ELSEVIER

Journal of Electromyography and Kinesiology 12 (2002) 45–57

JOURNAL OF ELECTROMYOGRAPHY AND KINESIOLOGY

www.elsevier.com/locate/jelekin

Statistical power and measurement allocation in ergonomic intervention studies assessing upper trapezius EMG amplitude A case study of assembly work

Svend Erik Mathiassen [a, b,*], Alex Burdorf [c], Allard J. van der Beek [c, d]

[a] *Division of Production Ergonomics, School of Technology and Society, Malmö University, SE-20506 Malmö, Sweden*
[b] *National Institute for Working Life, Sweden*
[c] *Department of Public Health, Erasmus University Rotterdam, The Netherlands*
[d] *Department of Social Medicine, Institute for Research in Extramural Medicine, Vrije Universiteit Medical Centre, Amsterdam, The Netherlands*

Received in revised form 1 April 2001; accepted 27 June 2001

Abstract

The present study aimed at exploring the statistical power of ergonomic intervention studies using electromyography (EMG) from the upper trapezius muscle. Data from a previous study of cyclic assembly work were reanalyzed with respect to exposure variability between subjects, between days, and within days. On basis of this information, the precision and power of different data collection strategies were explored. A sampling strategy comprising four registrations of about two min each (i.e. two work cycles) for one day per subject resulted in coefficients of variation between subjects on the 10-, 50-, and 90-APDF-percentiles of 0.44, 0.31, and 0.29, respectively. The corresponding necessary numbers of subjects in a study aiming at detecting a 20% exposure difference between two independent groups of equal size were 154, 78, and 68, respectively ($p \leq 0.05$, power 0.80). Multiple measurement days per subject would improve power, but only to a marginal extent beyond 4 days of recording. Increasing the number of recordings per day would have minor effects. Bootstrap resampling of the data set revealed that estimates of variability and power were associated with considerable uncertainty. The present results in combination with an overview of other occupational studies showed that common-size investigations using trapezius EMG percentiles are at great risk of suffering from insufficient statistical power, even if the expected intervention effect is substantial. The paper suggests a procedure of how to retrieve and use exposure variability information as an aid when studies are planned, and how to allocate measurements efficiently.

Keywords: Bootstrapping; Ergonomics; Exposure variability; Neck and shoulder; Study design

1. Introduction

Changes in mechanical exposure (physical work load) have been used as proxies for expected health effects in a large number of ergonomic intervention studies. This approach has been prominent in controlled experimental studies of specific characteristics of a job, e.g. tools [4,43], work station design [31] or work pace [38], as well as in field studies of production systems with a fast turn-over of production processes or labour [7,20]. One important reason is that changes in exposure are in general easier to assess and interpret as the specific result of an intervention than health outcomes. In recent years, it has become increasingly evident that direct recordings of mechanical exposure are superior to self-reports or observations as regards accuracy and resolution [49,56]. Substantial efforts have been invested in developing personal monitors allowing mechanical exposure data to be continuously sampled and stored for hours during occupational work [3,24].

Surface electromyography (EMG) has been extensively used in working life research. In 1995, a review identified 97 internationally published papers using recordings of EMG from the upper trapezius muscle, several of which compared EMG amplitudes during differ-

* Corresponding author. Tel.: +46-(0)40-6657702; fax: +46-(0)40-6657706.
E-mail address: svend.erik.mathiassen@ts.mah.se (S.E. Mathiassen).

1050-6411/02/$ - see front matter
PII: S1050-6411(01)00028-1

ent conditions in simulated or real work [39]. Thus, upper trapezius EMG amplitude may serve as a prominent example of a directly recorded exposure variable in ergonomic intervention studies.

However, the use of EMG requires considerable resources in terms of equipment, competence and time. Studies assessing upper trapezius EMG amplitude have therefore in general been conducted on small groups, rarely exceeding 15 subjects, as shown by a comprehensive selection of manual handling studies (Table 1). In addition, a majority of the selected studies reported an exposure standard deviation between subjects beyond 50% of the group mean EMG amplitude. The combination of a small study size and a large exposure variability leads to the risk of low statistical power, i.e. an insufficient ability of the study to detect statistically significant differences in exposure. Analyses of power and optimal allocation of measurement resources are widely accepted study design tools, giving guidance to the necessary investment of measurement resources to reach an acceptable chance of success in a planned study, as well as to an efficient use of these resources. So-far, however, these tools have been only sporadically discussed in the context of ergonomic intervention studies in general, and even less so with a specific focus on EMG [5,37].

Conventional power analysis accepts that group mean values of exposure may fluctuate according to an observed variance within and between subjects, but it does not take into account that this exposure variance is in itself a stochastic variable, subject to random error [9]. The influence on power estimates of this additional source of uncertainty has not previously been discussed with reference to ergonomic intervention studies.

The present investigation had the general purpose of illustrating a procedure for considering statistical power and resource allocation in the design of studies comparing two independent groups. These issues were explored with respect to assessments of upper trapezius EMG amplitude. Data were obtained from a previous study with a repeated-measures design which allowed values of mean exposure and variance components to be derived, as well as their statistical distributions.

2. Material and methods

2.1. Parent data and their empirical distribution

EMG data were obtained from a previous study of eight females performing light, cyclic assembly work at controlled pace in a laboratory mock-up of an industrial setting [38]. EMG was collected at 500 Hz from the right upper trapezius muscle during 136 s (two work cycles) at one hour intervals for four times a day during four separate days, in all 16 registrations per subject (Fig. 1). Off-line, EMG was rms-converted (moving window, 100 ms) and normalized using the mean rms-EMG amplitude of one submaximal reference contraction preceding work (%RVE, [39]). The reference contraction required about 15% of the maximal capacity. Further details can be found in the original study [38]. The normalized EMG amplitude of each recording was processed to give the 10-, 50- and 90-percentiles of the cumulative amplitude distribution function (the APDF, [28]). Thus, for each APDF-percentile, the data set consisted of 128 values, or 'quanta' (eight subjects, 4 days per subject, four quanta per day).

For each percentile, the total exposure variability among the 128 values was partitioned according to a nested model [47]:

$$E_{sdq}=\mu+\alpha_s+\beta_{sd}+e_{sdq}$$

where: E_{sdq} value of exposure recording ('quantum') q, collected day d in subject s; μ general mean; α_s residual due to the effect of subject, s=1, 2, …, n_s; β_{sd} residual due to the effect of day within subject, d=1, 2, …, n_d; and e_{sdq} residual due to the effect of quantum within day within subject, q=1, 2, …, n_q. All α_s, β_{sd}, and e_{sdq} are assumed to be independently and identically distributed with zero means, i.e.

α_s~i.i.d.$(0,\sigma_s^2)$, β_{sd}~i.i.d.$(0,\sigma_d^2)$, and e_{sdq}~i.i.d.$(0,\sigma_q^2)$.

Variance components were obtained by resolving expected mean squares in a two-way nested ANOVA (Table 2).

Thus, the parent set of parameters comprised, for each of the three APDF-percentiles, estimates of the overall mean value (m) and three variance components: variance between subjects (s_s^2), between days within subject (s_d^2), and between quanta within day within subject (s_q^2).

In order to explore the precision of the observed parameters, a bootstrap technique was applied. Bootstrapping generates an empirical distribution of some target parameter by executing a large number of virtual replications of the original experiment [14]. In the present case, eight subjects were randomly selected with replacement from the parent group of eight. For each of the eight drawn subjects, 4 days were randomly selected with replacement from that subject's original 4 days of measurement. Finally, for each of these selected days, four registrations were drawn with replacement among the four available registrations for that specific subject and day. This 'nested resampling' procedure resulted in a virtual, secondary data set of the same size as the parent one, and the mean values and variance components of this secondary data set were derived as described above. The resampling procedure was repeated 1000 times to yield 1000 empirical values of each of the parameters m, s_q^2, s_d^2, and s_s^2 for each APDF-percentile. Bias-corrected 95% confidence intervals for the observed values of