ELSEVIER

Journal of Biomechanics 41 (2008) 1623–1631

JOURNAL OF BIOMECHANICS

www.elsevier.com/locate/jbiomech
www.JBiomech.com

Biomechanical testing in experimental bone interventions—May the power be with you

Olli V. Leppänen[a,b], Harri Sievänen[c,*], Teppo L.N. Järvinen[a,b]

[a]*Medical School and the Institute of Medical Technology, University of Tampere, 33 014 Tampere, Finland*
[b]*Division of Orthopaedics and Traumatology, Department of Trauma, Musculoskeletal Surgery and Rehabilitation, Tampere University Hospital, Tampere, Finland*
[c]*The Bone Research Group, UKK Institute, P.O. Box 30, 33 500 Tampere, Finland*

Accepted 17 March 2008

Abstract

Total variation in any measured variable, in conjunction with expected treatment effect, defines the minimum sample size (minSS) required to detect the expected effect with statistical confidence should the effect truly exist. A comprehensive literature survey of 3472 original studies was carried out to identify studies with biomechanical testing of whole bones. Total variation in common biomechanical traits and expected treatment effects in typical interventions were statistically determined. According to this survey, total variation in biomechanical traits between different species of experimental animals was similar, justifying the use of rat femur as a model in further analyses. Due to poorer precision, stiffness and energy absorption assessment require substantially larger sample size than breaking load. Due to same reason, minSS for femoral neck compression test is considerably larger than for femoral shaft three-point bending test. For the bending test, minSS to show a 10% treatment effect in the breaking load with 80% statistical power is 11 rats/group, while corresponding minSS is 23 for the stiffness, and 53 for the energy absorption. For the femoral neck compression test, minSSs are 16, 51, and 134 rats/group, respectively. Among the reviewed studies, the mean sample size was 11 animals/group. This underscores the need for considerably larger sample sizes in experimental bone interventions which employ mechanical traits as primary outcome variables. In particular, poor precision and generally small expected treatment effects compromise the utility of stiffness and energy absorption assessments in experimental bone interventions.

Keywords: Sample size; Statistical power; Precision; Mechanical testing; Experimental studies

1. Introduction

Skeleton has primarily evolved to allow efficient *locomotion* (Burr, 1997; Frost, 1997; Parfitt, 1998), and accordingly, bone mechanical competence depicts its ultimate phenotype (Jarvinen et al., 2005; van der Meulen et al., 2001). Biomechanical testing provides a direct method to study mechanical traits of bones among various experimental animals with different testing protocols (Andreassen and Oxlund, 2000; Cullinane et al., 2002; Fleming et al., 2000; Ikeda et al., 2003; Jerome et al., 1999; Judex and Zernicke, 2000; Klein et al., 2001, 2003; Les et al., 2002; Luppen et al., 2002). Besides careful technical execution (Turner and Burr, 1993), general utility of biomechanical testing relies on its precision. Precision is consistently required for studies using bone densitometry, but this is not the case for biomechanical testing of bones. Obviously, precision, in the stringent sense, cannot be determined because of destructive nature of the method. However, precision can be reasonably assessed from within-pair variation in biomechanical and structural traits of bone (Eckstein et al., 2004; Jamsa et al., 1998; Jarvinen et al., 1998b; Leppanen et al., 2006; Peng et al., 1994).

Total variation observed in any trait reflects both biological variation and methodological variation. Precision of the method affects thus the statistical power of the study; i.e., the probability that the study could detect the

*Corresponding author. Tel.: +358 3 282 9111; fax: +358 3 282 9200.
E-mail address: harri.sievanen@uta.fi (H. Sievänen).

0021-9290/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jbiomech.2008.03.017

expected treatment effect with statistical confidence if the effect truly existed. For obvious scientific and ethical reasons, the sample size should be planned carefully before starting any experiment (Eisman, 2006), so that the number of animals/group is large enough to provide adequate statistical power. This is crucial as underpowered studies can seldom address any research question meaningfully, but only lead to inconclusive results (Ioannidis, 2005).

Detectable treatment effect is inversely related to sample size, meaning that large samples are needed to reveal small effects (Altman et al., 2001). In practice, appropriate determination of sample size requires realistic estimates of the treatment effect and the total variation in individual responses to given treatment; desired level of statistical significance for the expected results (type I error); and desired statistical power (type II error).

Objectives of the present study were fourfold: (1) to evaluate total variation in various biomechanical traits of whole bones; (2) to estimate treatment effects on biomechanical traits in experimental interventions based on ovariectomy, increased activity, and inactivity using rat femur as a model; (3) to characterize methodological variation in biomechanical testing of rat femur to illuminate contribution of biological and methodological variation to total variation in biomechanical traits; and (4) to devise a scheme for minimum sample size (minSS) needed to detect a treatment effect in biomechanical traits with statistical significance.

2. Materials and methods

2.1. Total variation in biomechanical traits

We reviewed all 3472 original studies published between 1999 and 2003 in *Bone*, *Calcified Tissue International*, *Journal of Bone and Mineral Research*, or *Journal of Orthopaedic Research*. We considered this sample of four major bone journals representative of contemporary status of experimental bone intervention studies. Inclusion criteria were (1) mechanical testing of whole bones was performed; (2) bones were extracted either from rat, mouse, dog, rabbit, or monkey; and (3) the study had an intact control group. Accordingly, 123 studies (see Supplementary data) were included.

Number of animals (n), mean and standard deviation (S.D.) of breaking load, stiffness, and energy absorption were collected for each control group, whenever available. When standard error of mean (S.E.M.) was given instead of S.D., S.D. was calculated as

$$\text{S.D.} = \sqrt{n} \times \text{S.E.M.} \tag{1}$$

Percentage variation (σ_T) in each study and trait was calculated as

$$\sigma_T = \frac{\text{S.D.}}{\text{Mean}} \times 100\%. \tag{2}$$

The mean total percentage variation ($\bar{\sigma}_T$) for each type of intervention and trait was calculated as

$$\bar{\sigma}_T = \sqrt{\frac{\sum \sigma_T^2}{k}}, \tag{3}$$

where k is the number of separate studies.

2.2. Treatment effects on biomechanical traits

To obtain appropriate estimates of treatment effects in typical experimental bone interventions, we chose the common rat model and included studies that met the following criteria: (1) femoral shaft three-point bending test or femoral neck compression test[1] was performed; and (2) intervention was either *ovariectomy*, *increased activity* (climbing, treadmill training, voluntary wheel-running), or *inactivity* (neurectomy, hindlimb suspension, limb taping). Altogether, data from 40 studies (see Supplementary data) were included.

To estimate the effect size of each intervention (in z-scores and %-values) meta-analytic principles described in Eqs. (4)–(12) (Eqs. (1)–(3) also apply as appropriate) were employed (see Appendix A) (Cooper and Hedges, 1994).

2.3. Sample size estimation

The minimum sample size (minSS) needed to show a specified treatment effect (δ) in mechanical traits with statistical significance of $p = 0.05$ (if the effect truly existed) was calculated using an approximation of Neyman's solution (Snedecor and Cochran, 1967) as

$$\text{minSS} = 7.9\frac{\bar{\sigma}_T^2}{\delta^2} \quad \text{for the statistical power of 0.80} \tag{13}$$

and

$$\text{minSS} = 10.5\frac{\bar{\sigma}_T^2}{\delta^2} \quad \text{for the statistical power of 0.90.} \tag{14}$$

2.4. Methodological variation in biomechanical traits

The relationship between total variation (σ_T), biological variation (σ_B) in the tested trait and the precision ($CV\%_{rms}$) is given by

$$\bar{\sigma}_T = \sqrt{\bar{\sigma}_B^2 + CV\%_{rms}^2}. \tag{15}$$

Several studies (Ammann et al., 2000; Jamsa et al., 1998; Jarvinen et al., 1998a, b; Leppanen et al., 2006; Peng et al., 1994) have shown that the precision of biomechanical tests can be assessed using paired specimens. This approach presumes that contralateral bones are equal, which may not always be true (Banse et al., 1996; Hanson and Markel, 1994). However, when the paired bones are extracted from healthy animals developed under normal circumstances, no systematic difference would apparently exist. Thus, the precision was calculated as

$$CV\%_{rms} = \sqrt{\frac{\sum(100 \times (\text{right} - \text{left}/\text{right} + \text{left}))^2}{n}}, \tag{16}$$

where n is the number of femur pairs, and right and left denote the measured values from respective femora.

Precision of femoral shaft three-point bending and femoral neck compression tests was determined using test results from femora of 60 Sprague-Dawley rats (age: 17–69 weeks, and body weight: 240–630 g). Excised and defleshed bones were wrapped in saline-soaked gauze bandages to prevent dehydration and stored at −20 °C. This procedure does not affect bone mechanical properties (Pelker et al., 1984; Sedlin and Hirsch, 1966). The research protocol was accepted by Ethics Committee for Animal Experiments of the University of Tampere. The study conformed to NIH Guide for the Care and Use of Laboratory Animals.

At the testing day, the femora were thawed at the room temperature and kept in the saline-soaked gauzes except during measurements.

[1]Note that the mechanical testing of the femoral neck is rather a cantilever bending test than a compression test. However, the latter term has become established in the literature, and for the sake of consistency, the femoral neck test is called a compression test in the present study.