Scand J Work Environ Health 1995;21(1):15-23
https://doi.org/10.5271/sjweh.3
Issue date: Feb 1995

Reducing random measurement error in assessing postural load on the back in epidemiologic surveys

by Burdorf A

The following articles refer to this text: 2001;27(2):125-132; 2007;33(5):351-357; 2014;40(3):252-265

Key terms: bias; epidemiology; physical load; reliability

This article in PubMed: www.ncbi.nlm.nih.gov/pubmed/7784859

Print ISSN: 0355-3140 Electronic ISSN: 1795-990X

Scand J Work Environ Health 1995;21:15–23

Reducing random measurement error in assessing postural load on the back in epidemiologic surveys

by Alex Burdorf, PhD[1]

Burdorf A. Reducing random measurement error in assessing postural load on the back in epidemiologic surveys. *Scand J Work Environ Health 1995;21:15–23.*

Objectives The goal of this study was to design strategies to assess postural load on the back in occupational epidemiology by taking into account the reliability of measurement methods and the variability of exposure among the workers under study.

Methods Intermethod reliability studies were evaluated to estimate the systematic bias (accuracy) and random measurement error (precision) of various methods to assess postural load on the back. Intramethod reliability studies were reviewed to estimate random variability of back load over time.

Results Intermethod surveys have shown that questionnaires have a moderate reliability for gross activities such as sitting, whereas duration of trunk flexion and rotation should be assessed by observation methods or inclinometers. Intramethod surveys indicate that exposure variability can markedly affect the reliability of estimates of back load if the estimates are based upon a single measurement over a certain time period. Equations have been presented to evaluate various study designs according to the reliability of the measurement method, the optimum allocation of the number of repeated measurements per subject, and the number of subjects in the study.

Conclusion Prior to a large epidemiologic study, an exposure-oriented survey should be conducted to evaluate the performance of measurement instruments and to estimate sources of variability for back load. The strategy for assessing back load can be optimized by balancing the number of workers under study and the number of repeated measurements per worker.

Key terms bias, epidemiology, physical load, reliability.

Mechanical load is regarded as a primary cause of musculoskeletal disorders. In epidemiologic studies no instruments are available with which to measure mechanical load directly on specific segments of the human body. Hence biomechanical models are commonly used to estimate the forces and moments acting on the location of interest. For example, a biomechanical model has been developed to predict compression and shear forces at the level of the L_5-S_1 intervertebral disc; it takes into account the weight and position of the trunk, the arms, the upper and lower legs, and any external load if present (1). Similar approaches using biomechanical models to estimate postural load on other body segments have been published for the shoulder (2) and the neck (3). When biomechanical models are applied to predict postural load, exposure to postural load can be estimated by measuring the angular position of the body segment of interest.

Several techniques have been developed to measure the distributions of angular position of body segments of workers performing their regular activities. These techniques range from subjective methods such as questionnaires to objective methods with real-time recording of the posture of the human body (4). Questionnaires can be used to collect reasonably simple data, but detailed and complex data cannot be sought without risk of substantial measurement error (5). Therefore, questionnaires may be too coarse a method with which to arrive at a quantitative conclusion on an individual's exposure. Sophisticated measurement techniques like three-dimensional video analysis may provide satisfactory information, but their applicability in epidemiologic studies is hampered since they are often more elaborate, expensive or time-consuming (6). Moreover, these techniques may contribute information too detailed to summarize exposure patterns in parameters useful in epidemiologic studies.

In epidemiologic studies the validity and the practicability of a method for assessing postural load must be balanced. Feasibility considerations may argue for the

1 Institute of Public Health, Erasmus University, PO Box 1738, 3000 DR Rotterdam, The Netherlands.

Reprint requests to: Dr Alex Burdorf, Institute of Public Health, Erasmus University, PO Box 1738, 3000 DR Rotterdam, The Netherlands.

use of alternative methods that are less precise or less valid or both. This article discusses the choice of a measurement method in relation to exposure assessment strategy. The assessment of postural load on the back is taken as an example, but many considerations also hold for the assessment of postural load on other body segments. This paper starts with a brief review of the instruments of exposure assessment currently used in epidemiologic studies on associations between postural load on the back and the occurrence of back disorders. Errors in exposure assessment are examined by means of validity and reliability aspects. Finally, the reliability of a particular measurement method is evaluated in relation to the design of a measurement strategy in a particular epidemiologic study.

Assessing the validity and reliability of measurement methods

Validity and reliability studies are essential tools in addressing misclassification and measurement error in categorical and continuous exposure measures, respectively. Moreover, estimates of parameters of validity and reliability can be used to evaluate the attenuation of associations between exposure and disease due to nondifferential measurement error (5). A general definition of validity is the degree to which a measurement measures what it purports to measure (7). The validity of a measurement technique can be derived from a comparison with an instrument that measures accurately (ie, zero bias) and precisely (ie, small random error), like a gold standard that measures the true exposure value, on the average, with a random error sufficiently small considering its purpose.

Fairly often a perfect measurement instrument is not available or it is infeasible so that alternative methods are used to ascertain the validity of a measurement technique. An appropriate approach is to conduct a reproducibility study to obtain indirect information on error distributions (5, 8). In reproducibility studies two or more separate assessments of exposure are performed on the same individuals, either by different instruments, if intermethod reliability is to be assessed, or by repeated measurements with the same instrument, if intramethod reliability is to be estimated. The rationale behind this procedure can be made more explicit by considering the relationship between true scores (T) and observed scores (X) with two imperfect measurement techniques in a simplified model:

$$X_{1i} = \alpha_1 + \beta_1 T_i + \varepsilon_{1i} \wedge X_{2i} = \alpha_2 + \beta_2 T_i + \varepsilon_{2i}, \qquad \text{equation 1}$$

where X_{1i} = observed score for subject i at time t with method 1, X_{2i} = observed score for subject i at time $t + \Delta t$ with method 2, T_i = true score for subject i at time t in random sample of n subjects, α_1 = constant bias, β_1 = relative bias, and ε_{1i} = random measurement error.

In studies of intramethod reliability methods 1 and 2 are the same and $\Delta t > 0$. In intermethod studies methods 1 and 2 are different and Δt can be zero. Method 1 is considered perfect if $\alpha_l = 0$ and $\beta_l = 1$, with ε_l being sufficiently small.

In the classical test theory the described parallel test model is used to obtain the reliability coefficient, $\rho_{X_1X_2}$, by correlating parallel measurements on a population of N subjects. Under strict assumptions the reliability coefficient can yield information about the validity coefficient of either method 1 or method 2, although these latter coefficients, $\rho^2_{TX_1}$ and $\rho^2_{TX_2}$, cannot be estimated directly from empirical data. In formula form (9):

$$\rho_{X_1X_2} = \rho^2_T / (\rho^2_T + \rho^2_\varepsilon) = \rho^2_{TX_1}\, \rho^2_{TX2}. \qquad \text{equation 2}$$

Essential assumptions are: normally distributed parameters T, X_1, X_2, ε_l and ε_2; X_1 and X_2 have equal true scores ($\alpha_l = \alpha_2$ and $\beta_1 = \beta_2 = 1$); random errors with $E(\varepsilon_l) = E(\varepsilon_2) = 0$; equal error variances with $\sigma_{\varepsilon_1}{}^2 = \sigma_{\varepsilon_2}{}^2 = \sigma_\varepsilon{}^2$; and uncorrelated errors with $\rho_{\varepsilon_1\varepsilon_2} = 0$ and $\rho_{T\varepsilon_1} = \rho_{T\varepsilon_2} = 0$. The assumption of equal variances may be incorrect, particularly when an exposure measure X_1 of moderate precision (eg, derived from a diary) is compared with measure X_2 with greater precision (eg, derived from an observation method). If the other assumptions are not violated, it can be demonstrated that (9):

$$\rho_{X_1X_2} < \rho_{TX_1} < \sqrt{\rho_{X_1X_2}}. \qquad \text{equation 3}$$

This equation presents an upper and lower boundary for the validity coefficient of exposure measure X_1. The intermethod reliability coefficient is estimated by the Pearson correlation coefficient r of continuous variables and the Spearman correlation coefficient r_s of categorical variables with an ordinal scale (5).

One has to bear in mind that the reliability coefficient obtained by parallel testing depends on the range of the true scores in the sample. Reliability will increase with a wider range. Therefore, the use of the reliability coefficient as a measure of agreement has been criticized (10). Caution is also needed because the reliability coefficient in a particular sample of n subjects is an estimate of the true reliability coefficient in the total population of N subjects. Reliability in another sample of subjects with a different exposure distribution may result in a different coefficient.

The parallel test method can also be used within the context of an intramethod reliability study by applying a particular instrument to the same subjects at two or more points in time. This is of special interest when exposure variability due both to imperfect measurements with random error and to variability of exposure over time is evaluated. The Pearson correlation coefficient r is not appropriate for intramethod surveys since systematic bias (constant or relative) is not reflected in the correlation coefficient. In intramethod studies the reliability ρ_X is estimated by the intraclass coefficient r_I for continuous variables and the weighted Cohen's kappa κ_ω for categorical variables (5). These estimates treat systematic bias as part of the random error. When bias is absent, however, ρ_X is equivalent to r, provided each pair of measurements is considered twice (11). Liu and his colleagues have shown that the intramethod reliability coefficient ρ_X in a study with two measurements of each worker can be expressed as (12):

$$\rho_X = 1/(1 + \sigma_\alpha^2 / \sigma_r^2), \qquad \text{equation 4}$$

where σ_α^2 denotes the intraindividual variance and σ_r^2 the interindividual variance of the exposure variable measured. The ratio $\sigma_\alpha^2/\sigma_r^2$ is also called the variance ratio λ. Thus the reliability of a dual measurement survey is estimated by the intraclass coefficient r_I, which can be rewritten as the term $1/(1 + \lambda)$. The objective of intramethod surveys is to estimate the performance of a particular measurement method in relation to its reliability and exposure variability.

Measurement errors in methods for assessing back load

Reviews of epidemiologic studies on back disorders have demonstrated that job title is the exposure variable most frequently used. Among the epidemiologic studies that attempted to assess postural load on the back, a questionnaire approach was the most common (4, 6). In the past decade observational techniques have been increasingly used (6). These methods vary from pencil-and-paper techniques based upon multiple observations of workers at specific intervals throughout a representative period of work activities (13) to video-computerized systems for the real-time recording of trunk postures and movements (14). However, epidemiologic studies addressing the reliability of the measurement method applied are few (6).

Table 1 presents findings of intermethod reliability studies on various aspects of postural load (15—21).

Table 1. Findings of intermethod reliability studies on the exposure assessment of postural load on the back at the workplace. (na = not available)

Reference	Measure of exposure	Comparison	Subjects (N)	Reliability coefficient	Systematic difference[a]
Baty et al (15)	Duration of trunk flexion (> 15 degrees) in percentage of work time	Questionnaire versus observation	46	r = 0.3	5.5%
	Duration of trunk flexion (> 15 degrees) in percentage of work time	Observation versus inclinometer	7	r = 0.63*	−17.7%*
	Duration of standing & walking in percentage of work time	Questionnaire versus observation	46	r = 0.4*	−0.7%
Burdorf (16)	Duration of sitting in percentage of work time	Diary versus observation	81	r = 0.66*	4.3%
	Duration of trunk flexion and rotation (> 20 degrees) in percentage of work time	Diary versus observation	28	r = 0.41*	−24.4%*
Burdorf et al (17)	Duration of trunk flexion (> 20 degrees) in percentage of work time	Observation versus inclinometer	14	r_s = 0.62*	16.9%*
	Duration of trunk flexion (> 20 degrees) in percentage of work time	Observation versus inclinometer	16	r_s = 0.57*	0.2%
Burdorf & Laan (18)	Duration of trunk flexion (> 20 degrees) in percentage of work time	Questionnaire versus observation	35	r = 0.27	−18.0%*
	Duration of trunk flexion (> 20 degrees) in percentage of work time	Diary versus observation	35	r = 0.42*	−29.0%*
Dallner et al (19)	Duration of trunk flexion (> 60 degrees) in percentage of work time	Diary versus questionnaire	211	κ_ω = 0.36	na
	Duration of trunk rotation (> 45 degrees) in percentage of work time	Diary versus questionnaire	211	κ_ω = 0.63	na
	Duration of sitting in percentage of work time	Diary versus questionnaire	211	κ_ω = 0.76	na
Nordin et al (20)	Frequency of trunk flexion (> 72 degrees) during 1 h	Observation versus inclinometer	10	r_s = 0.99*	na
Wiktorin et al (21)	Duration of trunk flexion (20–60 degrees) in percentage of work time	Questionnaire versus inclinometer	97	r_s = 0.10	na
	Duration of sitting in percentage of work time	Questionnaire versus posimeter	97	r_s = 0.85*	na

[a] Estimated as the difference of the sample means of both methods.
* $P < 0.05$.

Although some studies refer to the validity of a specific measurement instrument, the correlations have been interpreted as intermethod reliability coefficients. Questionnaires are of limited use when postural load is assessed. Only duration of sitting during a normal shift was estimated with reasonable reliability and without systematic bias (16, 21). Therefore, assessing gross postural activity by questionnaire, defined as either duration of sitting or duration of standing or walking, may be appropriate. When a questionnaire is considered for this particular purpose, a diary kept on the subjects over several shifts can offer the advantage of collecting information on variation in exposure (16, 19). Postural load due to trunk posture is best assessed by direct observation or inclinometer measurements. Comparisons of both methods showed reliability coefficients of about 0.60 for duration of trunk flexion over 20 degrees (15, 17). One study mentioned an extremely strong correlation of 0.99 for the frequency of trunk flexions over 72 degrees (20). This extraordinary result may be due to a highly skewed exposure distribution with a few subjects with high exposure.

The occurrence of systematic bias (ie, $\alpha \neq 0$ or $\beta \neq 1$) in several reliability studies is problematic since it may lead to an overestimation or underestimation of the risk per unit of exposure. Systematic bias can result from a dependency of the relative bias on the exposure magnitude. Several authors have pointed at a discrepancy in the definition of the angular position of the trunk as another origin of bias (15, 17).

Intramethod reliability studies focusing on changes in postural load over time are few. Harber and his colleagues demonstrated considerable variability in trunk flexion both within and between supermarket checkers, whereby the coefficients of variation for within-worker variance tended to be higher (22). Another study presented variance ratios for trunk flexion and trunk rotation among five occupational title groups. After log-transformation of the exposure data, variance ratios were reported that varied from 0.2 to 7.1 (23). When normally distributed exposure parameters are assumed and this particular range of variance ratios is used, the associated reliability coefficients, expressed as intraclass coefficient r_1, would range from 0.12 to 0.83. A few studies have been published on intra- and interobserver agreement for observation methods. One study reported an interrater agreement of 86% for duration of trunk flexion and rotation (20 degrees) (24). Another study presented agreement of over 90% for the duration of trunk flexion (>15 degrees) within and between observers (25). In a study with the assessment of the same parameter of exposure, an interobserver agreement of 81% was reported (15). Recently, van Beek and his colleagues described interobserver reliabilities for trunk postures (assessed at four levels) of more than 80% agreement (26). These studies suggest that intraobserver and interobserver agreement are high when trained observers are used to collect data on exposure to postural load. Although the measures of agreement were not presented by estimates of reliability coefficients, it can reasonably be assumed that the measurement error of an observation method due to observer variability is small compared with the true variability in exposure due to workplace conditions (23).

Random measurement error and the attenuation of odds ratios

The information on the reliability of a particular measurement method can be used to evaluate the influence of random measurement error on the risk estimate in epidemiologic studies. In the case of musculoskeletal disorders cross-sectional studies are often published that present the odds ratio as a risk estimate. The amount of measurement error in a continuous exposure variable, such as percentage of work time with trunk flexion, can be expressed by the variation from either the "true" value or the same measurement repeated (11). The effect of random measurement error on the coefficient in a regression model is given by (8, 11):

$$\beta_0 = \beta_T \rho_X , \qquad \text{equation 5}$$

This equation pertains to linear regression models and logistic regression models. The equation states that the observed regression coefficient β_0 is weaker than the true regression coefficient β_T by the factor equal to the reliability coefficient ρ_X of an intramethod study. In intermethod studies the reliability coefficient is noted as $\rho_{X_1X_2}$. Thus the attenuation of risk estimates derived from logistic regression coefficients can be calculated simply by the equation (11, 27):

$$OR_0 = OR_T^{\rho_X} = OR_T^{\rho^2_{TX_1}}, \qquad \text{equation 6}$$

where OR_0 is the observed odds ratio per unit increment of exposure in the logistic regression model, OR_T is the unbiased true odds ratio, and ρ_X is the reliability coefficient estimated by the intraclass coefficient r_1. Figure 1 graphically illustrates the influence of four intraclass coefficients with values 1.0, 0.8, 0.5 and 0.1, respectively, on the true odds ratio. These coefficients correspond to variance ratios of 0, 0.25, 1 and 9, respectively, and reflect the results obtained in a study on the variability of exposure to trunk flexion and rotation (23). The figure shows the attenuation towards unity for all intraclass coefficients and a dramatic weakening of the odds ratio at intraclass coefficients of 0.5 and below. An intraclass coefficient of 0.5 corresponds to a validity coefficient of 0.7, under the strict assumptions of equation 2.

In general, this approach to estimating the attenuation in the odds ratio holds true for one exposure variable measured with random nondifferential error. In the specific situation of calculating odds ratios as risk estimates, an additional assumption is that the outcome variable of interest must be a rare disease. When more than one covariate in the logistic model is measured with error, then the estimates of any of the covariate effects can be influenced by measurement error (8, 28). Appropriate models taking into account multiple variables with random nondifferential measurement error have been discussed extensively by Rosner and his colleagues (29, 30). In these models a correction procedure for observed odds ratios can be complicated by the fact that both attenuation and overestimation may occur (28).

Optimizing measurement strategies

The previous section was restricted to the classical approach of a reliability study prior to the epidemiologic study that applies the same method of exposure measurement on a large scale. Such a study usually consists of an inter- or intramethod reproducibility survey of a limited number of measurements to derive a reliability coefficient. This information is then used to evaluate the influence of random error in single measurements of exposure in the epidemiologic study. Examples concerning postural load on the back are the surveys presented in table 1. An important drawback of this approach is that one measurement of exposure to postural load may not be sufficient to estimate accurately the true exposure of each subject. The few studies on patterns of exposure to back load have demonstrated considerable variability in trunk flexion both within and between workers (22, 23). This problem can be addressed by choosing the appropriate number of measurements per subject to distinguish one worker from another and, at the same time, assuming that the true exposure is estimated unbiased by the average value of a number of measurements. In recent publications this principle has been used to review effects of random measurement error in various study designs (5, 11, 12, 31).

Liu and his colleagues have presented an equation to calculate the reliability coefficient of the average value or, alternatively, the number of repeats required given a particular reliability coefficient. The reliability coefficient of the average exposure, $\rho_{\bar{x}}$, depends on the number of repeats per subject (12):

$$\rho_{\bar{x}} = 1/[1+ (\sigma_\alpha^2/k\sigma_r^2)] = \sigma_r^2/(\sigma_r^2 + \sigma_\alpha^2/k), \qquad \text{equation 7}$$

where k denotes the number of repeats per subject. This equation demonstrates that the reliability is determined by the number of measurements per subject and the ratio of the intraindividual variance, σ_α^2, and the interindividual variance, σ_r^2. Figure 2 graphically illustrates the relation between $\rho_{\bar{x}}$ and k at four levels of the variance ratio λ with values 0.25, 1, 4 and 9, respectively. These variance ratios correspond with intramethod reliability coefficients r_I of 0.8, 0.5, 0.2 and 0.1, respectively. Figure 2 demonstrates that, in studies in which k is very large or the intraindividual variance is considerably smaller than

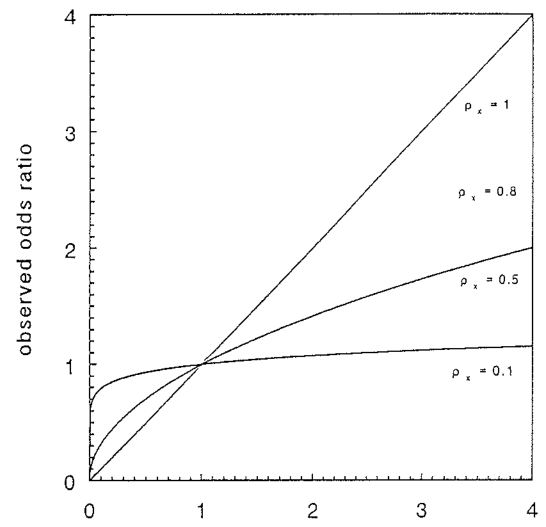

Figure 1. Attenuation of the true odds ratio, estimated by logistic regression, at different levels of reliability of the measurement method.

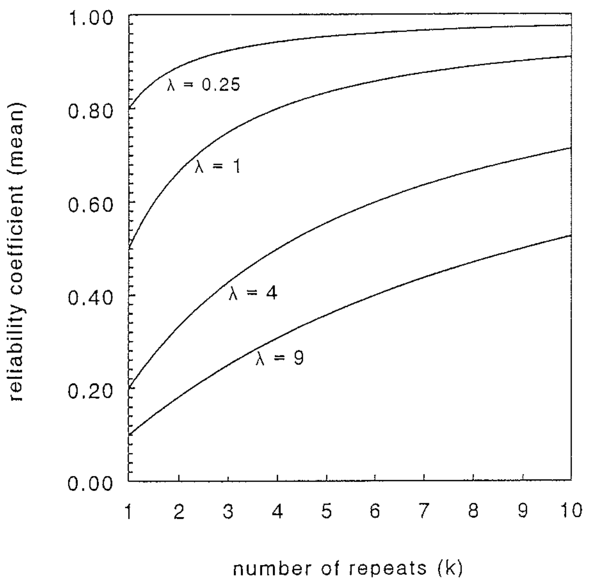

Figure 2. Relation between the number of repeat measurements per subject and the reliability of the subject's average exposure at different magnitudes of the variance ratio.

the interindividual variance, the attenuation of the odds ratio is small. The attenuation of the odds ratio can be calculated with equation 6 by substituting the reliability coefficient of the average of several measurements, $\rho_{\bar{X}}$, for the reliability coefficient of the single measurement, ρ_X.

Equation 7 can be rewritten as a function of the number of repeat measurements:

$$k = [\rho_{\bar{X}}/(1-\rho_{\bar{X}})] \times (\sigma_\alpha^2/\sigma_r^2) . \qquad \text{equation 8}$$

This equation yields an answer to the question of how many measurements are required to achieve a specified degree of reliability. For example, to obtain $\rho_{\bar{X}} = 0.8$ in study populations with variance ratios of 0.25, 1, 4, and 9, the number of repeat measurements required is 1, 4, 16, and 36, respectively. Thus equations 7 and 8 are instrumental when one is designing an appropriate measurement strategy for large epidemiologic studies. Information on the variance ratio can be obtained from an exposure-oriented survey prior to the start of the epidemiologic study. In such a survey repeat measurements should be collected randomly in time on a random sample of workers.

Along the same lines the optimum allocation for the number of subjects (sample size) and repeat measurements per subject can be evaluated. The familiar formula for determining the sample size when cases and referents in a 1:1 ratio are to be compared on a continuous exposure variable is (31):

$$n_T = [2(t_\alpha + t_\beta)^2\sigma_p^2] / [\mu_1 - \mu_0)^2]. \qquad \text{equation 9}$$

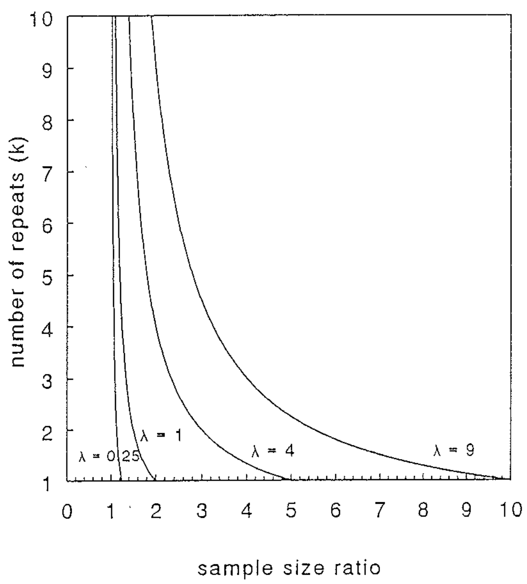

Figure 3. Relation between the sample size ratio n_k/n_T and the number of repeat measurements per subject in a case-referent study at different magnitudes of the variance ratio.

The numerator contains the t-statistics for the desired level of significance, α, and the power, 1-β, and the pooled variance of exposure. The denominator denotes the difference in exposure to be detected. When the intraindividual variance is equal to zero among the cases and the referents, the pooled variance is estimated by:

$$\sigma_p^2 = (\sigma_{r1}^2 + \sigma_{r0}^2) / 2. \qquad \text{equation 10}$$

Substituted into equation 9, the sample size n_T is calculated. The assumption is that the true exposure of each subject is estimated unbiased by a single measurement per subject. If the intraindividual variance is greater than zero, then the unbiased exposure for each subject can be estimated by averaging k measurements for each subject. In the situation of equal variance ratios for the two groups, the pooled variance can be expressed as (12):

$$\sigma_{p1}^2 = [\sigma_{r1}^2 + (\sigma_{\alpha 1}^2 / k) + \sigma_{r0}^2 + (\sigma_{\alpha 0}^2 / k)] / 2 \qquad \text{equation 11}$$

and, consequently, the sample size n_k is equal to:

$$n_k = n_T (1 + \lambda/k). \qquad \text{equation 12}$$

This formula shows that the required sample size must be increased by $(1 + \lambda/k)$ due to intraindividual variance. Equation 7 shows that $1/\rho_{\bar{X}}$ is equal to the term $(1 + \lambda/k)$. Hence a simple equation for the required sample size is derived (11):

$$n_k = n_T / \rho_{\bar{X}}. \qquad \text{equation 13}$$

The sample size calculated under the assumption of intraindividual variance equal to zero (n_T) need only be divided by the reliability coefficient of the average of k repeat measurements to obtain the actual sample size required (n_k).

Figure 3 shows the relation between the sample size ratio, expressed as n_k/n_T, and the number of repeat measurements per subject. This figure demonstrates that a larger sample size can be counterbalanced by increasing the number of repeat measurements. The efficiency of either decision depends on the actual value of the variance ratio and the study costs. Suppose the researcher is interested in the risk for low-back pain associated with trunk flexion, expressed by the average percentage of work time with trunk flexion over 20 degrees. The minimum difference of significant importance is set at 5% and the standard deviation of exposure to trunk flexion among workers with low-back pain and those without low-back pain is about 10%. A two-sided α of 5% ($t_\alpha = 1.96$) and a 90% power ($t_\beta = 1.28$) are defined. According to equation 8, a sample size n_T of 84 is obtained. In groups with a variance ratio of 0.25 the required sample size is 105 with one measurement per worker and 95 with two measurements per worker. The extra costs of 95 measurements are usually less than offset by the decrease of the sample size with 10 workers. However, two measurements per subject in groups with a variance ratio of four reduce the sample size by about 40% (from 420 to 252). If the cost of one additional measurement per worker is smaller than the cost of monitoring one additional worker, more measurements per worker could be an attractive option.

Discussion

The importance of validly assessing exposure in epidemiologic studies has been stressed by many authors (5, 8, 11, 12, 27—30). Several studies have shown that assessing exposure to postural load on the back is subject to systematic bias and random measurement error (21—23). Both can lead to spurious conclusions about the relation between postural load and the occurrence of back disorders. Therefore, exposure assessment should be a major topic in the design of epidemiologic studies on back disorders (4).

The choice of an appropriate strategy for assessing postural load on the back at the workplace is influenced, for example, by the type of epidemiologic study, the amount and detail of data required, the validity of the measurement method, and the variability of postural load within and between workers. Three methods commonly used in musculoskeletal epidemiology are the questionnaire, the diary, and the observation technique. Questionnaires have been applied to collect information on past and present exposure to various aspects of back load, such as duration of daily periods of sitting and postures maintained with a twisted trunk (6). Diaries offer the advantage of minor recall bias since they focus on the assessment of present exposure. Their application may take account of infrequent exposure and the variability of exposure (19). Observation techniques deliver the most-detailed information on exposure patterns but are expensive and time-consuming, both of which hamper their use in epidemiologic studies. Although these considerations may guide the researcher towards a particular measurement technique, the decision should also focus on measurement error associated with the application of various measurement methods and its consequences for the risk estimate in the epidemiologic study planned.

The reliability of a measurement method can be evaluated in intermethod reproducibility studies. The results of intermethod studies, as presented in table 1, show that questionnaires and diaries often lack accuracy. The systematic bias can be estimated as the difference between the sample means of methods 1 and 2. In theory, individual exposure assessments can be adjusted when this bias is constant in time and independent of the exposure and disease status of the subjects. Generally, this will not be the case, as the magnitude and direction of systematic bias (at the individual level) is difficult to predict. Therefore, these studies suggest that questionnaires and diaries may be warranted only when gross postural activities such as sitting, standing and walking are assessed (18, 19). Postural load due to trunk posture is best assessed with observation techniques or direct instrumentation.

Another approach for assessing the reliability of a measurement method is an intramethod reproducibility survey. Reliability coefficients in these surveys include the effects of instrumental features and exposure variability. The latter source of imperfect exposure assessment often surpasses the effect of instrumental characteristics (23). Again, systematic bias can introduce insuperable problems. When systematic bias is absent, the reliability coefficient yields information crucial to the design of measurement strategies. An exposure-oriented survey could combine the inter- and intramethod approach by conducting measurements with two methods with each pair of measurements repeated once. This procedure allows both the occurrence of systematic bias (instrumental characteristics) and random error (precision and exposure variability) to be estimated.

This article does not address the important question of whether a higher frequency or a longer duration of measurement is to be preferred. It is obvious that a longer duration of measurement will generally decrease the exposure variability. The answer to this question requires detailed information on the temporal variation of back load, which depends on individual behavior and the particular characteristics of the job and tasks involved. Currently, no general guidelines are available to evaluate the trade-off between repeated and prolonged measurements in a group of workers. Computer simulations on detailed quantitative exposure distributions, such as Monte Carlo techniques, have to be performed to explore the effect of different sample sizes and duration on the estimation of the mean exposure of individuals.

The equations presented in this article enable the influence of random error on the odds ratio in cross-sectional studies to be evaluated. In principle, these equations only apply to situations with a rare disease. In cross-sectional studies on back disorders this rare disease assumption is seldom met, and, consequently, odds ratios overestimate risk ratios (32). Hence it does not make sense to correct an odds ratio for the amount of random measurement error if the adjusted "true" odds ratio is not a valid risk estimate. Alternative methods to correct for random measurement error have been developed that are not restricted by the rare disease assumption, based on probit regression (33) and discriminant analysis (34). However, despite methodological drawbacks in the familiar procedure of odds ratios derived from logistic regression, the equations presented can guide towards an appropriate measurement strategy.

An important feature of the intramethod reliability coefficient is its use in evaluating various design options regarding the sample size of the study and the number of measurements per worker. With the use of various criteria for optimizing the study, traditional power calculations can be applied to consider the best measurement strategy (35, 36). This article described straightforward procedures with which to optimize the measurement strategy through the minimum attenuation of the odds ratio and maximum discriminatory power of the study. A core element in these procedures is an unbiased estimate of the intramethod reliability coefficient or variance ratio. It should be remembered that the reliability coefficient depends on the exposure distributions among the workers monitored, and this dependability hampers the application from one population to another (5). In order to obtain a reliability coefficient with reasonable precision, the sample size in an exposure-oriented survey must be sufficiently large. A disadvantage of the application of an intramethod reliability coefficient is that it requires a fully random measurement strategy whereby measurements are randomly sampled within worker's exposure experiences and workers are randomly sampled within the occupational population. In particular situations other measurement strategies may be more appropriate, for example, sampling in specific exposure strata or adopting a group-based exposure assessment strategy rather than assessment at the individual level.

In epidemiologic studies on back disorders the procedures described may intuitively appeal through their focus on the role of repeat measurements. Currently, in epidemiologic studies on back disorders, fairly simple parameters of exposure to postural load are being adopted, and the characterization of exposure is usually limited to the worker's average exposure. Valid assessment of a worker's average exposure requires workers to be monitored repeatedly or, alternatively, monitored with increasing averaging time of measurement. The appropriate number of measurements per subject or the optimum measurement duration is relative to the load variation in the populations under study. This paper provides a quantitative framework to evaluate various design options in measurement strategies in relation to the precision of the odds ratio.

Acknowledgments

The author thanks Professor R van Strik for his vital contribution to the understanding and application of the statistical equations in this paper.

References

1. University of Michigan, Center for Ergonomics. Three dimensional static strength prediction program, version 2.0. Ann Arbor, MI: Center for Ergonomics, University of Michigan, 1993.
2. Högfors C, Sigholm G, Herberts P. Biomechanical model of the human shoulder joint I: elements. J Biomech 1987;20:157—66.
3. Snijders CJ, Hoek van Dijke GA, Roosch ER. A biomechanical model for the analysis of the cervical spine in static postures. J Biomech 1991;24:783—92.
4. Hagberg M. Exposure variables in ergonomic epidemiology. Am J Ind Med 1992;21:91—100.
5. Armstrong BK, White E, Saracci S. Principles of exposure measurement in epidemiology. Oxford: Oxford University Press, 1992.
6. Burdorf A. Exposure assessment of risk factors for disorders of the back in occupational epidemiology. Scand J Work Environ Health 1992;18:1—9.
7. Last JM. A dictionary of epidemiology. New York, NY: Oxford University Press, 1983.
8. Thomas D, Stram D, Dwyer J. Exposure measurement error: influence on exposure-disease relationships and methods of correction. Annu Rev Publ Health 1993;14: 69—93.
9. Carmines EG, Zeller RA. Reliability and validity assessment. Beverly Hills, CA: Sage Publications, 1979.
10. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;1:307—10.
11. De Klerk NH, English DR, Armstrong BK. A review of the effects of random measurement error on relative risk estimates in epidemiological studies. Int J Epidemiol 1989;18:705—12.
12. Liu K, Stamler J, Dyer A, McKeever J, McKeever P. Statistical methods to assess and minimize the role of intra-individual variability in obscuring the relationship between dietary lipids and serum cholesterol. J Chronic Dis 1978;31:399—418.
13. Karhu O, Kansi P, Kuorinka I. Correcting working postures in industry: a practical method for analysis. Appl Ergon 1977;8:199—201.
14. Punnett L, Fine LJ, Keyserling WM, Herrin GD, Chaffin DB. Back disorders and nonneutral trunk postures of automobile assembly workers. Scand J Work Environ Health 1991;17:337—46.
15. Baty D, Buckle PN, Stubbs DA. Posture recording by direct observation, questionnaire assessment and instrumentation: a comparison based on a recent field study. In: Corlett N, Wilson J, ed. The ergonomics of working postures. London: Taylor and Francis, 1986:283—92.
16. Burdorf A. Vergelijking van een dagboek en observaties voor het schatten van houding en beweging van werkers [Comparison of a diary and an observation method for assessing basic postures of workers]. Rotterdam: Institute of Public Health, 1992.
17. Burdorf A, Derksen J, Naaktgeboren B, Van Riel M. Measurement of trunk bending during work by direct observation and continuous measurement. Appl Ergon 1992;23:263—7.
18. Burdorf A, Laan J. Comparison of methods for the assessment of postural load on the back. Scand J Work Environ Health 1991;17:425—9.
19. Dallner M, Gamberale F, Winkel J. Validation of a questionnaire for estimation of physical load in relation to a workplace diary. In: Quéinnec Y, Daniellou F, editors. Designing for everyone. London: Taylor and Francis, 1991:233—5.

20. Nordin M, Hultman G, Philipsson R, Ortelius A, Andersson GBJ. Dynamic measurements of trunk movements during work tasks. In: Corlett N, Wilson J, editors. The ergonomics of working postures. London: Taylor and Francis, 1986:74—81.
21. Wiktorin C, Karlqvist L, Winkel J. Validity of self-reported exposures to work postures and manual materials handling. Scand J Work Environ Health 1993;19:208—14.
22. Harber P, Bloswick D, Peña L, Beck J, Lee J, Baker D. The ergonomic challenge of repetitive motion with varying ergonomic stresses. J Occup Med 1992;34:518—28.
23. Burdorf A. Sources of variance in exposure to postural load on the back in occupational groups. Scand J Work Environ Health 1992;18:361—7.
24. Kivi P, Matilla M. Analysis and improvement in the building industry: application of the computerised OWAS method. Appl Ergon 1991;22:43—8.
25. Foreman TK, Davies JC, Troup JDG. A posture and activity classification system using a microcomputer. Int J Ind Ergon 1988;2:285—9.
26. van der Beek AJ, Van Gaalen LC, Frings-Dresen MHW. Working postures and activities of lorry drivers: a reliability study of on-site observation and recording on a pocket computer. Appl Ergon 1992;23:331—6.
27. Wu ML, Whittemore AS, Jung DL. Errors in reported dietary intakes: I. short-term recall. Am J Epidemiol 1986;124:826—35.
28. Kupper LL. Effects of the use of unreliable surrogate variables on the validity of epidemiologic research studies. Am J Epidemiol 1984;120:643—8.
29. Rosner B, Spiegelman D, Willett WC. Correction of logistic regression relative risk estimates and confidence intervals for measurement error: the case of multiple covariates measured with error. Am J Epidemiol 1990;132:734—45.
30. Rosner B, Spiegelman D, Willett WC. Correction of logistic regression relative risk estimates and confidence intervals for random within-person measurement error. Am J Epidemiol 1992;136: 1400—13.
31. McKeown-Eyssen GE, Thomas DC. Sample size determination in case-control studies: the influence of the distribution of exposure. J Chronic Dis 1985;38:559—68.
32. Axelson O, Fredrikson M, Ekberg K. A comment on the implications of using odds ratios or prevalence ratios in cross-sectional studies. Rev Epidemiol Sante Publique 1992;40 suppl 1:72.
33. Carroll RJ, Spiegelman CH, Gordan Lan KK, Bailey KT, Abbott RD. On errors-in-variables for binary regression models. Biometrika 1984;71:19—25.
34. Armstrong BG, Whittemore AS, Howe GR. Analysis of case-control data with covariate measurement error: application to diet and colon cancer. Stat Med 1989;8:1151—63.
35. Greenland S. On sample size and power calculations for studies using confidence intervals. Am J Epidemiol 1988;128: 231—6.
36. Spiegelman D, Gray R. Cost-efficient study designs for binary response data with Gaussian covariate measurement error. Biometrics 1991;47:851—69.

Received for publication: 2 March 1994