods for estimation of a two-parameter Weibull distribution for reliability of dental materials. *Dent. Mater.* **2015**, *31*, e33–e50. [CrossRef] [PubMed]
14. Greenland, S. Bayesian perspectives for epidemiological research: I. Foundations and basic methods. *Int. J. Epid.* **2006**, *35*, 765–775. [CrossRef] [PubMed]
15. Pocock, S.J.; Hughes, M.D.; Lee, R.J. Statistical problems in the reporting of clinical trials. *N. Eng. J. Med.* **1987**, *317*, 426–432. [CrossRef] [PubMed]
16. Polit, D.F. Blinding during the analysis of research data. *Int. J. Nursing Stud.* **2011**, *48*, 636–641. [CrossRef] [PubMed]
17. Gøtzsche, P.C. Blinding during data analysis and writing of manuscripts. *Contr. Clin. Trials* **1996**, *17*, 285–293. [CrossRef]
18. Pocock, S.J. *Clinical Trials: A Practical Approach*; John Wiley & Sons: New York, NY, USA, 1983.
19. Doll, R. Controlled trials: The 1948 watershed. *Br. Med. J.* **1998**, *317*, 1217–1220. [CrossRef]
20. Schulz, K.F.; Chalmers, I.; Altman, D.G. The landscape and lexicon of blinding in randomized trials. *Ann. Internal. Med.* **2002**, *136*, 254–259. [CrossRef]
21. Schulz, K.F.; Altman, D.G.; Moher, D. CONSORT 2010 Statement: Updated guidelines for reporting parallel group randomization trials. *Ann. Internal. Med.* **2010**, *152*, 726–732. [CrossRef] [PubMed]
22. Moher, D.; Hopewell, S.; Schulz, K.F.; Montori, V.; Gøtzsche, P.C.; Devereaux, P.J.; Elbourne, D.; Egger, M.; Altman, D. CONSORT 2010 Explanation and Elaboration: Updated guidelines for parallel group randomized trials. *J. Clin. Epidemiol.* **2010**, *340*, c869. [CrossRef] [PubMed]
23. Miller, L.E.; Stewart, M.E. The blind leading the blind: Use and misuse of blinding randomized controlled trials. *Contemp. Clin. Trials* **2011**, *32*, 240–243. [CrossRef] [PubMed]
24. Schulz, K.F.; Grimes, D.A. Blinding in randomised trials: Hiding who got what. *Lancet* **2002**, *359*, 696–700. [CrossRef]
25. Klein, J.R.; Roodman, A. Blind analysis in nuclear and particle physics. *Annu. Rev. Nucl. Part. Sci.* **2005**, *55*, 141–163. [CrossRef]
26. *Dentistry-Ceramic Materials*; ISO. 6872:2008; International Organization for Standardization: Geneva, Switzerland, 2008.
27. R Foundation for Statistical Computing. *A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2010. Available online: http://www.R-project.org (accessed on 22 June 2016).
28. Holm, S. A simple sequentially rejective multiple test procedure. *Scand. J. Stat.* **1979**, *6*, 65–70.
29. Ryan, B.F.; Joiner, B.L. *Minitab 14 Statistical Software*; Minitab, Inc.: State College, PA, USA; pp. 1972–2005.
30. Johnson, V.E. Revised standards for statistical evidence. *Proc. Natl. Acad. Sci. USA* **2013**, *48*, 19313–19317. [CrossRef] [PubMed]
31. Pocock, S.J. Clinical trials with multiple outcomes: A statistical perspective on their design, analysis and interpretation. *Control. Clin. Trials* **1997**, *18*, 530–545. [CrossRef]
32. Watthanacheewakul, L. Analysis of Variance with Weibull Data. In Proceedings of the International MultiConference of Engineers and Computer Scientists, Hong Kong, China, 16–18 March 2011; Volume 3, pp. 2051–2056.
33. Hannigan, A.; Lynch, C.D. Statistical methodology in oral and dental research: Pitfalls and recommendations. *J. Dent.* **2013**, *41*, 385–392. [CrossRef] [PubMed]

© 2016 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (http://creativecommons.org/licenses/by/4.0/).

3. Diskussion

Im Folgenden werden die zwei vorgelegten Artikel einzeln diskutiert.

3.1. Einfluss der Prüfkörperherstellung und der Prüfmethode auf die Biegefestigkeit von monolithischen Zirkonoxidmaterialien

Soweit den Autoren bekannt ist, existierten zum Zeitpunkt dieser Studie keine Daten in der Literatur, die monolithische Zirkonoxid-Keramiken mit unterschiedlichen Biegeprüfmethoden testen und miteinander vergleichen. Um unsere Daten diskutieren zu können, wird im Folgenden auch auf Studien mit anderen keramischen Materialen Bezug genommen.

Die in dieser Studie aufgestellte Nullhypothese beinhaltet, dass die angewendeten Biegeprüfungen die Biegefestigkeitsresultate nicht beeinflussen würden. Es konnte jedoch ein deutlicher Unterschied zwischen den mit der Biaxialen-Biegeprüfung und den mit der Vier-Punkt-Biegeprüfung getesteten Gruppen festgestellt werden, wobei die Vier-Punkt-Biegeprüfung die niedrigsten Biegefestigkeitswerte zeigte. Bei dieser Biegeprüfung wird ein größerer Volumenanteil des Materials einer Spannung ausgesetzt. Die Wahrscheinlichkeit einer Rissentstehung wird somit, verglichen mit der Drei-Punkt-Biegeprüfung höher, was sich auch in niedrigeren Biegefestigkeitswerten wiederspiegeln kann [10]. In unseren Daten waren die Gruppen der Vier-Punkt-Biegeprüfung in allen bis auf eine statistisch unterschiedlich von denjenigen Gruppen, die mit der Drei-Punkt-Biegeprüfung getestet wurden. Diese Beobachtung ist in Übereinstimmung mit den Daten einer früheren Studie, in der Veneer-Keramiken für Zirkonoxid untersucht wurden [26].

In einer anderen Studie mit Zirkonoxid-Kernmaterialien präsentierte die Biaxiale-Biegeprüfung höhere und statistisch abgrenzbare Biegefestigkeitswerte als die Drei-Punkt-Biegeprüfung [27]. In allen bis auf eine Gruppe zeigten auch unsere Daten die größte mittlere Biegefestigkeit in den biaxialen Gruppen. Eine klare Unterscheidung zwischen Drei-Punkt-Biegeprüfung und Biaxialer-Biegeprüfung konnte jedoch nur in zwei Gruppen getroffen werden. Die höheren Werte der biaxialen Prüfmethode können auf den vernachlässigbaren Effekt von unerwünschten Fehlern oder Unebenheiten an den Kanten der Prüfkörper zurückzuführen sein [28,29]. In einer weiteren Studie untersuchten Jin *et al.* [30] den Einfluss aller drei Prüfmethoden an unterschiedlichen Keramikarten. Obwohl in dieser Studie kein Zirkonoxidmaterial getestet wurde, identifizierten die Autoren die kleinsten Biegefestigkeitswerte mit der Vier-Punkt-Biegeprüfung und konnten ebenfalls keine klare Unterscheidung zwischen der Biaxialen- und der Drei-Punkt-Biegeprüfung vorfinden. Eine Rangordnung über die Biegefestigkeitswerte in unserer Studie kann wie folgt etabliert werden: Biaxiale-Biegeprüfung > Drei-Punkt-Biegeprüfung > Vier-Punkt-Biegeprüfung.

Betrachtet man den Einfluss der Biegeprüfung auf den Weibullmodul so wurde festgestellt, dass nur maschinell nach dem Sintern poliertes, mit der Drei-Punkt-Biegeprüfung getestetes Zirkonoxid C und ebenfalls nach dem Sintern poliertes, mit der Biaxialen-Biegeprüfung getestetes Zirkonoxid Z einen höheren Weibullmodul aufwies. Weder die Untersuchung der Oberflächenrauhigkeit noch die Rasterelektronenmikroskopaufnahmen können diese Beobachtungen erklären.
Die zweite aufgestellte Hypothese beinhaltete, dass die Art der Prüfkörperherstellung weder die Biegefestigkeit, noch die Zuverlässigkeit des Gefüges von den getesteten Zirkonoxidmaterialien beeinflussen würde. Die Zuverlässigkeit eines Gefüges wird mit dem Weibullmodul beschrieben, dabei steht ein hoher Weibullmodul für eine statistisch gesehen höhere Zuverlässigkeit. Signifikant höhere Werte der Biegefestigkeit zeigten die Prüfkörper, welche nach dem Sintern poliert wurden, wohingegen kein genereller Trend der Werte für die Weibullmoduli beobachtet werden konnte. Der Weibullmodul nahm für das in der Drei-Punkt-Biegeprüfung getestete und nach dem Sintern polierte Zirkonoxidmaterial C sogar höhere Werte an im Vergleich zu den vor dem Sintern polierten Materialien. Somit kann der erste Teil der aufgestellten Hypothese nicht bestätigt werden. Der zweite Teil der Hypothese kann generell innerhalb der Biegefestigkeitsprüfgruppen und Zirkonoxidmaterialgruppen akzeptiert werden. Bei genauer Betrachtung kann ein statistisch gesehen, jedoch nicht signifikant, kleinerer Weibullmodul bei den nach dem Sintern polierten und in der Vier-Punkt-Biegeprüfung getesteten Gruppen der Zirkonoxidmaterialen C und Z festgestellt werden. Das lässt die Vermutung zu, dass das maschinelle Polieren nach dem Sintern Fehler im Gefüge induzieren könnte. Dahingegen ist die Wahrscheinlichkeit einen kritischen Gefügefehler aufzudecken natürlich größer, wenn ein größerer Anteil des zu testenden Materials unter Spannung gesetzt wird. Als kritischer Gefügefehler wird der Fehler oder das Element eines Materials bezeichnet, welcher zum Versagen des Materials führen kann. Dieser Effekt wird auch als „Weakest-Link“ Hypothese bezeichnet [31]. Die in den Rasterelektronenmikroskop-Aufnahmen zu sehenden Oberflächenkratzer der Prüfkörper lassen vermuten, dass diese als Entstehungsort von Rissen agiert haben könnten. Die Weibullmoduli bewegten sich zwischen 5.1 und 16.5, was sich mit den in der Literatur auffindbaren Werten für Y-TZP Keramiken deckt [25,32]. Die Biegefestigkeit von Zirkonoxidmaterialen wird in der Literatur unterschiedlich angegeben, je nachdem welche Methode zur Oberflächenveränderung oder Oberflächenbehandlung durchgeführt wurde. Verschiedene Autoren beschreiben eine Abnahme der Biegefestigkeit und des Weibullmoduls von Zirkonoxidmaterialien, wenn die Oberfläche einem Schleifprozess unterworfen wurde

[14,18,33,34]. Es wird jedoch auch eine Zunahme im Mittelwert der Biegefestigkeit und eine Abnahme der Zuverlässigkeit des Materials beobachtet, wenn feinkörnige Diamantbohrer für den Schleifprozess verwendet werden [35] und wenn die Prüfkörper manuell unter Wasserkühlung bearbeitet werden [36], wobei in der letztgenannten Studie jedoch keine Angaben über den Weibullmodul gemacht werden. Eine Zunahme der Biegefestigkeit und eine Abnahme des Weibullmoduls wird beobachtet, wenn die Oberfläche der Prüfkörper mit Korund sandgestrahlt wurde [18]. Diese Zunahme der Biegefestigkeit kann durch die bei Zirkonoxidmaterial auftretende Phasentransformation an der Oberfläche des Materials erklärt werden. Man beobachtet einen erhöhten Anteil an monokliner Phase, wenn Prüfkörper nach dem Sintern einer Oberflächenbearbeitung unterzogen wurden [18,37]. Dieser Mechanismus wird auch als „Transformation Toughening“ bezeichnet. Er beschreibt ein Phasenumwandlungs- und Verstärkungsverhalten von Zirkonoxidmaterialen, welches zu komprimierenden Oberflächenspannungen führt und so die Biegefestigkeit des Materials positiv beeinflusst [38,39]. Die nach dem Sintern polierten Prüfkörper wiesen jedoch in dieser Studie keinen erhöhten Anteil an monokliner Phase auf. Der Poliervorgang kann möglicherweise diese monokline Oberflächenschicht beeinflusst haben, indem sie während des Polierprozesses abgetragen wurde. Guazatto *et al.* [17] haben ebenfalls diesen Effekt beobachtet. Die oben geführte Diskussion möglicher Ursprungsorte von Fehlern im Material wird auch von dem Anderson-Darling-Anpassungstest (AD) unterstützt (Tabelle 4). Kleine AD Werte zeigen eine bessere Übereinstimmung der vorliegenden Daten mit einer Verteilungsannahme.

Es konnte jedoch keine optimale Verteilungsannahme der Biegefestigkeitswerte festgestellt werden, die eine perfekte Passung bei allen Testgruppen aufwies. Jedoch zeigen sowohl die Weibull-, als auch die Normalverteilungsannahme eine angemessene und nützliche Annäherung, um alle Testgruppen miteinander vergleichen zu können. Dies wird zudem durch eine starke negative Assoziation zwischen dem Weibullmodul und COV (rho = -0.972, $p < 0.001$) bestärkt. Weibullmodul-Annahmen rangierten von 5.1 bis 16.5, wohingegen die der COV von 0.007-0.223 nahelegen, dass sowohl die Weibull-, als auch die Normalverteilungsannahme auf alle Testgruppen anwendbar sind.

Die dritte Hypothese bestand aus der Annahme, dass die getesteten Zirkonoxidmaterialen keine Unterschiede bezüglich ihrer Biegefestigkeit aufweisen. Diese Hypothese kann bestätigt werden, da kein besonderer Unterschied zwischen den Materialien festgestellt werden konnte. Es wurde beobachtet, dass sich die Zirkonoxidmaterialen unterschiedlich verhalten, abhängig von der Art der Biegefestigkeitprüfmethode und der Poliermethode.

Wie die Literatur zeigt [19], besitzt das Zirkonoxidmaterial D eine signifikant größere Korngröße als die Materialen C und Z. Es wurde jedoch kein Zusammenhang zwischen Korngröße und den getesteten Parametern in unserer Studie festgestellt.

3.2 Zwei unabhängige, prospektiv geplante und verblindete statistische Weibullanalysen der Biegefestigkeitsdaten von Zirkonoxidmaterialien.

Streng genommen ist die Weibullverteilungsannahme nur ein besonderer Fall einer allgemeineren Herangehensweise zur statistischen Annäherung an Festigkeitsverteilungen [31]. Die Weibullverteilung wird in der praktischen Anwendung jedoch vor allem wegen ihrer Flexibilität bevorzugt. Sie bietet eine gute statistische Passung an Festigkeitsdaten [31,22] und gibt der wahrscheinlich zugrunde liegenden Verteilung von Biegefestigkeitsdaten eine solide Basis. Obwohl die Weibullverteilung für die Analyse von Festigkeitsdaten von spröden Werkstoffen angemessener ist als die Annahme einer Normalverteilung [31], wird die Normalverteilung zur Dateninterpretation sehr gerne aufgrund ihrer leichten Anwendbarkeit und Verfügbarkeit in den gängigen statistischen Standardprogrammen herangezogen. Es ist schwierig die Adäquanz und die Implikation einer Weibull- und Normalverteilungsannahme in der Praxis zu unterscheiden. Dies gilt insbesondere für kleine Stichprobengrößen [31]. Eine Faustregel besagt, dass mindestens 30 Messwerte für jede getestete Gruppe vorliegen müssen, um die wahre Stichprobenverteilung erkennen zu können [45]. Die Schätzung des Weibullparameters s der zwei-parametrischen Weibullverteilung und des Mittelwerts der Normalverteilungsannahme benötigt weniger Beobachtungen als die Schätzung des Weibullmoduls m und sd [44]. Nohut [43] fordert eine Stichprobengröße von mindestens 150 pro Gruppe, da sonst keine Unterscheidung zwischen einer Normalverteilungs- oder Weibullverteilungsannahme möglich sei. Abernethy [45] unterbreitet einen vorsichtigen Ansatz zur Weibullanalyse. Er schlägt die Anwendung der zwei-parametrischen Weibullverteilung als Arbeitshypothese vor, ungeachtet wie sich die Kurvatur in den Probability Plots verhält. Durch die hier vorliegende solide Anzahl von 40 möglichen Beobachtungen in jeder Gruppe, analysierte der Statistiker C die wahre zugrunde liegende Stichprobenverteilung. Er stellte fest, dass keine optimal auf alle Gruppen anwendbare statistische Passung vorgeschlagen werden konnte. Die zwei-parametrische Weibullverteilung lieferte in vier von sechs Gruppen des vorliegenden Datensatzes die bessere statistische Passung als die Annahme einer Normalverteilung der Daten, deshalb wurden beide Verteilungsannahmen in der Analyse berücksichtigt. Die Entscheidung für die verschiedenen Verteilungsannahmen bleibt ein ungelöstes Problem. Da weder die Normalverteilungs- noch

die Weibullverteilungsannahme das vorliegende Datenset perfekt beschreibt, liegt die Wahrheit vermutlich zwischen beiden Annahmen. Dies berücksichtigend ist der hier präsentierte Ansatz für zwei verschiedene Verteilungsannahmen eine Weiterführung zu Abernethy's [45] konservativer Herangehensweise. Das Zugrundelegen von zwei Verteilungsannahmen ermöglicht eine realistische und nicht überoptimistische Interpretation der Daten. In unserer durchgeführten Analyse beeinflussen sowohl ZM als auch SP die Biaxialen-Biegefestigkeitswerte. Hierbei hatte SP im Vergleich zu ZM einen größeren Einfluss auf die mittlere Biegefestigkeit und den geschätzten Parameter s als ZM. Vor allem die Poliermethode nach dem Sintern der Prüfkörper scheint den Parameter s zu erhöhen. Innerhalb dieser Gruppe konnten die Zirkonoxidmaterialien C und Z nicht unterschieden werden, außerdem verkleinerte das Material D den Weibullmodul m. Es konnte kein klares Muster für den Weibullmodul m festgestellt werden. Es scheint eine gegenseitige Beeinflussung von ZM und SP vorzuliegen. Die Kombination Gruppe G6 (D/nachher) zeigt einen kleinen Wert für m und eine große Varianz der Biegefestigkeitswerte. Das Vorliegen von hohen s und m Werten hat eine positive Auswirkung für die Bewertung eines Werkstoffes, demnach schnitten die Kombinationen von ZM und SP in den Gruppen G2 (C/nachher) und G4 (Z/nachher) am besten ab.
Nach der Entschlüsselung der Daten liegt die Vermutung nahe, dass die Handhabung der Prüfkörper in Gruppenclustern bei der Prüfkörperherstellung zu den oben genannten Abweichungen in der Gruppe G6 (D/nachher) geführt haben könnte. Es war uns nicht möglich den ausschlaggebenden alleinigen Faktor zu finden. Die Probability Plots (Bild 3) und die Histogramme (Bild 4) lassen eher auf eine bimodale anstatt auf eine unimodale Verteilung der Daten schließen, was darauf hinweist, dass dieses Cluster andere Eigenschaften aufweist. Interessanterweise hatte dieses Cluster eine höhere sd (niedrigere m) Schätzung, was diese Vorgehensweise für zukünftige Studien überdenken lässt. Folgende Unsicherheiten können die Herstellung der Prüfkörper der Hauptstudie [41] beeinflusst haben: das Material (Hersteller, Lot Nummer), Abweichungen der Sintertemperatur (nicht gleichmäßige Temperatur innerhalb des Ofens, zu frühes Öffnen der Ofentür), Messfehler (Präzision der Mikrometerschraube), Positionierung der Prüfkörper in der Testmaschine, Messfehler an der Testmaschine. Aufgrund der Kapazität des Sinterofens und der Poliermaschine, war es nicht möglich alle Prüfkörper in einem Durchgang zu behandeln, deshalb wurden die Prüfkörper oft in Clustern gehandhabt. Wenn eine Ungenauigkeit aufgetreten ist, so wären wahrscheinlich alle Prüfkörper eines Clusters davon betroffen.

Die unterschiedlichen Schätzmethoden des globalen Weibulltests führten zu voneinander abweichenden Entscheidungen. Die Schätzmethode der LS wies auf einen Unterschied der Moduli zwischen den sechs getesteten Gruppen hin, wohingegen ML keinen Beweis für Unterschiede der Gruppen fand. Nach Nelson [44] impliziert eine Stichprobengröße von 40 in jeder getesteten Gruppe, dass wenn der wahre Wert des Moduls 10 wäre, die Schätzwerte des Moduls mit einer Wahrscheinlichkeit von 99.5% in dem Intervall von (6.2, 16.0) gefunden werden können. Es lässt sich aus Tabelle 3 feststellen, dass die Schätzungen des Moduls m genau in diesem Bereich gefunden werden können. Dies lässt vermuten, dass keine wahren Unterschiede der Weibullmoduli der sechs getesteten Gruppen vorliegen, wie es auch die Schätzmethode Weibull/ML vorschlägt. Eine weitere Möglichkeit die LS/ML Diskrepanz aufzuklären, besteht in der Verwendung eines härteren Signifikanzniveaus von 0.005. Auch Johnson [62] fordert eine Anpassung des geläufigen, aber unangemessen hohen Signifikanzniveaus von 0.05 auf 0.005. Johnson's Vorschlag zielt auf die Reproduzierbarkeit von in der Literatur auffindbaren Ergebnissen ab. Die Anwendung eines niedrigeren $\alpha = 0.005$ Niveaus würde zu einer Übereinstimmung der globalen Schlussfolgerungen der LS/ML Schätzmethoden für die Weibullverteilungsannahme führen. Es gäbe keinen Beweis für Unterschiede der Moduli zwischen den sechs Gruppen und die charakteristische Festigkeit könnte wie folgt angeordnet werden G1 < (G3, G5) < (G2, G4) < G6. Für die Normalverteilungsannahme würde die Rangordnung der Mittelwerte der Rangordnung der charakteristischen Weibullfestigkeit entsprechen. Jedoch wäre die Standardabweichung in der Gruppe G6 größer als in den anderen getesteten Gruppen. Unter dieser Berücksichtigung würden die Weibull- und Normalverteilungsergebnisse sich immer noch unterscheiden.
Vielleicht ist es eine falsche Wahrnehmung, dass die Ergebnisse einer statistischen Analyse eindeutig sein müssen. Deshalb führten wir zwei unabhängige und verblindete statistische Analysen durch, wobei die Statistiker keine Kenntnis über die Zuordnung der Faktorenebenen und die Bedeutung des Hauptergebnisses hatten. Dabei galt nicht etwa das Ziel identische Ergebnisse von beiden Statistikern zu erhalten, sondern die zugrunde liegende, wahre Darstellung der Daten präziser beschreiben zu können. Deshalb wurden auch keine starren Richtlinien für die statistische Analyse im Vorfeld aufgestellt. Es wurde auf einen detaillierten Analyseplan, einschließlich starr und eindeutig definierter Ziele, und jegliche unflexible Spezifikationen der statistischen Analysemethoden verzichtet.
Gøtzsche [50] nimmt eine Datenanalyse als einen sehr subjektiven Prozess war, der anfällig für Voreingenommenheit ist. Er schlägt vor, dass zwei fertige statistische Manuskripte verfasst werden sollten, bevor eine Entschlüsselung der verblindeten Daten erfolgt. Hier wird

ein singulärer Endreport präsentiert, der die identischen Ergebnisse enthält und mögliche Unterschiede einzeln diskutiert. Dies soll Gemeinsamkeiten und Diskrepanzen besser darstellen. Es war das erste Mal für Statistiker C und D, dass eine „unabhängige statistische Analyse“ und eine „parallele Manuskripterstellung“ auf ein „Verblinden des Datensatzes“ angewendet wurde. Vor allem das „Verblinden des Datensatzes“ war eine ungewohnte Technik für den Betreuer A und die Statistiker C und D. Die übliche Herangehensweise an die Datenanalyse besteht darin, dass der Statistiker vor der Analyse umfassende Kenntnis über die präsentierten Messwerte und Grundlagen des durchgeführten Experiments hat, um die bestmögliche Analyse für ein Projekt zu gewährleisten. Das „Verblinden des Datensatzes“ brachte Statistiker C und D in eine ungewohnte Situation und führte möglicherweise zu einer objektiveren Analyse.

Die „unabhängige statistische Analyse“ hatte einen Einfluss auf beide Statistiker. Das Wissen über eine anschließende Diskussion und die Beurteilung der erzielten Ergebnisse von einem unabhängigen Statistiker, führte zu einer sehr genauen und aufmerksamen Analyse der Daten mit spontaner Eigenüberprüfung der involvierten Statistiker. Die „unabhängige statistische Analyse“ kann zu einem wirkungsvollen Werkzeug für eine fundierte und umfassende statistische Analyse werden, die eine solitäre Überprüfung der Beweisführung überbietet. Erstaunlicherweise schlugen die Statistiker C und D unterschiedliche Wege sowohl für Visualisierung der Daten als auch für die Schätz- und Testmethoden ein. Sie erreichten jedoch komplementäre Ergebnisse.

Die „parallele Manuskripterstellung“ war ebenfalls kein Standardprocedere. Sie zwang jedoch beide Statistiker ihre finalen Schlussfolgerungen klar darzulegen, sich für eine explizite Interpretation der Daten zu entscheiden und diese prägnant in Tabellen und Abbildungen zu zeigen. Um in der fünften Phase der Studie (Abbildung 1) ein Zusammenführen beider Manuskripte in einen Endreport zu ermöglichen, mussten an den einzelnen Manuskriptbestandteilen Kürzungen vorgenommen werden. Für eine ähnliche zukünftige Studie wird empfohlen, vorab eine Vereinheitlichung der Bezeichnungen und die Länge der unabhängig angefertigten Abschnitte zu besprechen, um das Zusammenführen der Manuskripte zu vereinfachen.

Eine mögliche Einschränkung einer verblindeten statistischen Analyse ist ihre immanente Ineffizienz [50]. Es nimmt viel Zeit in Anspruch zwei Statistiker mit einer Analyse zu betrauen und zwei unabhängige Manuskripte vorzustellen.

Miller & Stewart [56] betonen, dass die Anforderung an den Statistiker mit einem verschlüsselten Datensatz zu arbeiten, der Studiendurchführung eine größere Komplexität

gibt. Um die benötigte Zeit zur Datenanalyse zu planen, wurden die Empfehlungen von Pocock [51,63] berücksichtigt. Er forderte genug Zeit für die Messdatenanalyse, da eine gute statistische Analyse nicht im Schnellverfahren erreicht werden kann. Bei der Studienplanung sollte eben dieser Zeitfaktor positiv berücksichtig werden. Deshalb wurden in diesem Studiendesign beiden Statistikern drei Wochen Zeit für die Analyse der Daten und das Schreiben der unabhängigen Manuskripte gegeben, was für eine tiefgehende statistische Analyse ausreichend Zeit bedeuten sollte. Tatsächlich brauchte jeder der Statistiker mindestens 25 Stunden Zeit für die Analyse und die Fertigstellung des Entwurfs. Die Bereitstellung von drei Wochen Zeit erscheint viel im Vergleich zu einer normalen, nicht parallel geführten statistischen Analyse. Es erschien jedoch in diesem Fall durchaus gerechtfertigt.

Beide Statistiker schlugen unterschiedliche Wege zu Analyse des Datensatzes ein. Der Statistiker D bevorzugte den Ansatz der Permutationstests für die Weibullstatistik.

Diese innovative Idee wird durch den Ansatz unterstützt, dass Permutationstests nicht parametrisch sind, jede beliebige Testgröße zulassen und unter der Nullhypothesenannahme keine Vermutung über die Verteilung dieser Testgröße anstellen (H_0, keine Unterschiede zwischen den Gruppen). R kann zur Programmierung verwendet werden und die Unabhängigkeit von spezifischer statistischer Software wird ermöglicht [22,45,46]. Ein Defizit dieses Ansatzes war, dass die Permutationstests für Post-hoc-Tests in Tabelle 4c separat programmiert werden mussten und die Anwendung der Bonferroni-Holm Korrektur für die Paartests (Tabelle 4c) notwendig wurde. Interessanterweise kam Statistiker D zu anderen Schlussfolgerungen, hauptsächlich aus dem Fehlen der Anwendung einer globalen Analyse. Einzelne Vergleiche in jeder Faktorebene waren nachteilig, da sie zu Mehrfachvergleichen, schwieriger Interpretation und kompliziertem Design führten. Das Stützen auf eine ANOVA Analyse unter Berücksichtigung beider Faktoren wäre für das spezifische Problem besser geeignet gewesen, jedoch wären hierzu normalverteilte Variablen mit gleichen Varianzen nötig. Es gibt Methoden zur Umwandlung von weibullverteilten zu normalverteilten Variablen und diese können eine Alternative darstellen [64].

Der eingeschlagene Weg der Analyse von Statistiker C umfasste sowohl die Weibull- als auch die Normalverteilung unter Verwendung der grafischen und statistischen Möglichkeiten in Minitab. Verwendet wurden globale Likelihood Ratio Tests zusammen mit integrierten Post-Hoc Konfidenzintervallen. Keine weitere Programmierung war nötig. Letztendlich wurde dieser Ansatz auch für die Analyse eines größeren Datensatzes von Biegefestigkeitsmesswerten verwendet [41].

Tabelle 5 illustriert den finalen Arbeitsablauf für eine lösungsorientierte statistische Analyse von Biegefestigkeitsdaten, er wurde in dieser Studie entwickelt und auch erfolgreich in einer weiteren Studie angewandt [41]. Der Workflow besteht aus sieben Schritten beginnend mit der Visualisierung der Messwerte (Schritt 1), Auswahl von mindestens einer angemessenen Verteilungsannahme (Schritt 2), Schätzung der Parameter einer Verteilung (Schritt 3), Überprüfung der Goodness-of-Fit (Schritt 4) und Schätzung des 95%CI (Schritt 5). Um zu verstehen, welche Gruppen als besser abschneidend als andere bezeichnet werden können, müssen hinsichtlich der Lageparamter und Streuung angemessene statistische Tests verwendet werden (Schritt 6). Diese Tests verlangen die Festlegung auf eine Verteilungsannahme, welche auf alle Testgruppen übertragen werden kann. In der Praxis ist es jedoch fast unmöglich eine einzige, auf alle Gruppen gleich gut anwendbare Verteilungsannahme zu finden. Die entwickelte Arbeitshypothese bestand aus der Ausweitung von Abernethy's [45] vorsichtiger Herangehensweise auf zwei unterschiedliche, vernünftige Verteilungsannahmen (Weibull und Normalverteilung in diesem Fall) (Schritt 2). Als Konsequenz müssen im letzten Schritt die Ergebnisse der verschiedenen Verteilungsannahmen verglichen werden (Schritt 7).

In dieser Studie wurde nicht darauf abgezielt, die ganze Bandbreite von möglichen statistischen Herangehensweisen für die Analyse von Festigkeitsmesswerten abzufragen. Natürlich können nur zwei teilnehmende Statistiker nicht den vollen Bereich von verschiedenen, untersucherabhängigen Analysen aufzeigen. Zum Bespiel würden einige Wissenschaftler die gleichwertige volumenbasierte Betrachtung innerhalb der Weibullanalyse bevorzugen [12,31]. Andere würden eine Log-Normal Verteilungsannahme berücksichtigen, argumentierend, dass eine Log-Normal Verteilung eine gute Alternative für die Stabilisierung und Reduzierung der Streuung ist und zudem eine sinnvolle Linearisierung der Probability Plots ermöglicht. Eine normale ANOVA, angewandt auf Log-Normal Daten, wäre solide genug, um offene Fragen schnell und unkompliziert zu beantworten.

Gegenüber den oben genannten Unzulänglichkeiten stellte der Ansatz der Durchführung einer verblindeten Analyse eine praktikable statistische Vorgehensweise für die Analyse des finalen Datensatzes [41] dar. Er ergänzt die Empfehlungen von Hannigan und Lynch [65] und sollte definitiv berücksichtigt werden.

4. Zusammenfassung und Ausblick

In Kombination beider durchgeführter Studien konnten weiterführende Erkenntnisse in dem Verhalten von Zirkonoxidmaterialien der zweiten Generation in unterschiedlichen Biegefestigkeitsversuchen und unterschiedlicher Herstellungsweise der Prüfkörper, sowie die Plausibilität möglicher Interpretationsansätze von statistischen Daten ermittelt werden. Der Biaxiale-Biegeversuch zeigte die höchsten Biegefestigkeitsdaten, der Vier-Punkt-Biegeversuch die niedrigsten. Die Oberflächenbearbeitung beinflusste die Biegefestigkeitsdaten, wohingegen der Weibullmodul nicht relevant beeinflusst wurde. Die Bearbeitung der Zirkonoxidprüfkörpern nach dem Sintern führte zu einer Alteration der Oberflächenbeschaffenheit, jedoch war kein erhöhter monokliner Phasenanteil nachweisbar. Die dem Material inhärente Eigenschaft zur Phasenumwandlung besteht immer, deshalb sollten vorliegende Biegefestigkeitswerte mit weiteren Untersuchungsmethoden wie Röntgendiffraktographie, Profilometrie oder Rasterelektronenmikroskopie ergänzt werden, um einen Hinweis auf mögliche Diskrepanzen zu erhalten. Die Biegefestigkeitsdaten der Zirkonoxidmaterialien bewegten sich innerhalb derselben Grenzen und kein singulärer Effekt eines Zirkonoxidmaterials konnte beobachtet werden. Diese Untersuchung enthält einen Überblick über das Verhalten im Biegefestigkeitsversuch von Zirkonoxidmaterialien der zweiten Generation. Derzeit wird bereits an Werkstoffen der dritten Generation gearbeitet. Es erscheint als interessantes Ziel Materialien der dritten Generation mit denen der zweiten Generation zu vergleichen. Was die statistische Interpretation der Daten betrifft, liefert der Ansatz des „verblindeten Datensatzes“, der „unabhängigen statistischen Anlayse“ und des „parallelen Manuskriptschreibens“ von zwei unabhängigen Statistikern eine wertvolle Bereicherung der statistischen Analyse. Wie man aus der vorliegenden Untersuchung folgern kann, unterschieden sich beide Herangehensweisen der Statistiker voneinander. Die ausschlaggebende Differenz war der Unterschied von verschiedenen Verteilungsannahmen (Normal- oder Weibullverteilungsannahme) und die Anwendung anderer Schätzmethoden (LS/ML). In der Conclusio des parallelen Ansatzes der statistischen Analyse wird ein bereits erprobter Arbeitsablauf für die Analysierung von Biegefestigkeitsdaten weiterer Studien vorgeschlagen.

5. Literaturverzeichnis

1. Christel, P.; Meunier, A.; Heller, M.; Torre, J.P.; Peille, C.N. Mechanical properties and short-term in vivo evaluation of yttrium-oxide-partially-stabilized zirconia. J. Biomed. Mater. Res. Part A 1989, 23, 45–61.
2. Garvie, R.C.; Hannink, R.H.; Pascoe, R.T. Ceramic steel? Nature 1975, 258, 703–704.
3. Pecho, O.E.; Ghinea, R.; Ionescu, A.M.; Cardona, J.D.L.C.; Paravina, R.D.; Pérez, M.D.M. Color and translucency of zirconia ceramics, human dentine and bovine dentine. J. Dent. 2012, 40, e34–e40.
4. Aboushelib, M.N.; de Jager, N.; Kleverlaan, C.J.; Feilzer, A.J. Microtensile bond strength of different components of core veneered all-ceramic restorations. Dent. Mater. 2005, 21, 984–991.
5. Conrad, H.J.; Seong, W.-J.; Pesun, I.J. Current ceramic materials and systems with clinical recommendations: A systematic review. J. Prosthet. Dent. 2007, 98, 389–404.
6. Al-Amleh, B.; Lyons, K.; Swain, M. Clinical trials in zirconia: A systematic review. J. Oral Rehabil. 2010, 37, 641–652.
7. Sax, C.; Hammerle, C.H.; Sailer, I. 10-year clinical outcomes of fixed dental prostheses with zirconia frameworks. Int. J. Comput. Dent. 2011, 14, 183–202.
8. Johansson, C.; Kmet, G.; Rivera, J.; Larsson, C.; Vult Von Steyern, P. Fracture strength of monolithic all-ceramic crowns made of high translucent yttrium oxide-stabilized zirconium dioxide compared to porcelain-veneered crowns and lithium disilicate crowns. Acta Odontol. Scand. 2013, 72, 145–153.
9. Beuer, F.; Stimmelmayr, M.; Gueth, J.-F.; Edelhoff, D.; Naumann, M. In vitro performance of full-contour zirconia single crowns. Dent. Mater. 2012, 28, 449–456.
10. Quinn, G.D.; Morrell, R. Design data for engineering ceramics: A review of the flexure test. J. Am. Ceram. Soc. 1991, 74, 2037–2066.
11. Shetty, D.K.; Rosenfield, A.R.; Duckworth, W.H.; Held, P.R. A biaxial-flexure test for evaluating ceramic strengths. J. Am. Ceram. Soc. 1983, 66, 36–42.
12. Quinn, J.B.; Quinn, G.D. A practical and systematic review of Weibull statistics for reporting strengths of dental materials. Dent. Mater. 2010, 26, 135–147.
13. Bhamra, G.; Palin,W.M.; Fleming, G.J.P. The effect of surface roughness on the flexure strength of an alumina reinforced all-ceramic crown material. J. Dent. 2002, 30, 153–160.
14. Curtis, A.R.;Wright, A.J.; Fleming, G.J.P. The influence of surface modification techniques on the performance of a Y-TZP dental ceramic. J. Dent. 2006, 34, 195–206.

15. de Castro, H.L.; Corazza, P.H.; Paes-Junior, T.D.A.; Della Bona, A. Influence of Y-TZP ceramic treatment and different resin cements on bond strength to dentin. Dent. Mater. 2012, 28, 1191–1197.
16. Della Bona, A.; Donassollo, T.A.; Demarco, F.F.; Barrett, A.A.; Mecholsky, J.J., Jr. Characterization and surface treatment effects on topography of a glass-infiltrated alumina/zirconia-reinforced ceramic. Dent. Mater. 2007, 23, 769–775.
17. Guazzato, M.; Albakry, M.; Quach, L.; Swain, M.V. Influence of surface and heat treatments on the flexural strength of a glass-infiltrated alumina/zirconia-reinforced dental ceramic. Dent. Mater. 2005, 21, 454–463.
18. Kosmač, T.; Oblak, C.; Jevnikar, P.; Funduk, N.; Marion, L. The effect of surface grinding and sandblasting on flexural strength and reliability of Y-TZP zirconia ceramic. Dent. Mater. 1999, 15, 426–433.
19. Stawarczyk, B.; Frevert, K.; Ender, A.; Roos, M.; Sener, B.;Wimmer, T. Comparison of four monolithic zirconia materials with conventional ones: Contrast ratio, grain size, four-point flexural strength and two-body wear. J. Mech. Behav. Biomed. Mater. 2016, 59, 128–138.
20. Dentistry—CeramicMaterials; ISO 6872; International Organization for Standardization: Geneva, Switzerland, 2008.
21. Toraya, H.; Yoshimura, M.; Somiya, S. Calibration curve for quantitative analysis of the monoclinic-tetragonal ZrO2 system by X-ray diffraction. J. Am. Ceram. Soc. 1984, 67.
22. Roos, M.; Stawarczyk, B. Evaluation of bond strength of resin cements using different general-purpose statistical software packages for two-parameter Weibull statistics. Dent. Mater. 2012, 28, e76–e88.
23. Stawarczyk, B.; Özcan, M.; Hämmerle, C.H.F.; Roos, M. The fracture load and failure types of veneered anterior zirconia crowns: An analysis of normal and Weibull distribution of complete and censored data. Dent. Mater. 2012, 28, 478–487.
24. Bütikofer, L.; Stawarczyk, B.; Roos, M. Two regression methods for estimation of a two-parameter Weibull distribution for reliability of dental materials. Dent. Mater. 2015, 31, e33–e50.
25. Stawarczyk, B.; Özcan, M.; Trottmann, A.; Hämmerle, C.H.F.; Roos, M. Evaluation of flexural strength of hipped and presintered zirconia using different estimation methods of Weibull statistics. J. Mech. Behav. Biomed. Mater. 2012, 10, 227–234.
26. Fischer, J.; Stawarczyk, B.; Hämmerle, C.H.F. Flexural strength of veneering ceramics for zirconia. J. Dent. 2008, 36, 316–321.

27. Chai, J.; Chu, F.C.; Chow, T.W.; Liang, B.M. Chemical solubility and flexural strength of zirconia-based ceramics. Int. J. Prosthodont. 2007, 20, 587–595.
28. Anusavice, K.J.; Kakar, K.; Ferree, N. Which mechanical and physical testing methods are relevant for predicting the clinical performance of ceramic-based dental prostheses? Clin. Oral Implants Res. 2007, 18, 218–231.
29. Ban, S.; Anusavice, K.J. Influence of test method on failure stress of brittle dental materials. J. Dent. Res. 1990, 69, 1791–1799.
30. Jin, J.; Takahashi, H.; Iwasaki, N. Effect of test method on flexural strength of recent dental ceramics. Dent. Mater. J. 2004, 23, 490–496.
31. Danzer, R.; Lube, T.; Supancic, P.; Damani, R. Fracture of ceramics. Adv. Eng. Mater. 2008, 10, 275–298.
32. Papanagiotou, H.P.; Morgano, S.M.; Giordano, R.A.; Pober, R. In vitro evaluation of low-temperature aging effects and finishing procedures on the flexural strength and structural stability of Y-TZP dental ceramics. J. Prosthet. Dent. 2006, 96, 154–164.
33. Kosmač, T.; Oblak, C.; Jevnikar, P.; Funduk, N.; Marion, L. Strength and reliability of surface treated Y-TZP dental ceramics. J. Biomed. Mater. Res. 2000, 53, 304–313.
34. Luthardt, R.G.; Holzhüter, M.; Sandkuhl, O.; Herold, V.; Schnapp, J.D.; Kuhlisch, E.;Walter, M. Reliability and properties of ground Y-TZP-zirconia ceramics. J. Dent. Res. 2002, 81, 487–491.
35. Denry, I.L.; Holloway, J.A. Microstructural and crystallographic surface changes after grinding zirconia-based dental ceramics. J. Biomed. Mater. Res. Part B Appl. Biomater. 2006, 76, 440–448.
36. Qeblawi, D.M.; Muñoz, C.A.; Brewer, J.D.; Monaco, E.A., Jr. The effect of zirconia surface treatment on flexural strength and shear bond strength to a resin cement. J. Prosthet. Dent. 2010, 103, 210–220.
37. Karakoca, S.; Yilmaz, H. Influence of surface treatments on surface roughness, phase transformation, and biaxial flexural strength of Y-TZP ceramics. J. Biomed. Mater. Res. Part B Appl. Biomater. 2009, 91, 930–937.
38. Gupta, T.K. Strengthening by surface damage in metastable tetragonal zirconia. J. Am. Ceram. Soc. 1980, 63, 117–117.
39. Hannink, R.H.J.; Kelly, P.M.; Muddle, B.C. Transformation toughening in zirconia-containing ceramics. J. Am. Ceram. Soc. 2000, 83, 461–487.

40. Stawarczyk, B.; Özcan, M.; Hallmann, L.; Ender, A.; Mehl, A.; Hämmerle, C.H.F. The effect of zirconia sintering temperature on flexural strength, grain size, and contrast ratio. Clin. Oral Investig. 2012, 17, 269–274.
41. Schatz, C.; Strickstrock, M.; Roos, M.; Edelhoff, D.; Eichberger, M.; Zylla, I.M.; Stawarczyk, B. Influence of specimen preparation and test methods on the flexural strength results of monolithic zirconia materials. Materials 2016, 9, 180.
42. McCabe, J.F.; Carrick, T.E. A statistical approach to the mechanical testing of dental materials. Dent. Mater. 1986, 2, 139–142.
43. Nohut, S. Influence of sample size on strength distribution of advanced ceramics. Ceram. Int. 2014, 40, 4285–4295.
44. Nelson, W. Applied Life Data Analysis; John Wiley & Sons, Inc.: New York, NY, USA, 1982.
45. Abernethy, R. The New Weibull Handbook, 5th ed.; Robert, B.A., Ed.; Robert. Abernethy: North Palm Beach, FL, USA, 2009.
46. Rinne, H. The Weibull Distribution: A Handbook; Chapman & Hall/CRC Press: Giessen, Germany, 2009.
47. Greenland, S. Bayesian perspectives for epidemiological research: I. Foundations and basic methods. Int. J. Epid. 2006, 35, 765–775.
48. Pocock, S.J.; Hughes, M.D.; Lee, R.J. Statistical problems in the reporting of clinical trials. N. Eng. J. Med. 1987, 317, 426–432.
49. Polit, D.F. Blinding during the analysis of research data. Int. J. Nursing Stud. 2011, 48, 636–641.
50. Gøtzsche, P.C. Blinding during data analysis and writing of manuscripts. Contr. Clin. Trials 1996, 17, 285–293.
51. Pocock, S.J. Clinical Trials: A Practical Approach; John Wiley & Sons: New York, NY, USA, 1983.
52. Doll, R. Controlled trials: The 1948 watershed. Br. Med. J. 1998, 317, 1217–1220.
53. Schulz, K.F.; Chalmers, I.; Altman, D.G. The landscape and lexicon of blinding in randomized trials. Ann. Internal. Med. 2002, 136, 254–259.
54. Schulz, K.F.; Altman, D.G.; Moher, D. CONSORT 2010 Statement: Updated guidelines for reporting parallel group randomization trials. Ann. Internal. Med. 2010, 152, 726–732.

55. Moher, D.; Hopewell, S.; Schulz, K.F.; Montori, V.; Gøtzsche, P.C.; Devereaux, P.J.; Elbourne, D.; Egger, M.; Altman, D. CONSORT 2010 Explanation and Elaboration: Updated guidelines for parallel group randomized trials. J. Clin. Epidemiol. 2010, 340, c869.
56. Miller, L.E.; Stewart, M.E. The blind leading the blind: Use and misuse of blinding randomized controlled trials. Contemp. Clin. Trials 2011, 32, 240–243.
57. Schulz, K.F.; Grimes, D.A. Blinding in randomised trials: Hiding who got what. Lancet 2002, 359, 696–700.
58. Klein, J.R.; Roodman, A. Blind analysis in nuclear and particle physics. Annu. Rev. Nucl. Part. Sci. 2005, 55, 141–163.
59. R Foundation for Statistical Computing. A Language and Environment for Statistical Computing; R Foundation for Statistical Computing: Vienna, Austria, 2010. Available online: http://www.R-project.org (accessed on 22 June 2016).
60. Holm, S. A simple sequentially rejective multiple test procedure. Scand. J. Stat. 1979, 6, 65–70.
61. Ryan, B.F.; Joiner, B.L. Minitab 14 Statistical Software; Minitab, Inc.: State College, PA, USA; pp. 1972–2005.
62. Johnson, V.E. Revised standards for statistical evidence. Proc. Natl. Acad. Sci. USA 2013, 48, 19313–19317.
63. Pocock, S.J. Clinical trials with multiple outcomes: A statistical perspective on their design, analysis and interpretation. Control. Clin. Trials 1997, 18, 530–545.
64. Watthanacheewakul, L. Analysis of Variance with Weibull Data. In Proceedings of the International MultiConference of Engineers and Computer Scientists, Hong Kong, China, 16–18 March 2011; Volume 3, pp. 2051–2056.
65. Hannigan, A.; Lynch, C.D. Statistical methodology in oral and dental research: Pitfalls and recommendations. J. Dent. 2013, 41, 385-392.

Danksagung

Ich möchte mich von Herzen bei den Menschen bedanken, die den Erstellungsprozess dieser Arbeit fortwährend unterstützt und ermöglicht haben. Folgenden Personen gebührt mein aufrichtiger Dank:

PD Dr. Bogna Stawarczyk für die Ermöglichung der Durchführung der Versuche, ihr fortlaufendes Engagement, Motivation und geduldige Hilfestellung bei der Erstellung des Manuskripts.

PD Dr. Malgorzata Roos, für die immer freundliche und kompetente statistische Unterstützung, sowie die Realisierung der verblindeten statistischen Analyse.

Frau Marlis Eichberger für die herzliche Hilfestellung bei der Herstellung der Prüfkörper.

Frau Monika Strickstrock für die Unterstützung bei den Oberflächenmessungen der Prüfkörper.

Lebenslauf

(...)