Journal of Biomechanics 83 (2019) 150–156

Contents lists available at ScienceDirect

Journal of Biomechanics

journal homepage: www.elsevier.com/locate/jbiomech
www.JBiomech.com

Paired versus two-group experimental design for rheological studies of vocal fold tissues

Chet C. Xu [a,b,*,1], Dateng Li [c], Ted Mau [a], Elhum McPherson [a], Mindy Du [a], Song Zhang [d,1]

[a] Department of Otolaryngology-Head and Neck Surgery, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA
[b] Graduate Program in Biomedical Engineering, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA
[c] Department of Statistical Science, Southern Methodist University, Dallas, TX 75275, USA
[d] Department of Clinical Sciences, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA

ARTICLE INFO

Article history:
Accepted 23 November 2018

Keywords:
Mixed-effects model
Sample size
Soft tissue
Statistical efficiency
Viscoelastic shear properties

ABSTRACT

Vibratory function of the vocal folds is largely determined by the rheological properties or viscoelastic shear properties of the vocal fold lamina propria. To date, investigation of the sample size estimation and statistical experimental design for vocal fold rheological studies is nonexistent. The current work provides the closed-form sample size formulas for two major study designs (i.e. paired and two-group designs) in vocal fold research. Our results demonstrated that the paired design could greatly increase the statistical power compared to the two-group design. By comparing the variance of estimated treatment effect, this study also confirms that ignoring within-subject and within-vocal fold correlations during rheological data analysis will likely increase type I errors. Finally, viscoelastic shear properties of intact and scarred rabbit vocal fold lamina propria were measured and used to illustrate theoretical findings in a realistic scenario and project sample size requirement for future studies.

1. Introduction

Located at the narrowest portion of the upper airway, the human vocal folds undergo flow-induced, self-sustained oscillation to generate acoustic output during voice production or phonation (Titze, 2000). There are three main layers in the vocal folds: the epithelium, the connective tissue layer called the lamina propria, and the muscle (Titze, 2000). Human vocal fold mucosa, which consists of the epithelium and the superficial layer of lamina propria, is the major vibratory portion during oscillation. Its viscoelastic shear properties or rheological properties critically determine the vibratory characteristics of the vocal folds since the mucosal wave propagating on the vocal fold surface layer during phonation is a shear wave (Chan and Rodriguez, 2008; Chan and Titze, 1999). Pathological conditions such as vocal fold scarring can cause changes in tissue rheological properties, leading to dysphonia (Hansen and Thibeault, 2006; Hirano, 2005; Woo et al., 1994).

Given their defining role in the acoustics and biomechanics of voice production, knowledge of the vocal fold lamina propria rheological properties is important for the diagnosis and treatment of vocal fold disorders. Furthermore, the rheological studies of vocal fold tissues in pathological states may help in understanding their etiology. For these reasons, various methods have been developed to measure the viscoelasticity of vocal fold tissues under shear deformation (Chan and Rodriguez, 2008; Chan and Titze, 1999; Jiao et al., 2009; Kazemirad et al., 2014). Evaluation of the lamina propria rheological properties is also becoming a standard functional test in animal studies of vocal fold wound healing, scarring treatment, and regenerative medicine (Bartlett et al., 2016; Hansen and Thibeault, 2006; Kutty and Webb, 2009).

There are two major experimental designs for vocal fold animal studies. In a paired design, one vocal fold in each animal receives the experimental treatment, while the other serves as normal or sham control (Cobell et al., 2006; Rousseau et al., 2004; Thibeault et al., 2002; 2011; Xu et al., 2010). Alternatively, in a two-group design, the animals are assigned to either experimental or control group with both vocal folds of each animal receiving the same treatment (Choi et al., 2012; Svensson et al., 2011; Thibeault et al., 2009). In spite of the growing number of studies on rheological characterization of human and animal vocal fold tissues, the determination of sample size and the implication of study design in statistical inference were rarely mentioned, let alone fully investigated.

In this study we present the closed-form sample size formulas for the rheological studies of vocal fold tissues, followed by an

* Corresponding author.
E-mail address: chetcxu@gmail.com (C.C. Xu).
[1] These authors contributed equally to the study.

https://doi.org/10.1016/j.jbiomech.2018.11.040
0021-9290/

analytic evaluation of the statistical properties of the paired and two-group designs. A second objective was to examine the impact of ignoring data correlations on the outcome of rheological data analysis. Finally, a rabbit vocal fold model was used to illustrate theoretical findings in a realistic scenario.

2. Methods

2.1. Statistical derivation based on mixed-effects modeling

2.1.1. The mixed-effects model

Rheological properties of vocal fold tissues are functions of frequency and can be quantified by the complex shear modulus, which includes an elastic (storage) component—the elastic shear modulus G', and a viscous component—the dynamic viscosity η' (Chan and Titze, 1999). In most rheological tests, a vocal fold specimen is measured multiple times, each at a different frequency. Such repeated measurements on the same vocal fold tissue generate intra-vocal fold correlation. Additionally, intra-animal or human subject correlation exists among measurements obtained from the left and right vocal folds of the same subject. To include this hierarchical correlation structure into the vocal fold rheological data analysis, we employ the mixed-effects model approach (Xu et al., 2017; 2018). The mixed-effects model refers to the use of both fixed and random effects in the same analysis (Laird and Ware, 1982; West et al., 2014). The within-subject and within-vocal fold correlations can be accounted for by including subject and vocal fold random effects in the model (Xu et al., 2017; 2018). The mixed-effects models previously developed for the statistical analysis of vocal fold rheological properties are summarized in Table 1.

As shown in Table 1, frequency is always one of the fixed effects in the models because G' and η' are functions of frequency. Our recent study also showed that it was preferable to treat frequency as a categorical variable (Xu et al., 2018). Another fixed effect is the treatment effect, for the assessment of possible significant relationship between tissue rheological properties and a treatment. It has been demonstrated in multiple publications that on a log scale, changes in tissue rheological properties due to scarring or vocal fold injection were largely constant across the test frequency range (Choi et al., 2012; Hirano et al., 2004; Rousseau et al., 2004; Thibeault et al., 2002). Thus, treatment effect was assumed to be independent of frequency, and the interaction term between treatment and frequency was not included in the mixed-effects model of the present study.

In a typical rheological test, m measurements are obtained on each vocal fold during the frequency sweep experiment performed on a total of n animals. Let Y_{ijk} be the continuous measurement outcome obtained at the kth ($k = 1, \cdots, m$) frequency from the jth (j = 1 for left and 2 for right) vocal fold of the ith (i = 1,$\cdots$,n) animal. Y_{ijk} were log-transformed in this study to increase their distributional normality (Xu et al., 2017) and were modeled by a linear mixed-effects model:

$$Y_{ijk} = \alpha_k + \beta_1 r_{ij} + \gamma_i + \xi_{ij} + e_{ijk}, \quad (1)$$

where r_{ij} takes value 1 or 0 indicating that the jth vocal fold of the ith animal receives the experimental or control treatment; α_k is the frequency-specific intercept representing the mean value under control condition at the kth ($k = 1, \cdots, m$) frequency; β_1 represents the treatment effect, which is the mean difference between the treatment and control groups; γ_i and ξ_{ij} are the animal and vocal fold random effects, respectively; e_{ijk} represents the residual error. γ_i, ξ_{ij} and e_{ijk} are assumed to be independent normal variables with zero mean and variances σ_γ^2, σ_ξ^2, and σ_e^2, respectively. The null hypothesis of interest is $H_0 : \beta_1 = 0$ (i.e. there is no treatment effect).

Eq. (1) can accommodate different experimental designs through the specification of r_{ij}. For instance, $r_{i1} = 1$ and $r_{i2} = 0$ indicate a paired design where the ith animal receives the experimental treatment on the left vocal fold while the right vocal fold serves as control. For the two-group design, r_{i1} and r_{i2} always have the same value: $r_{i1} = r_{i2} = 1$ if animal i is randomly assigned to the experimental group, or $r_{i1} = r_{i2} = 0$ if animal i is randomized to the control group.

2.1.2. Closed-form sample size formulas

Eq. (1) suggests that marginally, Y_{ijk} follows a normal distribution with mean $\alpha_k + \beta_1 r_{ij}$ and variance $\tau^2 = \sigma_\gamma^2 + \sigma_\xi^2 + \sigma_e^2$. For measurements obtained from the same vocal fold specimen, the within-vocal fold correlation is $Corr\ (Y_{ijk},\ Y_{ijk'}) = \rho_1 = \frac{\sigma_\gamma^2 + \sigma_\xi^2}{\tau^2}$ for $k \neq k'$ because they share both random effects γ_i and ξ_{ij}. For measurements obtained from the same animal but different vocal folds, the within-animal correlation is $Corr\ (Y_{ijk},\ Y_{ij'k'}) = \rho_2 = \frac{\sigma_\gamma^2}{\tau^2}$ for $j \neq j'$ because they only share the animal random effect γ_i. Please note that $\rho_1 \geq \rho_2 \geq 0$ always holds. Measurements are assumed to be independent among animals.

For the paired design, without loss of generality, we suppose all animals receive experimental treatment on the left vocal fold ($j = 1$), and control on the right vocal fold ($j = 2$). The treatment effect β_1 can be estimated by

$$\widehat{\beta}_1^{(p)} = \frac{\sum_{i=1}^{n}\sum_{k=1}^{m}(Y_{i1k} - Y_{i2k})}{nm}, \quad (2)$$

where the superscript (p) stands for the paired design. $\widehat{\beta}_1^{(p)}$ has a normal distribution with mean β_1 and variance $V^{(p)} = \frac{2[1+(m-1)\rho_1 - m\rho_2]\tau^2}{nm}$ (see Appendix A).

Given the true value of treatment effect $\beta_1 = \beta_{10}$, the sample size required for a paired design to achieve pre-determined power $(1 - \gamma)$ at two-sided significance level α can be calculated by:

$$n^{(p)} = \frac{2\left(z_{1-\alpha/2} + z_{1-\gamma}\right)^2 [1 + (m-1)\rho_1 - m\rho_2]\tau^2}{\beta_{10}^2 m}. \quad (3)$$

Table 1
Mixed-effects models in previous reports for the statistical analysis of vocal fold rheological properties.

Application	Fixed effects	Random effects	Interaction term	Reference
To compare G' between human and animal vocal fold tissues*	Log frequency Species	Vocal fold tissue sample	Species • log frequency	Xu et al., 2017
To compare η' between human and animal vocal fold tissues*	Log frequency Species	Vocal fold tissue sample	No	Xu et al., 2017
To compare G' and η' between paired vocal folds in rabbit	Frequency (categorical) Laterality	Vocal fold tissue sample Animal subject	No	Xu et al., 2018

* Only one vocal fold from each animal or human subject was tested in the study. Therefore, there was no within-subject correlation to be considered in the statistical model.

For the two-group design, let treatment group consist of n_1 animals and control group consist of n_2 animals. Note that the total number of animals in the study is $n = n_1 + n_2$. To simplify notation, we code the animals such that $r_{i1} = r_{i2} = 1$ for $i = 1, \cdots, n_1$ (treatment group), and $r_{i1} = r_{i2} = 0$ for $i = n_1 + 1, \cdots, n$ (control group). Then the treatment effect β_1 can be estimated by

$$\widehat{\beta}_1^{(g)} = \frac{\sum_{i=1}^{n_1}\sum_{j=1}^{2}\sum_{k=1}^{m} Y_{ijk}}{2n_1 m} - \frac{\sum_{i=n_1+1}^{n}\sum_{j=1}^{2}\sum_{k=1}^{m} Y_{ijk}}{2n_2 m}, \tag{4}$$

where superscript (g) stands for the two-group design. $\widehat{\beta}_1^{(g)}$ is normally distributed with mean β_1 and variance $V^{(g)} = \frac{\tau^2+(m-1)\rho_1\tau^2+m\rho_2\tau^2}{2mn_1} + \frac{\tau^2+(m-1)\rho_1\tau^2+m\rho_2\tau^2}{2mn_2}$ (see Appendix B). For a balanced two-group design ($n_1 = n_2 = n/2$), the expression of variance can be simplified to $V^{(g)} = \frac{2[1+(m-1)\rho_1+m\rho_2]\tau^2}{nm}$.

The sample size required for a balanced two-group design to achieve pre-determined power $(1-\gamma)$ at two-sided significance level α can be calculated by:

$$n^{(g)} = \frac{2\left(z_{1-\alpha/2} + z_{1-\gamma}\right)^2[1+(m-1)\rho_1+m\rho_2]\tau^2}{\beta_{10}^2 m}. \tag{5}$$

The sample size formula for unbalanced two-group design (i.e. $n_1 \neq n_2$) is presented in Appendix C. The present study focuses on the comparison between the balanced two-group design and the paired design. The conclusion illustrates the general difference in statistical properties between the two designs, but the simpler formula for the balanced two-group design simplifies notation markedly. In the rest of this report, the term "two-group design" refers to balanced two-group design unless otherwise stated.

2.1.3. Comparison of statistical efficiency between paired and two-group designs

A study design that is able to estimate the treatment effect with a smaller variance (i.e. greater precision) for a given sample size is considered to be statistically more efficient. Hence, $1/V^{(p)}$ and $1/V^{(g)}$ represent the efficiency of the paired design and two-group design, respectively. With all other experimental parameters being the same, relative efficiency (R) of the paired design to two-group design is the ratio of $1/V^{(p)}$ to $1/V^{(g)}$:

$$R = \frac{1/V^{(p)}}{1/V^{(g)}} = \frac{1+(m-1)\rho_1+m\rho_2}{1+(m-1)\rho_1-m\rho_2} \geq 1. \tag{6}$$

2.1.4. Impact of ignoring intra-subject and intra-vocal fold correlations

A literature search revealed that many previous studies failed to consider correlations at subject and vocal fold levels during statistical analysis and therefore, mistakenly treated the rheological data as if they were completely independent. Such practice effectively sets $\rho_1 = \rho_2 = 0$ regardless of the experimental design. In this case, the variance of estimated treatment effect becomes $V^{(i)} = \frac{2\tau^2}{nm}$, where the superscript (i) stands for "independent". The ratio of $V^{(i)}$ to $V^{(p)}$is

$$\frac{V^{(i)}}{V^{(p)}} = \frac{2\tau^2}{2[1+(m-1)\rho_1-m\rho_2]\tau^2} = \frac{1}{1+(m-1)\rho_1-m\rho_2}. \tag{7}$$

On the other hand, the ratio of $V^{(i)}$ to $V^{(g)}$is

$$\frac{V^{(i)}}{V^{(g)}} = \frac{2\tau^2}{2[1+(m-1)\rho_1+m\rho_2]\tau^2} = \frac{1}{1+(m-1)\rho_1+m\rho_2}. \tag{8}$$

2.2. Application of theoretical findings to a rabbit vocal fold model

A rabbit vocal fold model was used to illustrate the theoretical results in a realistic scenario because rabbits are one of the most common animal models in laryngeal research (Bless and Welham, 2010). Viscoelastic shear properties of 6 pairs of intact and 4 pairs of scarred rabbit vocal fold lamina propria were measured with a custom-built rheometer. The empirical data obtained from these rabbits allowed us to estimate the design parameters for future studies in similar vocal fold models.

Scarring is a major cause of dysphonia following vocal fold injury and remains one of the most challenging voice problems (Hirano, 2005; Woo et al., 1994). By comparing the rheological data between the scarred and normal vocal folds, the levels of changes in viscoelastic shear properties induced by scarring could be determined.

All ten rabbits involved in the study were of the same gender, breed, age range (22 ± 2.6 weeks), weight range (3.3 ± 0.5 Kg), and diet. Hence, they were assumed to be from the same population. The animals were randomized to the experimental or control group (i.e. a two-group design). Similar size and age female New Zealand White rabbits have been used in many vocal fold studies (Miri et al., 2015; Svensson et al., 2011).

The experimental protocol was approved by the Institutional Animal Care and Use Committee of the University of Texas Southwestern Medical Center, in accordance with the U.S. Public Health Service Policy on Humane Care and Use of Laboratory Animals, the National Institutes of Health Guide for the Care and Use of Laboratory Animals (NIH Publication 85–23, revised 1985), and the Animal Welfare Act (7 U.S.C. et seq.).

2.2.1. Normal vocal fold lamina propria specimens

Normal vocal fold samples were harvested immediately postmortem from six female New Zealand White rabbits. None of these rabbits were sacrificed specifically for this study and none of them received manipulation on the larynges or intubation. Vocal fold specimens consisting of the epithelium and the lamina propria were carefully dissected from the larynges using phonomicrosurgical instruments as previously described (Xu et al., 2017). All specimens were inspected by an otolaryngologist and were free of visible pathology.

2.2.2. Scarred vocal fold lamina propria specimens

Four female New Zealand White rabbits were included in the laser induced scarring experiment. Details of vocal fold scar creation were described previously (Mau et al., 2014). Four weeks after laser injury, the animals were sacrificed and their vocal folds harvested.

2.2.3. Rheometric instrumentation and measurements

A custom-built, linear, simple-shear rheometer system (ElectroForce Systems Group, Eden Prairie, MN) (Fig. 1A) was used for the rheometric measurements as previously described (Chan and Rodriguez, 2008). A custom system was required because there was no commercial rheometer that could perform linear rheometry at frequencies matching the typical fundamental frequency range of human phonation (100–300 Hz). The system was calibrated with a standard ANSI S2.21 polyurethane material with known viscoelastic properties (Chan and Rodriguez, 2008). The system was limited by the signal to noise level, which was determined to be acceptable up to around 250 Hz for vocal fold measurements (Chan and Rodriguez, 2008).

The vocal fold specimens were kept hydrated in phosphate buffered saline solution at room temperature until rheometric measurements were performed, which were conducted within 6 h postmortem. The specimen was placed between two parallel, rectangular acrylic plates and subjected to translational simple shear

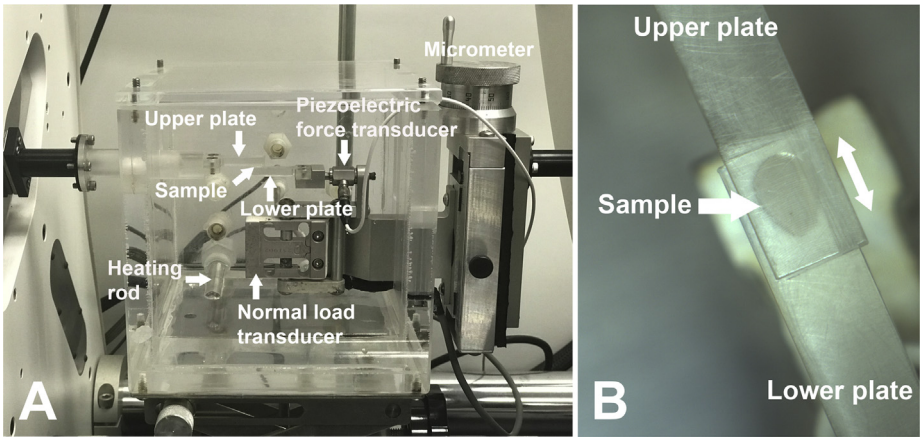

Fig. 1. (A) Picture of the controlled-strain, simple-shear rheometer system. Based on the ElectroForce 3200 mechanical test platform, this custom-built rheometer is capable of valid and reliable measurements of the linear viscoelastic properties at frequencies up to around 250 Hz. (B) A close-up of a rabbit vocal fold sample mounted between the two parallel acrylic plates. The double-headed arrow indicates the directions of sinusoidal translational shear. A schematic of this experimental setup can be found in Chan and Rodriguez, 2008.

(Fig. 1B). The upper tissue plate was attached to the shaft of a linear motor through an actuator, applying a translational displacement to the specimen at a prescribed frequency and magnitude with displacement feedback control. A linear variable differential transformer was used as the displacement transducer to estimate the shear strain of the sample. The resulting shear force at the lower plate was measured by a piezoelectric quartz force transducer (PCB Piezotronics, Depew, NY). Data were processed by a custom MATLAB program to calculate viscoelastic shear properties of the specimens. The water at the base of the acrylic environmental chamber was heated by the heating rod to maintain the ambient air temperature at around 37 °C with ~100% humidity.

3. Results

3.1. Viscoelastic shear properties of normal and scarred rabbit vocal folds

Viscoelastic shear properties of the rabbit vocal fold lamina propria were measured under a small strain (2% or 3%) to ensure stress-strain linearity (Xu et al., 2017; 2018). For each tissue sample, measurements were carried out at 15 test frequencies (i.e. $m = 15$): 1, 5, 10, 25, 50, 75, 100, 125, 130, 150, 175, 190, 200, 225, and 250 Hz. Elastic shear modulus (G') and dynamic viscosity (η') are plotted in Figs. 2 and 3, respectively. The rheologic properties of intact rabbit vocal folds were in good agreement with those of previous studies (Hertegard et al., 2003; Thibeault et al., 2002).

3.2. Comparison of rheological properties between normal and scarred vocal folds

As shown in Figs. 2 and 3, variation in G' and η' between the scarred and intact rabbit vocal folds were not perfectly consistent across the test frequency range. In order to determine whether the interaction term between treatment effect and frequency was required for the analysis of the rabbit data, a variable selection procedure (Xu et al., 2017) was conducted. We found that the treatment-frequency interaction effects were not statistically significant for either log G' ($p = 0.08$) or log η' ($p = 0.61$). Thus, Eq. (1) was suitable for the analysis of our rabbit vocal fold data.

Significant differences in G' ($p < 0.0001$) and η' ($p < 0.0001$) were found between the scarred and normal vocal folds, with scarred vocal folds having an average increase of nearly 200%

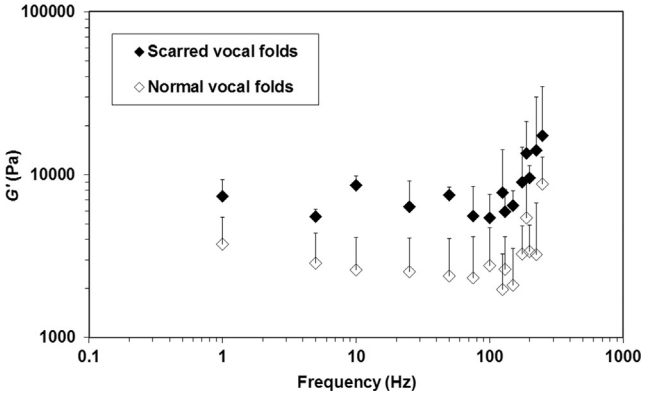

Fig. 2. Elastic shear modulus G' of normal rabbit vocal folds lamina propria (open symbols, n = 6 animals) and scarred vocal fold lamina propria (solid symbols, n = 4 animals) as a function of frequency. Only upper error bars representing standard deviations are shown for visual clarity. Scarred tissue samples show significantly higher values for G' compared with control tissues, indicating increased stiffness in shear deformation.

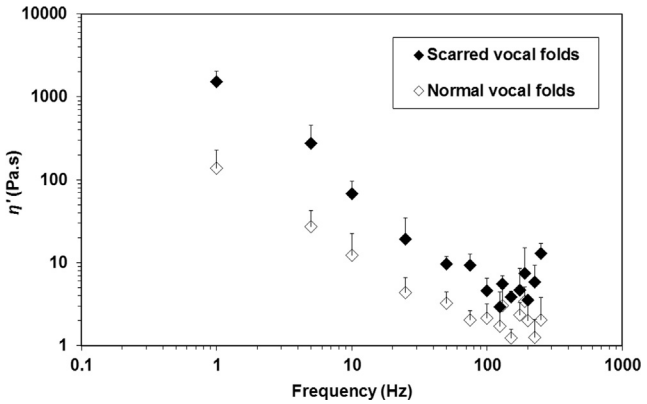

Fig. 3. Dynamic viscosity η' of normal rabbit vocal fold lamina propria (open symbols, n = 6 animals) and scarred vocal fold lamina propria (solid symbols, n = 4 animals) as a function of frequency. Only upper error bars representing standard deviations are shown for visual clarity. Scarred tissue samples show significantly higher values for η' compared with control tissues, indicating increased resistance to shear flow during oscillatory shear deformation.

and 350% in G' and η', respectively. The magnitudes of increase in G' and η' appeared to be similar to those of 2 months (Thibeault et al., 2002) and 6 months (Hirano et al., 2004; Rousseau et al., 2004) postinjury.

3.3. Sample size requirement for detecting different levels of treatment effects in rabbit

The parameters of covariance (τ^2, ρ_1, and ρ_2) of the rabbit vocal fold lamina propria were calculated for both log G' and log η' using the mixed-effects model (Table 2).

Using the data presented in Table 2, sample sizes for future studies in a similar rabbit vocal fold model can be calculated by Eqs. (3) and (5). Table 3 presents the sample size estimates to detect different treatment effects on G' and η' for both experimental designs. The number of measurements per vocal fold was set at 15. It is obvious that a larger treatment effect (β_1) is always associated with a smaller sample size requirement. According to Table 3, large increases in both G' and η' caused by scarring ($\geq$200%) can be detected using a minimum of 3 and 6 rabbits for paired and two-group designs, respectively. This confirmed that the number of animals in our rabbit study (4 experimental, 6 control) was sufficient for identifying the differences in rheological properties between the scarred and normal vocal folds.

3.4. Comparison of statistical efficiency between paired and two-group designs

As shown in Eq. (6), R is always larger or equal to 1 because ρ_2 is non-negative by definition. The paired design is therefore more efficient than the two-group design under the same experimental conditions. Eq. (6) also indicates that a stronger within-animal correlation (ρ_2) will result in a greater relative efficiency of the paired design to the two-group design while a stronger within-vocal fold correlation (ρ_1) is in turn associated with a smaller relative efficiency. R also depends on the number of measurements (m), as shown in Table 4.

By comparing Eqs. (3), (5), and (6), we can conclude that $R = n^{(g)}/n^{(p)}$, which indicates $n^{(g)} \geq n^{(p)}$ because $R \geq 1$. As shown in Table 4, when $m = 15$, a two-group design requires 66% and 35% more rabbits to obtain the same statistical power compared to a paired design in order to detect changes in G' and η', respectively. The theoretical efficiency gains of the paired design may translate slightly differently into practice because the animal numbers are rounded up to the nearest integer for paired design, and to the nearest larger even integer for balanced two-group design (Table 3).

3.5. Issue of ignoring intra-subject and intra-vocal fold correlations among rheological data

For the two-group design, it is obvious that $\frac{V^{(i)}}{V^{(g)}} \leq 1$ as shown in Eq. (8). For the paired study design, the influence of ignoring data correlation is not as straightforward and it depends on the values of m, ρ_1, and ρ_2. Eq. (7) shows that $\frac{V^{(i)}}{V^{(p)}} \leq 1$ when $\frac{\rho_2}{\rho_1} \leq \frac{m-1}{m}$. In reality, it is uncommon to observe $\frac{\rho_2}{\rho_1} > \frac{m-1}{m}$, which requires $\sigma_\gamma^2 \geq (m-1)\sigma_\xi^2$ or that the variability at the subject level is $(m-1)$ times greater than that at the vocal fold level. This was confirmed by our rabbit vocal fold rheological data: $\frac{\rho_2}{\rho_1}$ equaled to 0.287 for log G' and 0.274 for log η', respectively. Both were much smaller than $\frac{m-1}{m}$, which is 0.933 when $m = 15$. Thus, we can conclude that ignoring the correlations almost always results in an under-estimated variance of treatment effect. In rabbit, the variance of treatment effects will be under-estimated by 42–85% if the within-animal and within-vocal fold correlations are ignored (Table 5). Such severely

Table 2
τ^2, ρ_1, and ρ_2 of log-transformed elastic shear modulus G' and dynamic viscosity η' for rabbit vocal fold lamina propria.

Study design parameters	log G'	log η'
Variance of Y_{ijk} (τ^2)	0.624	0.349
Intra-vocal fold correlation (ρ_1)	0.300	0.073
Intra-animal correlation (ρ_2)	0.086	0.020

Table 5
Comparison of variances of estimated treatment effect in a rabbit vocal fold model ($m = 15$). $V^{(i)}$—the estimated variance of treatment effect when data correlations are ignored; $V^{(p)}$—the estimated variance of treatment effect for the paired design; $V^{(g)}$—the estimated variance of treatment effect for the two-group design.

	log G'	log η'
$V^{(i)}/V^{(p)}$	0.256	0.581
$V^{(i)}/V^{(g)}$	0.154	0.431

Table 3
Numbers of animals needed to detect different levels of treatment effect β_1 on G' and η', respectively, of rabbit vocal fold lamina propria ($m = 15$)*.

Level of increase or decrease	Paired design		Two-group design**	
	G'	η'	G'	η'
50% increase or 33% decrease ($\beta_1 = \pm 0.41$ on log scale)	21	5	34	8
100% increase or 50% decrease ($\beta_1 = \pm 0.70$ on log scale)	7	2	12	4
150% increase or 60% decrease ($\beta_1 = \pm 0.92$ on log scale)	5	1	8	2
200% increase or 67% decrease ($\beta_1 = \pm 1.10$ on log scale)	3	1	6	2

* Assuming 90% power and 0.05 two-sided significance level. Sample sizes were rounded up to the nearest integer for paired study design, and to the nearest larger even integer for balanced two-group study design.

** Sample size in the two-group design column denotes total number of animals in the study.

Table 4
Relative statistical efficiency (R) of the paired design to two-group design as a function of m (total number of measurements taken on each vocal fold specimen) for elastic shear modulus G' and dynamic viscosity η' in a rabbit vocal fold model.

m	6	8	10	12	15
R based on G'	1.52	1.57	1.61	1.63	1.66
R based on η'	1.19	1.24	1.27	1.31	1.35

under-estimated variances for the treatment effect will lead to greatly inflated type I errors (i.e. the risk of incorrectly declaring a significant treatment effect when there is none). Eqs. (7) and (8) also indicate that $V^{(i)}/V^{(g)}$ decreases with m, ρ_1, and ρ_2, while $V^{(i)}/V^{(p)}$ decreases with m and ρ_1 but increases with ρ_2.

4. Discussion

Rheological studies of vocal folds can benefit considerably from statistically designed experiments because experimental design has a major impact on sample size requirement. Correctly calculated sample sizes ensure experiments are adequately powered to detect treatment effect while avoiding wasting of resources. Using the mixed-effects model approach, the current work represents the first attempt to address the statistical implications of experimental designs for rheologic testing of vocal fold tissues. Closed-form formulas presented in this study enable researchers to estimate required sample size based on study design, number of planned measurements on each tissue sample, and parameters of covariance. We also found that if the correlations among rheological data are ignored during statistical analysis, type I errors will be substantially inflated. Although it was previously noted that the widely used regression-fitting methods are not suitable for the statistical analysis of data generated from repeated measurements (Xu et al., 2017; 2018), the current study demonstrated for the first time exactly how such inappropriate statistical methods affect the outcome of vocal fold rheological data analysis.

Our results indicated paired design could provide substantial gain in statistical power compared to two-group design. On the other hand, the benefits of a reduced sample size must be weighed against the potential limitations of using a paired design especially when it may be impractical or unattainable.

Treatment effect was assumed to be independent of frequency during statistical derivation and the assumption was validated by our rabbit vocal fold data. Thus, the sample size estimation formulas and conclusions about the two designs hold regardless of the frequency range of rheological measurement, which varies depending on the rheometer used in specific studies. On the other hand, significant frequency dependency of other causal effects such as species has been reported (Xu et al., 2017). Future studies of how the test frequency range might affect sample size estimation and statistical power are warranted.

As a final note, although a larger sample size enables researchers to find smaller differences statistically significant as demonstrated in Table 3, the differences found may be less scientifically meaningful. For this reason, researchers should exercise caution when determining the sample size and have a prior idea of what they would expect to be a scientifically meaningful difference.

5. Conclusions

This is the first study aimed to help researchers optimize statistical efficiency for the rheological studies of vocal fold tissues. A major finding of the present work was that the paired design could provide substantial gain in statistical power compared to the two-group design. Based on the mixed-effects model approach, closed-form formulas were derived for sample size estimation. We also found that ignoring correlation during vocal fold rheological data analysis will lead to inflated type I errors, which is the declaration of significant treatment effect when it does not exist. Lastly, our results indicated that with proper study design and data analysis methodology, a very small sample size (i.e. n = 3) is sufficient for detecting the large differences in G' and η' between scarred and normal vocal folds in a rabbit model.

Acknowledgements

This study was supported by the National Institutes of Health, NIDCD Grants R03DC011145 and R21DC013363. We thank Roger Chan and Miwako Kimura for their assistance in rheological data collection and the preparation of Fig. 1. We acknowledge Paula Timmons and Karen Pawlowski for their assistance in animal surgery. We would also like to thank Goldstein lab and Animal Resource Center in UT Southwestern Medical Center for providing animal cadavers and related records.

Conflict of interest statement

The authors declare that there is no conflict of interest in this work.

Appendix A. Calculation of $V^{(p)}$

Note that:

$$\begin{aligned} Var(Y_{i1k}-Y_{i2k}) &= Var(Y_{i1k}) + Var(Y_{i2k}) - 2Cov(Y_{i1k}, Y_{i2k}) \\ &= \tau^2 + \tau^2 - 2\tau^2\rho_2 = 2\tau^2(1-\rho_2), \end{aligned} \tag{A.1}$$

and

$$\begin{aligned} &Cov\left(Y_{i1k}-Y_{i2k}, Y_{i1k'}-Y_{i2k'}\right) \\ &= E\left((Y_{i1k}-Y_{i2k})\left(Y_{i1k'}-Y_{i2k'}\right)\right) - E(Y_{i1k}-Y_{i2k})E\left(Y_{i1k'}-Y_{i2k'}\right) \\ &= E\left(Y_{i1k}Y_{i1k'}\right) - E\left(Y_{i1k}Y_{i2k'}\right) - E\left(Y_{i2k}Y_{i1k'}\right) + E\left(Y_{i2k}Y_{i2k'}\right) - \beta_1^2 \\ &= \left(E\left(Y_{i1k}Y_{i1k'}\right) - E(Y_{i1k})E\left(Y_{i1k'}\right) + E(Y_{i1k})E\left(Y_{i1k'}\right)\right) \\ &\quad - \left(E\left(Y_{i1k}Y_{i2k'}\right) - E(Y_{i1k})E\left(Y_{i2k'}\right) + E(Y_{i1k})E\left(Y_{i2k'}\right)\right) \\ &\quad - \left(E\left(Y_{i2k}Y_{i1k'}\right) - E(Y_{i2k})E\left(Y_{i1k'}\right) + E(Y_{i2k})E\left(Y_{i1k'}\right)\right) \\ &\quad + \left(E\left(Y_{i2k}Y_{i2k'}\right) - E(Y_{i2k})E\left(Y_{i2k'}\right) + E(Y_{i2k})E\left(Y_{i2k'}\right)\right) - \beta_1^2 \\ &= \mathrm{Cov}\left(Y_{i1k}, Y_{i1k'}\right) + (\alpha_k + \beta_1)\left(\alpha_{k'} + \beta_1\right) - \mathrm{Cov}\left(Y_{i1k}, Y_{i2k'}\right) \\ &\quad - (\alpha_k + \beta_1)\alpha_{k'} - \mathrm{Cov}\left(Y_{i2k}, Y_{i1k'}\right) - \left(\alpha_{k'} + \beta_1\right)\alpha_k \\ &\quad + \mathrm{Cov}\left(Y_{i2k}, Y_{i2k'}\right) + \alpha_k\alpha_{k'} - \beta_1^2 \\ &= \rho_1\tau^2 - \rho_2\tau^2 - \rho_2\tau^2 + \rho_1\tau^2 = 2(\rho_1 - \rho_2)\tau^2. \end{aligned} \tag{A.2}$$

Thus, the variance of $\widehat{\beta}_1^{(p)}$ can be calculated:

$$\begin{aligned} V^{(p)} &= \mathrm{Var}\left(\widehat{\beta}_1^{(p)}\right) = Var\left(\frac{\sum_{i=1}^{n}\sum_{k=1}^{m}(Y_{i1k}-Y_{i2k})}{nm}\right) \\ &= \frac{\sum_{i=1}^{n} Var\left(\sum_{k=1}^{m}(Y_{i1k}-Y_{i2k})\right)}{n^2m^2} \\ &= \frac{n\sum_{k=1}^{m}\sum_{k'=1}^{m} Cov\left(Y_{i1k}-Y_{i2k}, Y_{i1k'}-Y_{i2k'}\right)}{n^2m^2} \\ &= \frac{n\left[2m\tau^2(1-\rho_2) + 2m(m-1)(\rho_1-\rho_2)\tau^2\right]}{n^2m^2} \\ &= \frac{2[1 + (m-1)\rho_1 - m\rho_2]\tau^2}{nm}. \end{aligned} \tag{A.3}$$

Appendix B. Calculation of $V^{(g)}$

For general randomized two-group study design:

$$\begin{aligned} V^{(g)} &= Var\left(\widehat{\beta}_1^{(g)}\right) = Var\left(\frac{\sum_{i=1}^{n_1}\sum_{j=1}^{2}\sum_{k=1}^{m}Y_{ijk}}{n_1 2m} - \frac{\sum_{i=n_1+1}^{n}\sum_{j=1}^{2}\sum_{k=1}^{m}Y_{ijk}}{n_2 2m}\right) \\ &= Var\left(\frac{\sum_{i=1}^{n_1}\sum_{j=1}^{2}\sum_{k=1}^{m}Y_{ijk}}{n_1 2m}\right) + Var\left(\frac{\sum_{i=n_1+1}^{n}\sum_{j=1}^{2}\sum_{k=1}^{m}Y_{ijk}}{n_2 2m}\right) \\ &= \frac{\sum_{i=1}^{n_1} Var\left(\sum_{j=1}^{2}\sum_{k=1}^{m}Y_{ijk}\right)}{4n_1^2m^2} + \frac{\sum_{i=n_1+1}^{n} Var\left(\sum_{j=1}^{2}\sum_{k=1}^{m}Y_{ijk}\right)}{4n_2^2m^2} \\ &= \frac{2mn_1\left[\tau^2+(m-1)\rho_1\tau^2+m\rho_2\tau^2\right]}{4n_1^2m^2} + \frac{2mn_2\left[\tau^2+(m-1)\rho_1\tau^2+m\rho_2\tau^2\right]}{4n_2^2m^2} \\ &= \frac{\tau^2+(m-1)\rho_1\tau^2+m\rho_2\tau^2}{2mn_1} + \frac{\tau^2+(m-1)\rho_1\tau^2+m\rho_2\tau^2}{2mn_2}. \end{aligned} \tag{B.1}$$

For balanced randomized two-group study design in which $n_1 = n_2 = n/2$:

$$V^{(g)} = Var\left(\widehat{\beta}_1^{(g)}\right) = \frac{2[1+(m-1)\rho_1+m\rho_2]\tau^2}{nm}. \tag{B.2}$$

Appendix C. Sample size formula for unbalanced two-group design

Let $\bar{r} = \frac{n_1}{n_1+n_2}$ indicate the proportion of subjects being assigned to the treatment group. Equation (B.1) can be rewritten as:

$$\begin{aligned} V^{(g)} &= \frac{\tau^2+(m-1)\rho_1\tau^2+m\rho_2\tau^2}{2mn\bar{r}} \\ &\quad + \frac{\tau^2+(m-1)\rho_1\tau^2+m\rho_2\tau^2}{2mn\left(1-\bar{r}\right)} \\ &= \frac{\tau^2+(m-1)\rho_1\tau^2+m\rho_2\tau^2}{2mn\bar{r}\left(1-\bar{r}\right)}. \end{aligned} \tag{C.1}$$

It follows immediately that the sample size formula for unbalanced two-group design is:

$$n^{(g)} = \frac{\left(z_{1-\alpha/2}+z_{1-\gamma}\right)^2[1+(m-1)\rho_1+m\rho_2]\tau^2}{\beta_{10}^2 2m\bar{r}\left(1-\bar{r}\right)}. \tag{C.2}$$

References

Bartlett, R.S., Guille, J.T., Chen, X., Christensen, M.B., Wang, S.F., Thibeault, S.L., 2016. Mesenchymal stromal cell injection promotes vocal fold scar repair without long-term engraftment. Cytotherapy 18, 1284–1296.

Bless, D.M., Welham, N.V., 2010. Characterization of vocal fold scar formation, prophylaxis, and treatment using animal models. Current Opinion Otolaryngol. Head Neck Surg. 18, 481–486.

Chan, R.W., Rodriguez, M.L., 2008. A simple-shear rheometer for linear viscoelastic characterization of vocal fold tissues at phonatory frequencies. J. Acoust. Soc. Am. 124, 1207–1219.

Chan, R.W., Titze, I.R., 1999. Viscoelastic shear properties of human vocal fold mucosa: measurement methodology and empirical results. J. Acoust. Soc. Am. 106, 2008–2021.

Choi, J.S., Kim, N.J., Klemuk, S., Jang, Y.H., Park, I.S., Ahn, K.H., Lim, J.Y., Kim, Y.M., 2012. Preservation of viscoelastic properties of rabbit vocal folds after implantation of hyaluronic acid-based biomaterials. Otolaryngol.–Head Neck Surg. 147, 515–521.

Cobell, W., Duflo, S.M., Magrufov, A., Thibeault, S.L., 2006. Fine needle aspiration of the vocal fold lamina propria in an animal model. Ann. Otol. Rhinol. Laryngol. 115, 764–768.

Hansen, J.K., Thibeault, S.L., 2006. Current understanding and review of the literature: vocal fold scarring. J. Voice 20, 110–120.

Hertegard, S., Dahlqvist, A., Laurent, C., Borzacchiello, A., Ambrosio, L., 2003. Viscoelastic properties of rabbit vocal folds after augmentation. Otolaryngol.–Head Neck Surg. 128, 401–406.

Hirano, S., 2005. Current treatment of vocal fold scarring. Current opinion Otolaryngol. Head Neck Surg. 13, 143–147.

Hirano, S., Bless, D.M., Rousseau, B., Welham, N., Montequin, D., Chan, R.W., Ford, C. N., 2004. Prevention of vocal fold scarring by topical injection of hepatocyte growth factor in a rabbit model. Laryngoscope 114, 548–556.

Jiao, T., Farran, A., Jia, X., Clifton, R.J., 2009. High frequency measurements of viscoelastic properties of hydrogels for vocal fold regeneration. Exp. Mech. 49, 235–246.

Kazemirad, S., Bakhshaee, H., Mongeau, L., Kost, K., 2014. Non-invasive in vivo measurement of the shear modulus of human vocal fold tissue. J. Biomech. 47, 1173–1179.

Kutty, J.K., Webb, K., 2009. Tissue engineering therapies for the vocal fold lamina propria. Tissue Eng. Part B, Rev. 15, 249–262.

Laird, N.M., Ware, J.H., 1982. Random-effects models for longitudinal data. Biometrics 38, 963–974.

Mau, T., Du, M., Xu, C.C., 2014. A rabbit vocal fold laser scarring model for testing lamina propria tissue-engineering therapies. Laryngoscope 124, 2321–2326.

Miri, A.K., Li, N.Y., Avazmohammadi, R., Thibeault, S.L., Mongrain, R., Mongeau, L., 2015. Study of extracellular matrix in vocal fold biomechanics using a two-phase model. Biomech. Model Mechanobiol. 14, 49–57.

Rousseau, B., Hirano, S., Chan, R.W., Welham, N.V., Thibeault, S.L., Ford, C.N., Bless, D. M., 2004. Characterization of chronic vocal fold scarring in a rabbit model. J. Voice 18, 116–124.

Svensson, B., Nagubothu, S.R., Cedervall, J., Chan, R.W., Le Blanc, K., Kimura, M., Ahrlund-Richter, L., Tolf, A., Hertegard, S., 2011. Injection of human mesenchymal stem cells improves healing of vocal folds after scar excision–a xenograft analysis. Laryngoscope 121, 2185–2190.

Thibeault, S.L., Gray, S.D., Bless, D.M., Chan, R.W., Ford, C.N., 2002. Histologic and rheologic characterization of vocal fold scarring. J. Voice 16, 96–104.

Thibeault, S.L., Klemuk, S.A., Chen, X., Quinchia Johnson, B.H., 2011. In Vivo engineering of the vocal fold ECM with injectable HA hydrogels-late effects on tissue repair and biomechanics in a rabbit model. J. Voice 25, 249–253.

Thibeault, S.L., Klemuk, S.A., Smith, M.E., Leugers, C., Prestwich, G., 2009. In vivo comparison of biomimetic approaches for tissue regeneration of the scarred vocal fold. Tissue Eng. Part A 15, 1481–1487.

Titze, I.R., 2000. Principles of Voice Production. National Center for Voice and Speech, Iowa City, IA.

West, B.T., Welch, K.B., Galecki, A.T., 2014. Linear mixed models: a practical guide using statistical software. Chapman and Hall/CRC, Boca Raton, FL.

Woo, P., Casper, J., Colton, R., Brewer, D., 1994. Diagnosis and treatment of persistent dysphonia after laryngeal surgery: a retrospective analysis of 62 patients. Laryngoscope 104, 1084–1091.

Xu, C.C., Chan, R.W., Sun, H., Zhan, X., 2017. A mixed-effects model approach for the statistical analysis of vocal fold viscoelastic shear properties. J. Mech. Behav. Biomed. Mater. 75, 477–485.

Xu, C.C., Chan, R.W., Weinberger, D.G., Efune, G., Pawlowski, K.S., 2010. Controlled release of hepatocyte growth factor from a bovine acellular scaffold for vocal fold reconstruction. J. Biomed. Mater. Res. Part A 93, 1335–1347.

Xu, C.C., Gao, A., Zhang, S., 2018. An investigation of left-right vocal fold symmetry in rheological and histological properties. Laryngoscope 128, E359–E364.