Gait & Posture 59 (2018) 182–187

Contents lists available at ScienceDirect

Gait & Posture

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

Full length article

Analysis of data collected from right and left limbs: Accounting for dependence and improving statistical efficiency in musculoskeletal research

Sarah Stewart[a,*], Janet Pearson[b], Keith Rome[a], Nicola Dalbeth[c], Alain C. Vandal[b,d]

[a] Department of Podiatry, Health & Rehabilitation Research Institute, Auckland University of Technology, Private Bag 92006, Auckland 1142, New Zealand
[b] Department of Biostatistics & Epidemiology, Faculty of Health and Environmental Sciences, Auckland University of Technology, Private Bag 92006, Auckland 1142, New Zealand
[c] Faculty of Medical and Health Sciences, The University of Auckland, Private Bag 92019, Auckland 1142, New Zealand
[d] Ko Awatea, Counties Manukau District Health Board, Private Bag 93311, Auckland 1640, New Zealand

ARTICLE INFO

Keywords:
Plantar pressure
Gait
Statistical analysis
Lower limb
Foot
Mixed effects models
Statistical efficiency

ABSTRACT

Objectives: Statistical techniques currently used in musculoskeletal research often inefficiently account for paired-limb measurements or the relationship between measurements taken from multiple regions within limbs. This study compared three commonly used analysis methods with a mixed-models approach that appropriately accounted for the association between limbs, regions, and trials and that utilised all information available from repeated trials.

Method: Four analysis were applied to an existing data set containing plantar pressure data, which was collected for seven masked regions on right and left feet, over three trials, across three participant groups. Methods 1–3 averaged data over trials and analysed right foot data (Method 1), data from a randomly selected foot (Method 2), and averaged right and left foot data (Method 3). Method 4 used all available data in a mixed-effects regression that accounted for repeated measures taken for each foot, foot region and trial. Confidence interval widths for the mean differences between groups for each foot region were used as a criterion for comparison of statistical efficiency.

Results: Mean differences in pressure between groups were similar across methods for each foot region, while the confidence interval widths were consistently smaller for Method 4. Method 4 also revealed significant between-group differences that were not detected by Methods 1–3.

Conclusion: A mixed effects linear model approach generates improved efficiency and power by producing more precise estimates compared to alternative approaches that discard information in the process of accounting for paired-limb measurements. This approach is recommended in generating more clinically sound and statistically efficient research outputs.

1. Introduction

Most rheumatic diseases, including rheumatoid arthritis, gout, osteoarthritis, psoriatic arthritis and spondyloarthropathy, present with a variety of musculoskeletal manifestations. Gout, osteoarthritis and psoriatic arthritis are often characterised by an asymmetrical pattern of distribution with regard to musculoskeletal symptoms, in that right and left limbs are not always affected equally. Clinical research in musculoskeletal rheumatology often involves the collection of data from right and left limbs from the same participant, resulting in limb-specific units of analysis, as opposed to person-specific units of analysis that occur when data are collected on single organ systems. However, person-specific and disease-specific factors, including age, gender, ethnicity, disease duration and the use of pharmacological therapy, result in a high level of within-subject dependence between limbs, meaning that data from right and left limbs are often highly correlated [1]. The same is true for multiple measurements taken from each limb, including from a range of joints or regions within limbs. This becomes problematic in the application of many commonly used statistical procedures, including linear models (such as the *t*-test and analysis of variance) that assume each data point is an independent observation [2].

It is not uncommon for researchers to pool data from right and left sides without accounting for the between-side correlation [3,4]. This approach is often considered a valid method if the dependent variables of interest are limb-specific rather than person-specific. Pooling of right and left limb data also provides an appealing option as it apparently

* Corresponding author.
E-mail addresses: sarah.stewart@aut.ac.nz (S. Stewart), janet.pearson@aut.ac.nz (J. Pearson), krome@aut.ac.nz (K. Rome), n.dalbeth@auckland.ac.nz (N. Dalbeth), alain.vandal@aut.ac.nz (A.C. Vandal).

http://dx.doi.org/10.1016/j.gaitpost.2017.10.018
Received 9 August 2017; Received in revised form 5 October 2017; Accepted 13 October 2017

Table 1
Peak plantar pressure (kPa) using Analysis Method 1 (general linear regression analysing right foot data only).

Parameter		Least-squares mean	Diff.	95% CI		p
				Lower	Upper	
Heel	Control	274.3				
	Gout	254.9	−19.5	−57.7	18.8	0.315
	Asymptomatic hyperuricemia	298.8	24.5	−12.8	61.8	0.196
Midfoot	Control	97.4				
	Gout	126.8	29.4	3.1	55.7	0.029
	Asymptomatic hyperuricemia	126.5	29.1	3.5	54.7	0.027
First metatarsal	Control	216.0				
	Gout	228.6	12.5	−31.4	56.5	0.571
	Asymptomatic hyperuricemia	245.9	29.9	−12.9	72.7	0.168
Second metatarsal	Control	296.5				
	Gout	309.1	12.6	−25.9	51.1	0.516
	Asymptomatic hyperuricemia	320.2	23.7	−13.8	61.2	0.212
Third to fifth metatarsals	Control	258.1				
	Gout	248.2	−9.9	−44.5	24.7	0.570
	Asymptomatic hyperuricemia	248.9	−9.2	−42.9	24.6	0.591
Hallux	Control	223.0				
	Gout	213.9	−9.1	−55.4	37.1	0.695
	Asymptomatic hyperuricemia	246.5	23.5	−21.5	68.6	0.302
Lesser toes	Control	104.5				
	Gout	127.0	22.4	−8.0	52.9	0.147
	Asymptomatic hyperuricemia	108.9	4.4	−25.3	34.1	0.769

Results are presented adjusted for age and BMI. Bolded P values indicate significant difference between groups at $P < 0.05$. Diff. = Difference in least-squares mean from control group; CI = Confidence Interval.

Table 2
Peak plantar pressure (kPa) using Analysis Method 2 (general linear regression analysing random left or right foot).

Parameter		Least-squares mean	Diff.	95% CI		p
				Lower	Upper	
Heel	Control	304.5				
	Gout	269.2	−35.3	−72.3	1.7	0.061
	Asymptomatic hyperuricemia	304.1	−0.4	−36.6	35.7	0.982
Midfoot	Control	84.2				
	Gout	133.8	49.6	25.8	73.4	< 0.001
	Asymptomatic hyperuricemia	110.8	26.6	3.4	49.9	0.025
First metatarsal	Control	223.8				
	Gout	242.5	18.7	−18.4	55.9	0.319
	Asymptomatic hyperuricemia	241.7	17.9	−18.4	54.2	0.329
Second metatarsal	Control	280.9				
	Gout	273.4	−7.5	−48.7	33.7	0.719
	Asymptomatic hyperuricemia	324.4	43.5	3.2	83.8	0.035
Third to fifth metatarsals	Control	239.3				
	Gout	225.4	−14.0	−48.2	20.3	0.419
	Asymptomatic hyperuricemia	250.1	11.0	−22.7	44.2	0.524
Hallux	Control	244.8				
	Gout	208.8	−36.0	−82.8	10.8	0.130
	Asymptomatic hyperuricemia	244.4	−0.4	−46.1	45.3	0.986
Lesser toes	Control	101.9				
	Gout	126.0	24.0	−3.5	51.6	0.086
	Asymptomatic hyperuricemia	104.5	2.54	−24.4	29.4	0.851

Results are presented adjusted for age and BMI. Bolded P values indicate significant difference between groups at $P < 0.05$. Diff. = Difference in least-squares mean from control group; CI = Confidence Interval.

doubles the sample size while maintaining the same number of participants. However, pooling data results in artificial deflation of confidence intervals and significance levels [5,6] that increases the probability of a type I error (rejecting the null hypothesis when it is true) compared to the nominal significance level α.

Several alternative methods have been used in musculoskeletal research to account for between-limb dependence, including undertaking a separate analysis of one or both limbs, whether this be the right and/or left [7–9], the most dominant side [10], the side with the most clinically evident symptoms of disease [11], or a randomly chosen side. However, such approaches result in a loss of valuable data and thus a reduction in statistical power and precision of estimates, and are overall inefficient methods of analysis. Furthermore, they may introduce a bias through the choice of which limb to use, particularly if a non-random selection approach is adopted.

Another commonly used method is to average data from right and left limbs. This becomes particularly problematic in rheumatic diseases that present with asymmetrical involvement, for example, osteoarthritis, gout and spondyloarthropathy, as averaging data may lessen the apparent magnitude of the disease and can lead to inaccurate inferences. Furthermore, without regarding the right and left sides as repeated within-subject measurements, efficiency and power are also lost. Similarly, averaging of repeated measurements is also common practice when measuring outcomes in quantitative research, whereby data is obtained over multiple trials (generally three) for each limb and their average used in subsequent analyses. Averaging is primarily undertaken to reduce measurement error; however, this method also removes useful information when the number of averaged trials may differ. Inefficiencies also arise when variables measured from multiple joints or regions within each limb are analysed separately without appropriately accounting for between-region correlations [12].

The issue of between-limb dependence in statistical analysis has been identified in several research fields, including ophthalmology [13], podiatry [14,15], orthopaedics [16] and rheumatology [17]. However, there is currently no consensus on the correct analytical approach of data collected from multiple trials from multiple limbs and/or regions within limbs. This article aims to compare three linear regression techniques, commonly used in current research under a generally incorrect assumption of independence between regions, with a mixed linear regression model that provides a more appropriate account for the association between limbs, regions, and trials, and that utilises all information available from repeated trials.

2. Methods

2.1. Data set

For the purpose of illustrating the various analysis methods in the current article, peak plantar pressure data, a continuous variable measured in kilopascals (kPa), was taken from a larger data set [18]. The aim of the original study was to compare the plantar pressure distribution during barefoot walking in people with gout (n = 25) or people with asymptomatic hyperuricaemia (n = 27) with that of healthy individuals with normal serum urate concentrations (n = 34). Plantar pressure data was collected for both right and left feet of each participant over three repeated walking trials. Peak plantar pressure