Ann. occup. Hyg., Vol. 46, No. 2, pp. 219–227, 2002
Published by Oxford University Press
DOI: 10.1093/annhyg/mef028

Sample Size Considerations for Studies of Intervention Efficacy in the Occupational Setting

DEANN LAZOVICH[1]*, DAVID M. MURRAY[2], LISA M. BROSSEAU[3], DAVID L. PARKER[3,4], F. THOMAS MILTON[5] and SIOBHAN K. DUGAN[3,4]

[1]Division of Epidemiology, University of Minnesota, 1300 South 2nd Street, 300, Minneapolis, MN 55454, USA; [2]Department of Psychology, 202 Psychology Building, The University of Memphis, Memphis, TN 38152-3230, USA; [3]Division of Environmental and Occupational Health, University of Minnesota, Box 807 Mayo, 420 Delaware Street South East, Minneapolis, MN 55455, USA; [4]Division of Chronic Disease and Environmental Epidemiology, Minnesota Department of Health, 717 Delaware Street South East, Minneapolis, MN 55440, USA; [5]Department of Wood and Paper Science, University of Minnesota, 2004 Folwell Avenue, St Paul, MN 55108, USA

Received 11 October 2000; in final form 18 July 2001

Objective. Due to a shared environment and similarities among workers within a worksite, the strongest analytical design to evaluate the efficacy of an intervention to reduce occupational health or safety hazards is to randomly assign worksites, not workers, to the intervention and comparison conditions. Statistical methods are well described for estimating the sample size when the unit of assignment is a group but these methods have not been applied in the evaluation of occupational health and safety interventions. We review and apply the statistical methods for group-randomized trials in planning a study to evaluate the effectiveness of technical/behavioral interventions to reduce wood dust levels among small woodworking businesses.

Methods. We conducted a pilot study in five small woodworking businesses to estimate variance components between and within worksites and between and within workers. In each worksite, 8 h time-weighted dust concentrations were obtained for each production employee on between two and five occasions. With these data, we estimated the parameters necessary to calculate the percent change in dust concentrations that we could detect (α = 0.05, power = 80%) for a range of worksites per condition, workers per worksite and repeat measurements per worker.

Results. The mean wood dust concentration across woodworking businesses was 4.53 mg/m^3. The measure of similarity among workers within a woodworking business was large (intraclass correlation = 0.5086). Repeated measurements within a worker were weakly correlated (*r* = 0.1927) while repeated measurements within a worksite were strongly correlated (*r* = 0.8925). The dominant factor in the sample size calculation was the number of worksites per condition, with the number of workers per worksite playing a lesser role. We also observed that increasing the number of repeat measurements per person had little benefit given the low within-worker correlation in our data. We found that 30 worksites per condition and 10 workers per worksite would give us 80% power to detect a reduction of ~30% in wood dust levels (α = 0.05).

Conclusions. Our results demonstrate the application of the group-randomized trials methodology to evaluate interventions to reduce occupational hazards. The methodology is widely applicable and not limited to the context of wood dust reduction.

Keywords: community trials; interventions; occupational health and safety; statistical methods

INTRODUCTION

Lack of rigorous study design has hampered the assessment of interventions to reduce job-related health risks in the occupational setting (Goldenhar and Schulte, 1994). Studies have typically lacked randomization to intervention and comparison conditions and have relied on small numbers of study

*Author to whom correspondence should be addressed.

participants, usually selected from a single worksite. Due to the shared environment, any study that attempts to evaluate an intervention by assigning workers within a worksite to intervention and comparison conditions may be compromised by contamination resulting from interaction between workers assigned to different conditions. Making matters worse, workers within a worksite often share common selection factors or exposures, so that observations taken on workers from the same worksite are likely to be correlated (Kish, 1995). Any such correlation will violate the independence of errors assumption that underlies the familiar analytical methods based on the General Linear Model (Searle, 1977). The solution to these problems is to employ a group-randomized trial design and analytical methods appropriate to the correlation expected in the data (Murray, 1998). Although such methods have been used widely in other areas of public health, including school-, worksite- and community-based health promotion projects, these studies have primarily sought to change individual lifestyle behaviors, like smoking or diet (Lasater *et al.*, 1997; Ockene *et al.*, 1997; Resnicow and Robinson, 1997), rather than work conditions that adversely affect health or safety.

The purpose of this paper is to introduce group-randomized trial methods to researchers working in the fields of environmental and occupational health. This task is made easier by the parallels between the analytical methods often used in group-randomized trials and those used recently to study exposure variability in cohort studies (Kromhout *et al.*, 1993, 1994, 1995, 1996; Kromhout and Heederik, 1995; Rappaport *et al.*, 1995, 1999; Tornero-Velez *et al.*, 1997a,b). These studies employed mixed model regression to examine between- and within-worker variability as they affect the precision of the exposure measures or as they affect the contrast in exposure between groups. These studies defined groups in different ways, but usually as a function of job title or type of work. One of the goals in these studies has been to identify groupings that maximize the differences in exposure, while maintaining high precision, so as to optimize the power to study dose–response relationships.

While we also use mixed model regression methods in group-randomized trials, the goals and study designs are quite different from those used in such cohort studies. In a group-randomized trial, whole worksites are randomized to intervention and comparison conditions and workers within those worksites are measured to assess the impact of the intervention. The group is the worksite, rather than a subset of the workers defined by job title or similar scheme. The goal is to minimize differences among groups within each study condition so as to optimize power to study intervention effects between conditions. Thus, the focus in a group-randomized trial is not variation in the exposure, but rather variation in the outcome.

Wood dust is a known carcinogen for cancer of the nasal cavity (IARC, 1995) and is also strongly associated with other respiratory illnesses (e.g. asthma). It is estimated that more than 1 200 000 workers are employed in the logging, woodworking and carpentry trades; 64% of these are employed in the manufacture of wood products (National Institute for Occupational Safety and Health, 1988). With this health concern in mind, we designed a group-randomized trial to evaluate the efficacy of technical assistance and worker education to reduce wood dust levels among small businesses (5–25 employees) that manufacture wood products in Minnesota. Because we were interested in changing not only worker behavior to lower dust, but also the work environment, thereby exposing all workers within a worksite to the intervention, we anticipated the need for a group-randomized trial. To determine the number of woodworking businesses that would be needed for statistical power to detect a specified reduction in wood dust between the intervention and comparison conditions, we needed estimates of the variance for worker exposure to inhalable dust from day to day, among workers within a woodworking business and between workers in different woodworking businesses. To obtain those estimates, we carried out a pilot study in five woodworking businesses (Brosseau *et al.*, 2001). In this paper we summarize the design and statistical considerations involved in a group-randomized trial and show how we used the data from the pilot study to estimate the sample size for the efficacy trial.

MATERIALS AND METHODS

Description of pilot study

A detailed description of the pilot study is provided elsewhere (Brosseau *et al.*, 2001); we briefly review the methods here. We invited a convenience sample of five small woodworking businesses in the metropolitan Twin Cities area to participate in our pilot study. These businesses were engaged in the manufacture of wood fixtures, cabinets or millwork and they employed on average eight workers (range 5–11). After receiving permission from the owner, we met with the workers to explain the purpose of the pilot study and the procedures for collecting dust measurements. Overall, 90% of the workers consented to participate in data collection.

Dust sampling was carried out with personal inhalable dust samplers (IOM sampling, SKC Inc., Eighty-Four PA), loaded with 5 μm PVC filters and personal sampling pumps (SKC Aircheck, SKC Inc), operated at 2 l/min. To the extent possible, we conducted our measurements in each worksite over 2 weeks, including each day of the week (e.g. Monday, Wednesday and Friday in the first week and Tuesday and Thursday in the second week), for a total of 5 days per worksite. On average, each worker within a woodworking business was sampled four times (range 2–5) over the course of 2 weeks.

At the conclusion of each day of sampling, filters and cassettes were desiccated and then weighed. Individual dust concentrations were measured for the total work time (usually 8 h) and concentrations (mg/m^3) were determined by dividing total cassette weight gain (mg), adjusted by averaged field weight gains (using blanks filters), by the total volume of sampled air (m^3) as:

$$\text{concentration (mg/m}^3\text{)} = [(W_2 - W_1) - (B_2 - B_1)] / V \times 1000$$

where W_1 is the initial weight of filter and cassette before sampling (mg), W_2 is the post-sampling weight of filter and cassette (mg), B_1 is the mean tare weight of blank filters and cassettes (mg), B_2 is the mean post-sampling weight of blank filters and cassettes (mg) and V is the volume of air sampled (l) [average flow in l/min × time sampled in min].

A final dust concentration for the worksite was calculated by taking the mean of the worker concentrations across all days of sampling. Prior to analysis, we applied a log transform on worksite dust concentration values to correct for a strong positive skew.

Statistical overview

As noted in the introduction, observations taken on workers in the same worksite tend to be positively correlated, due to common job selection, experience and mutual interaction. This positive intraclass correlation (ICC) is common to observations taken on the members of any identifiable group (Kish, 1995). In this case, it reflects an extra source of random variation due to the worksite, above and beyond random variation due to the workers and other sources. Because woodworking businesses are nested within study conditions in a group-randomized trial, the intervention effect must be assessed against the variation between worksites, rather than the variation within worksites (Cornfield, 1978). Unfortunately, the precision available to estimate the variation between worksites is expected to be less than for the variation within worksites, given the limited number of woodworking businesses relative to the number of workers. Without proper planning, these factors can combine to make it almost impossible to detect intervention effects in an otherwise well-designed and properly executed study (Murray, 1998).

The most common analysis for a nested cross-sectional design with two time intervals (pre-test and post-test) is a time × condition mixed model analysis of variance (ANOVA) or covariance (ANCOVA). This analysis is based on the General Linear Mixed Model (Harville, 1977; Laird and Ware, 1982; Laird *et al.*, 1987; Donner, 1984, 1985; Stiratelli *et al.*, 1984; Ware, 1985), which extends the General Linear Model to allow multiple sources of random variation. The mixed model ANOVA and ANCOVA have been widely recommended for this design because they carry the nominal Type I and II error rates across a variety of conditions common to group-randomized trials (Zucker, 1990; Koepsell *et al.*, 1991; Feldman and McKinlay, 1994; Murray and Wolfinger, 1994; Murray *et al.*, 1996; Hannan and Murray, 1996). More generally, mixed models are appropriate whenever the data includes multiple sources of random variation. Such studies are sometimes described as having nested, hierarchical or multi-level structures. Designs that have multiple sources of random variation include designs that have repeat measurements on the same participants, designs that employ cluster sampling and group-randomized trials.

Statistical notation

We follow the notational scheme of Murray (1998) to facilitate cross-references to that standard text for the design and analysis of group-randomized trials. Greek symbols are used when their meaning is well established, such as the mean (μ), intervention effect (Δ), variance (σ^2) and residual error (ε), and for algebraic functions, such as summation (Σ). The expression σ_e^2 is used to refer to the residual error variance. Similarly, the expression σ_Δ^2 is used to refer to the variance of the intervention effect.

Fixed and random effects are represented by an upper case first letter of the variable name while the number of levels of each effect is represented by a lower case first letter of the variable name. As such, Member, $\mathbf{M}_i$ ($i = 1 \ldots m$), identifies the unit of observation. Group, $\mathbf{G}_k$ ($k = 1 \ldots g$), identifies the unit of assignment. Time, T_j ($j = 1 \ldots t$), identifies the survey (e.g. pre-test, post-test). Repeat, R_p ($p = 1 \ldots r$) refers to the repeat measurements taken on a member in any given survey. Condition, C_l ($l = 1 \ldots c$), identifies the study conditions. Effects are fixed when all levels of interest are included in the study and inferences are limited to those levels (e.g. time, condition); effects are random when the levels included represent some larger population of levels and inferences are to be made on that larger population (e.g. group, member).

All random effects are in bold type, while fixed effects are in plain type. Because fixed and random effects are distinguished in this manner, all expected mean squares are presented in terms of variance components. For example, the variation among the levels of condition is defined as σ_c^2 instead of

$$\sum_{i=1}^{c} C_i^2 / (c-1)$$

Table 1. Expected mean squares for the adjusted analysis of data from a nested cross-sectional pre-test–post-test comparison group design

Source	df	E(MS)	MS
Condition	$c-1$	$\sigma^2_e + r\sigma^2_{m:tg:c} + tmr\sigma^2_{g:c} + gtmr\sigma^2_c$	MS_c
Group: C	$c(g-1)$	$\sigma^2_e + r\sigma^2_{m:tg:c} + tmr\sigma^2_{g:c}$	$MS_{g:c}$
Time	$t-1$	$\sigma^2_e + r\sigma^2_{m:tg:c} + mr\sigma^2_{tg:c} + cgmr\sigma^2_t$	MS_t
TC	$(t-1)(c-1)$	$\sigma^2_e + r\sigma^2_{m:tg:c} + mr\sigma^2_{tg:c} + gmr\sigma^2_{tc}$	MS_{tc}
TG:C	$(t-1)c(g-1)$	$\sigma^2_e + r\sigma^2_{m:tg:c} + mr\sigma^2_{tg:c}$	$MS_{tg:c}$
Member: TG:C	$tgc(m-1)$	$\sigma^2_e + r\sigma^2_{m:tg:c}$	$MS_{m:tg:c}$
Repeat: M:TG:C	$tgcm(r-1)$	σ^2_e	MS_e

Equations

Given a nested cross-sectional pre-test–post-test design with multiple observations on each worker measured in each survey, the model for the time × condition ANCOVA is:

$$Y_{pi:jk:l} = \mu + C_l + T_j + TC_{jl} + \sum_{o=1}^{x} \beta_o (X_{oip:jk:l} - \bar{X}_{o.....}) \quad (1)$$

$$+ \mathbf{G}_{k:l} + \mathbf{TG}_{jk:l} + \mathbf{M}_{i:jk:l} + \boldsymbol{\varepsilon}_{pi:jk:l}$$

Here the colons denote nesting. The observed value ($Y_{pi:jk:l}$) for the pth repeat measurement on the ith member nested within the kth group and lth condition and observed at the jth time is expressed as a function of the grand mean, μ, the effect of the lth condition, C_l, the effect of the jth time, T_j, the joint effect of the lth condition and the jth time, TC_{jl}, the realized value of the kth group, $\mathbf{G}_{k:l} \sim N(0,\sigma^2_{g:c})$, the realized value of the combination of the kth group and jth time, $\mathbf{TG}_{jk:l} \sim N(0,\sigma^2_{tg:c})$, and the realized value of the ith member, $\mathbf{M}_{i:jk:l} \sim N(0,\sigma^2_{m:tg:c})$. For each covariate, the portion of $Y_{pi:jk:l}$ that is explained by the difference between the observed value and sample mean on the covariate, ($X_{opi:jk:l} - \bar{X}_{o.....}$), is attributed to the covariate. Any difference between this predicted value and the observed value is allocated to the residual error, $\boldsymbol{\varepsilon}_{pi:jk:l} \sim N(0,\sigma^2_e)$, including any variation among the repeat measurements on each person in a single pre-test or post-test survey. (In papers published previously on the variability of exposure in cohort studies, what we term $\sigma^2_{m:tg:c}$ has often been labeled the between-subject variance and what we term σ^2_e has often been labeled the within-person variance.)

The intervention effect for this analysis is the adjusted net difference between the pre-test and post-test means in dust concentration in the intervention (I) and comparison (C) conditions:

$$\Delta = (\bar{Y}_{\cdot 2\cdot I} - \bar{Y}_{\cdot 1\cdot I}) - (\bar{Y}_{\cdot 2\cdot C} - \bar{Y}_{\cdot 1\cdot C}) \quad (2)$$

$$- \sum_{o=1}^{x} \beta_0 [(\bar{X}_{o\cdot\cdot 2\cdot I} - \bar{X}_{o\cdot\cdot 1\cdot I}) - (\bar{X}_{o\cdot\cdot 2\cdot C} - \bar{X}_{o\cdot\cdot 1\cdot C})]$$

The expected mean squares for this analysis are shown in Table 1. The null hypothesis of no intervention effect is assessed as $MS_{tc}/MS_{tg:c}$.

The variance of this intervention effect is:

$$\sigma^2_\Delta = 2 \times 2\left(\frac{MS_{tg:c}}{gmr}\right) = 2 \times 2\left(\frac{\sigma^2_e + r\sigma^2_{m:tg:c} + mr\sigma^2_{tg:c}}{gmr}\right) \quad (3)$$

There are 2 × 2 time × group means that define the intervention effect, r repeat measurements on each of m members in each time × group survey and g groups in each condition.

Here, σ^2_e is the within-member component of variance and $\sigma^2_{m:tg:c}$ is the between-member component of variance; together, these two components comprise the total variation in the dependent variable Y that is attributable to members: $\sigma^2_m = \sigma^2_e + \sigma^2_{m:tg:c}$. In the same way, $\sigma^2_{tg:c}$ is the within-group component of variance and $\sigma^2_{g:c}$ is the between-group component of variance; together, these two components comprise the total variation in the dependent variable Y that is attributable to groups: $\sigma^2_g = \sigma^2_{g:c} + \sigma^2_{tg:c}$. And because the variation attributable to members and groups together comprise the total random variation in the dependent variable Y, it follows that $\sigma^2_y = \sigma^2_e + \sigma^2_{m:tg:c} + \sigma^2_{g:c} + \sigma^2_{tg:c}$.

The between member component of variance as a fraction of the total variation attributable to members defines the within-member correlation over repeat observations, $r_{yy(m)}$:

$$r_{yy(m)} = \frac{\sigma^2_{m:tg:c}}{\sigma^2_e + \sigma^2_{m:tg:c}} \quad (4)$$

It follows that $\sigma^2_{m:tg:c} = \sigma^2_m(r_{yy(m)})$ and $\sigma^2_e = \sigma^2_m(1 - r_{yy(m)})$. Similarly, the between group component of variance as a fraction of the total variation attributable to groups defines the within-group correlation over repeat observations, $r_{yy(g)}$:

$$r_{yy(g)} = \frac{\sigma^2_{g:c}}{\sigma^2_{g:c} + \sigma^2_{tg:c}} \quad (5)$$

It follows that $\sigma^2_{g:c} = \sigma^2_g\,(r_{yy(g)})$ and $\sigma^2_{tg:c} = \sigma^2_g\,(1 - r_{yy(g)})$.

Finally, note that the total variation in the dependent variable *Y* that is attributable to groups, as a fraction of the total variation in *Y*, defines the classic intraclass correlation in a group-randomized trial that indexes the correlation among members nested within groups, which are in turn nested within conditions:

$$\mathrm{ICC}_{m:g:c} = \frac{\sigma^2_g}{\sigma^2_m + \sigma^2_g} \qquad (6)$$

It follows that $\sigma^2_m = \sigma^2_y\,(1 - \mathrm{ICC}_{m:g:c})$ and $\sigma^2_g = \sigma^2_y\,(\mathrm{ICC}_{m:g:c})$.

As a result:

$$\sigma^2_e = \sigma^2_m\,(1 - r_{yy(m)}) = \sigma^2_y\,(1 - \mathrm{ICC}_{m:g:c})(1 - r_{yy(m)})$$

$$\sigma^2_{m:tg:c} = \sigma^2_m\,(r_{yy(m)}) = \sigma^2_y\,(1 - \mathrm{ICC}_{m:g:c})(r_{yy(m)})$$

$$\sigma^2_{g:c} = \sigma^2_g\,(r_{yy(g)}) = \sigma^2_y\,(\mathrm{ICC}_{m:g:c})(r_{yy(g)})$$

$$\sigma^2_{tg:c} = \sigma^2_g\,(1 - r_{yy(g)}) = \sigma^2_y\,(\mathrm{ICC}_{m:g:c})(1 - r_{yy(g)}) \qquad (7)$$

Simple substitutions allow us to rewrite the variance of the intervention effect as a function of σ^2_y, $\mathrm{ICC}_{m:g:c}$, $r_{yy(m)}$ and $r_{yy(g)}$:

$$\sigma^2_\Delta = 2 \times 2\left(\frac{\sigma^2_y(1 - \mathrm{ICC}_{m:g:c})(1 - r_{yy(m)})}{gmr} + \frac{r\sigma^2_y\,(1 - \mathrm{ICC}_{m:g:c})(r_{yy(m)})}{gmr} + \frac{mr\sigma^2_y(\mathrm{ICC}_{m:g:c})(1 - r_{yy(m)})}{gmr}\right) \qquad (8)$$

Equation (3) or (8) can be used to plan a study that is large enough to allow for the extra variation and limited precision that is common to the nested cross-sectional design. To use equation (3), investigators need good estimates of σ^2_e, $\sigma^2_{m:tg:c}$ and $\sigma^2_{tg:c}$. To use equation (8), investigators need good estimates of σ^2_y, $\mathrm{ICC}_{m:g:c}$, $r_{yy(m)}$ and $r_{yy(g)}$. These parameters are best estimated from data gathered on workers in worksites like those that will participate in the efficacy trial; that is the reason we conducted the pilot study. Once investigators have an estimate for σ^2_Δ, they can then use equation (9) (below) to estimate the detectable difference as a function of the number of workers (*m*) per worksite, the number of worksites (*g*) per condition and the desired Type I and II error rates ($t_{critical:\alpha/2} + t_{critical:\beta}$):

$$\hat{\Delta} = \sqrt{\hat{\sigma}^2_\Delta\,(t_{critical:\alpha/2} + t_{critical:\beta})^2} \qquad (9)$$

More generally, equation (3) or (8) in combination with equation (9) can be used to plan any group-randomized trial that employs the same design and analytical plan. They can also be adapted to reflect modifications in that plan; for example, if no repeat measurements are planned, $r_{yy(m)}$ is set to 0 and r is fixed at 1. Murray (1998) presents expressions corresponding to equations (3) and (8) for a wide variety of other designs and analytical plans commonly employed in group-randomized trials; equation (9) is applicable to any design and analytical plan having a 1 df contrast as an intervention effect.

Analysis

We analyzed the data from the pilot study using a time × condition mixed model ANOVA with SAS PROC MIXED v.6.12 (SAS Institute, 1997), allowing for repeat observations on members. We did not make any regression adjustment for covariates in our analyses, so that the variance component estimates do not reflect any increase or decrease that might result from such an adjustment. We checked the residual error distribution and confirmed that it was normal.

We report here the mean dust concentrations and variances associated within and between worksites that we observed from these pilot data. As an illustration of these statistical methods, we used these estimates to also calculate a range of smallest detectable differences, expressed as a percent reduction in dust, to assess the interplay between the number of worksites (*g*), the number of workers (*m*) and the number of repeat measurements on each worker (*r*) and their effect on sample size in group-randomized trials conducted in an occupational setting. Finally, we express the detectable differences in standard deviation units which can be used to adapt the results to other dependent variables.

RESULTS

Table 2 presents the mean wood dust concentration across worksites as 4.53 mg/m^3 on the original scale and 1.05 ln(mg/m^3) on the log scale. Table 2 also presents the estimates for the four components of variance identified in equation (1), which ranged from 0.0585 to 0.4859. Using these estimates and equations (4) and (5), the over-time correlations for worksites and workers were estimated as 0.8925 and 0.1927, respectively. Table 2 also presents the intraclass correlation of workers within worksites, estimated as 0.5086; the estimate is interpretable as an average correlation among workers in the same worksite. It is quite large, but that is often the case for $\mathrm{ICC}_{m:g:c}$ in studies involving repeat observations on the same members (Murray, 1998). Fortunately, the over-time correlation within worksites was also quite large, at 0.8925, so that the impact of this large $\mathrm{ICC}_{m:g:c}$ was substantially reduced. The over-time correlation within workers was more modest, but also helped to reduce the impact of the large $\mathrm{ICC}_{m:g:c}$. The

Table 2. Results from the analysis of dust measurements from five pilot woodworking businesses for calculation of sample size estimates needed in a group-randomized trial to assess the efficacy of an intervention to reduce dust levels

Raw scale		
Mean dust concentration	mg/m^3	4.53
Log scale		
Mean dust concentration	mg/m^3	1.05
Workers per worksite	m	7
Variance due to worksites	$\hat{\sigma}^2_g = \hat{\sigma}^2_{g:c} + \hat{\sigma}^2_{tg:c}$	0.5444
Between worksite variance	$\hat{\sigma}^2_{g:c}$	0.4859
Within worksite variance	$\hat{\sigma}^2_{tg:c}$	0.0585
Correlation within worksites over time	$\hat{r}_{yy(g)} = \hat{\sigma}^2_{g:c} / (\hat{\sigma}^2_{g:c} + \hat{\sigma}^2_{tg:c})$	0.8925
Variation due to workers	$\hat{\sigma}^2_m = \hat{\sigma}^2_e + \hat{\sigma}^2_{m:tg:c}$	0.5259
Between worker variance	$\hat{\sigma}^2_{m:tg:c}$	0.1014
Within worker variance	$\hat{\sigma}^2_e$	0.4246
Correlation within workers over time	$\hat{r}_{yy(m)} = \hat{\sigma}^2_{m:tg:c} / (\hat{\sigma}^2_e + \hat{\sigma}^2_{m:tg:c})$	0.1927
Total variance	$\hat{\sigma}^2_y$	1.0703
ICC, workers within worksites	$\widehat{ICC}_{m:g:c} = \hat{\sigma}^2_g / (\hat{\sigma}^2_m + \hat{\sigma}^2_g)$	0.5086

over-time correlations reduce the effect of $ICC_{m:g:c}$ by reducing the variance of the intervention effect, as shown in equation (8).

In Table 3 we present detectable differences for wood dust, expressed both as percent reduction in log dust and as standard deviation units. These estimates are presented as a function of the number of workers per worksite and the number of worksites per condition, with the Type I error rate fixed at 5% (two-tailed), the power fixed at 80% and either 2 or 10 repeat observations per worker. As expected, as the number of worksites and/or workers increases, the size of the detectable difference decreased. And, as in all group-randomized trials, with other factors held constant, the improvement in power was greater by increasing the number of worksites per condition than by increasing the number of workers per worksite. For example, in Table 3A, with 10 worksites per condition, the detectable difference decreases from 62 to 48% as the number of workers increases from 5 to 25. On the other hand, if the number of workers in each worksite is held constant at 10, the detectable difference in dust levels ranges from 54% with just 10 worksites to 23% with 50 worksites in each condition. Table 3B presents detectable differences under the same conditions, except that the number of repeat measurements per worker was increased from 2 to 10. The increase in measurements per worker had minimal impact on the detectable difference except where there were few workers and few worksites.

Table 3C and D expresses the detectable differences in standard deviation units. These results may be used to estimate the detectable difference for any other dependent variable, given the same design and analytical plan and faith that the intraclass and over-time correlations are the same as for wood dust. Those are not assumptions that should be made for a final sample size calculation, but these data can provide an initial indication of what the detectable difference could be for other dependent variables.

Since we were interested in testing the efficacy of an intervention to decrease wood dust levels by ~30%, we used the data in Table 3 to set a target of 60 worksites to be randomized, 30 to the intervention condition (technical training/worker education) and 30 to a comparison condition (written recommendations). We planned to take 10 dust measurements at baseline and again 1 yr later. Given the low correlation from repeated measures within workers and the fact that worksites could have as few as five employees, we treated workers measured more than once in a worksite as independent observations. The total number of measurements (or workers) is exactly what the data in Table 3 would indicate was necessary to have the power to detect a 30% reduction in wood dust, avoiding the need to sample on additional days in the small worksites we were recruiting.

DISCUSSION

In 1998 the National Institute for Occupational Safety and Health (NIOSH) published the National Occupational Research Agenda (Rosenstock *et al.*, 1998). The priorities that form the agenda were arrived at through consensus building among NIOSH staff, researchers, stakeholders and health professionals and included the prevention of selected diseases and injuries, identification of issues pertaining to work environment and workforce (e.g. energy technologies, special populations at risk) and the need for application of research tools and approaches to occupational health and safety. This latter category included priorities ranging from surveillance and risk assessment to intervention effectiveness research. The Minnesota Wood Dust Study addresses several of the subheadings within these broad national priorities, in particular, the need for intervention effectiveness research.

Despite the importance assigned by NIOSH to evaluation of interventions to reduce health risks in the work place, few studies, as described in several recent reviews (Goldenhar and Schulte, 1994; Johnston *et al.*, 1994; Karas and Conrad, 1996; Zwerling *et al.*, 1997), have been conducted that adhere to accepted study design principles and provide an acceptable test of the efficacy of the intervention under evaluation. We identified only six studies in the occupational health and safety literature of 1999 in which groups of workers were randomized to an intervention or comparison condition (Marcus *et al.*, 1986; Martyny *et al.*, 1988; Parkinson *et al.*, 1989; Hillyer *et al.*, 1990; Gjerde *et al.*, 1991; Sorensen *et al.*, 1998); all but one of these was conducted in the 1980s. Only the most recent study (Sorensen, 1998) reported results that were analyzed according to the methods presented here. While the intervention in that study had an occupational component to it, the primary outcome was a change in individual smoking and dietary behavior.

Two of the more significant challenges of conducting group-randomized trials involve the total number of groups required and the effect of the intraclass correlation among participants within a group on the sample size. As seen in our study, we gained more power to detect differences between groups as we increased the number of woodworking businesses than if we increased the number of workers within a worksite. However, increasing the number of groups randomized can be quite costly and may be limited by the total number of groups available. The problem with the intraclass correlation is well illustrated by the value of 0.5086 that we observed among workers within worksites. That value is inflated in part because it does not reflect the substantial over-time correlation observed in our data. Recognizing that the group component of variance is reduced by that over-time correlation, the ICC that is operative in this design and analytical plan is better reflected as:

$$\mathrm{ICC}_{\mathrm{m:tg:c}} = \frac{\sigma^2_{\mathrm{g}}\,(r_{yy(\mathrm{g})})}{\sigma^2_{\mathrm{m}} + \sigma^2_{\mathrm{g}}\,(r_{yy(\mathrm{g})})} \qquad (10)$$

Using the estimates from Table 2, ICC is estimated as 0.1001. Though considerably smaller, that estimate is still much higher than has typically been observed in worksite health promotion studies (Feng *et al.*, 1999; Martinson *et al.*, 1999). It is not too surprising that worksites would exhibit a greater within-worksite correlation in occupational exposures than has been observed for individual behaviors, due to the influence of a variety of environmental factors, such as type of manufacturing processes, methods of exposure comparison, size of the facility and number of workers, to name a few. If we had ignored this correlation in the design or analysis of the efficacy trial, we would have greatly underestimated the variance, thereby considerably increasing the risk of falsely concluding that our intervention was effective (Type I error).

In addition to number of worksites and the intraclass correlation, there are other factors we did not take into account which may affect our ability to detect a statistically significant reduction in dust attributed to the intervention. First, matching groups on characteristics likely to be associated with the outcome prior to randomization has been recommended as a strategy for limiting potential selection bias (Murray, 1998). Since we planned to employ a sequential recruitment strategy, it would not be possible to know in advance about worksite characteristics on which to match. Another potential problem is that we based our estimate of the correlation between dust concentrations over time within a woodworking business on measurements taken over a short period of time. If the correlation between dust measurements between baseline and follow-up 1 yr later is not that high, then we will have underestimated the number of worksites needed to have the power to detect the specified reduction in dust levels. This could occur if woodworking businesses are engaged in a different aspect of the production process when follow-up measurements are taken a year later. However, we plan to collect data on the work tasks being performed at baseline and follow-up, so that we can adjust for these differences in the analysis, recognizing that such adjustment may further reduce the precision of our estimates. Other problems we may encounter, e.g. attrition over time or a smaller than anticipated reduction in wood dust, could also effect the precision of our estimate. Each of these limitations could be addressed by repeating the calculations summarized in Table 3 after varying the magnitude of the correlations, the size of the ICC, the number of worksites per condition, etc., so as to

Table 3. The smallest detectable difference as a function of the number of workers measured per worksite and the number of worksites per condition for a 5% two-tailed Type I error rate and 80% power

No. of workers	No. of worksites				
	10	20	30	40	50
(A) Expressed as a percent reduction in log dust for two repeat measurements per worker					
5	62.1	42.6	34.5	29.8	26.5
10	53.5	36.7	29.7	25.6	22.9
15	50.3	34.5	27.9	24.1	21.5
20	48.6	33.4	27.0	23.3	20.8
25	47.6	32.6	26.4	22.8	20.3
(B) Expressed as a percent reduction in log dust for 10 repeat measurements per worker					
5	52.7	36.2	29.3	25.2	22.5
10	48.2	33.1	26.8	23.1	20.6
15	46.6	32.0	25.9	22.3	19.9
20	45.7	31.4	25.4	21.9	19.5
25	45.2	31.0	25.1	21.7	19.3
(C) Expressed in standard deviation units for two repeat measurements per worker					
5	0.631	0.433	0.350	0.302	0.269
10	0.543	0.373	0.302	0.260	0.232
15	0.510	0.350	0.283	0.244	0.218
20	0.493	0.339	0.274	0.236	0.211
25	0.483	0.331	0.268	0.231	0.206
(D) Expressed in standard deviation units for 10 repeat measurements per worker					
5	0.535	0.367	0.297	0.256	0.229
10	0.489	0.336	0.272	0.234	0.209
15	0.473	0.324	0.262	0.226	0.202
20	0.464	0.319	0.258	0.222	0.198
25	0.459	0.315	0.255	0.220	0.196

provide a sensitivity analysis. Indeed, each panel in Table 3 provides a sensitivity analysis for two of the most important factors, the number of groups per condition (g) and the number of members per group (m). We have shown that the number of repeats per member has little effect on power, so that the other variables that might be manipulated to help us gauge the adequacy of our plan would be the intraclass and over-time correlations.

The methods presented here are equally applicable to any outcome and any intervention in which identifiable social groups are the unit of assignment. Group-randomized trials are the design of choice whenever the intervention operates at a group level, manipulates the social or physical environment or cannot be delivered to individuals (Murray, 1998). Equation (3) or (8), in combination with equation (9), can be used to plan any group-randomized trial that employs the same design. And while it is beyond the scope of this paper to review the many alternative designs that are applicable to group-randomized trials, we can refer interested readers to two recent textbooks that cover this material (Murray, 1998; Donner and Klar, 2000).

CONCLUSION

We have illustrated the importance of using the proper methods for determining sample size when the unit of assignment is a group, such as a worksite, and have provided the methods and sources for doing so. We urge others conducting intervention effectiveness research in occupational health and safety to consider these methods when designing future studies.

Acknowledgements—Funding for this research was provided entirely by a grant (5 R01 OH03408) from the National Institute for Occupational Safety and Health.

REFERENCES

Brosseau LM, Parker DL, Lazovich D, Dugan SK, Milton FT, Pan W. (2001) Inhalable dust exposures, tasks and use of ventilation in small woodworking shops: a pilot. Am Ind Hyg Assoc J; 62: 322–9.

Cornfield J. (1978) Randomization by group: a formal analysis. Am J Epidemiol; 108 (2): 100–2.

Donner A. (1984) Linear regression analysis with repeated measurements. J Chronic Dis; 37: 441–8.

Donner A. (1985) A regression approach to the analysis of data arising from cluster randomization. Int J Epidemiol; 14 (2): 322–6.

Donner A, Klar N. (2000) Design and analysis of cluster randomization trials in health research. London: Arnold. ISBN 0 34069153 0.

Feldman HA, McKinlay SM. (1994) Cohort versus cross-sectional design in large field trials: precision, sample size, and a unifying model. Statist Med; 13: 61–78.

Feng Z, Diehr P, Yasui Y, Evans B, Beresford S, Koepsell TD. (1999) Explaining community-level variance in group randomized trials. Statist Med; 18: 539–56.

Gjerde C, Ferguson K, Mutel C, Donham K, Merchant J. (1991) Results of an educational intervention to improve the health knowledge, attitudes, and self-reported behaviors of swine confinement workers. J Rural Health; 7: 278–86.

Goldenhar LM, Schulte PA. (1994) Intervention research in occupational health and safety. J Occup Med; 36: 763–75.

Hannan PJ, Murray DM. (1996) Gauss or Bernoulli? A Monte Carlo comparison of the performance of the linear mixed model and the logistic mixed model analyses in simulated community trials with a dichotomous outcome variable at the individual level. Eval Rev; 20: 338–52.

Harville DA. (1977) Maximum likelihood approaches to variance component estimation and to related problems. J Am Statist Assoc; 72: 320–38.

Hillyer JC, Brown KC, Sirles AT, Peoples L. (1990) A flexibility intervention to reduce the incidence and severity of joint injuries among municipal firefighters. J Occup Med; 32: 631–7.

IARC. (1995) IARC monographs on the evaluation of the carcinogenic risks to humans: Vol. 62, Wood dust and formaldehyde. Lyon: International Agency for Research on Cancer. ISBN 9 283 51262 2.

Johnston JJ, Cattledge GH, Collins JW. (1994) The efficacy of training for occupational injury control. Occup Med; 9: 147–58.

Karas BE, Conrad KM. (1996) Back injury prevention interventions in the workplace. An integrative review. AAOHN J; 44 (4): 189–96.

Kish L. (1995) Survey sampling. New York: John Wiley & Sons. ISBN 0 471 10949 5

Koepsell TD, Martin DC, Diehr PH, Psaty BM, Wagner EH, Perrin EB, Cheadle A. (1991) Data analysis and sample size issues in evaluations of community-based health promotion and disease prevention programs: a mixed-model analysis of variance approach. J Clin Epidemiol; 44: 701–13.

Kromhout H, Heederik D. (1995) Occupational epidemiology in the rubber industry: implications of exposure variability. Am J Ind Med; 27: 171–85.

Kromhout H, Symanski E, Rappaport SM. (1993) A comprehensive evaluation of within- and between-worker components of occupational exposure to chemical agents. Ann Occup Hyg; 37: 253–70.

Kromhout H, Swuste P, Boleij JSM. (1994) Empirical modelling of chemical exposure in the rubber-manufacturing industry. Ann Occup Hyg; 38: 3–22.

Kromhout H, Loomis DP, Mihlan GJ, Peipins LA, Kleckner RC, Iriye R, Savitz DA. (1995) Assessment and grouping of occupational magnetic field exposure in five electric utility companies. Scand J Work Environ Health; 21: 43–50.

Kromhout H, Tielemans E, Preller L, Heederik D. (1996) Estimates of individual dose from current measurements of exposure. Occup Hyg; 3: 23–39.

Laird NM, Ware JH. (1982) Random-effects models for longitudinal data. Biometrics; 38: 963–74.

Laird NM, Lange N, Stram D. (1987) Maximum likelihood computations with repeated measures: application of the EM algorithm. J Am Statist Assoc; 82: 97–105.

Lasater TM, Becker DM, Hill MN, Gans KM. (1997) Synthesis of findings and issues from religious-based cardiovascular disease prevention trials. Ann Epidemiol; 7 (suppl. 7): S46–S53.

Marcus AC, Baker DB, Froines JR, Brown ER, McQuiston T, Herman NA. (1986) ICWU cancer control education and evaluation program: research design and needs assessment. J Occup Med; 28: 226–36.

Martinson BC, Murray DM, Jeffery RW, Hennrikus DJ. (1999) Intraclass correlation for measures from a worksite health promotion study: estimates, correlates and applications. Am J Health Promot; 13 (6): 347–57.

Martyny J, Buchan RM, Keefe TJ, Blehm KD. (1988) The impact of OSHA Onsite Consultation Program on small businesses in Colorado. Appl Ind Hyg; 3: 12–4.

Murray DM. (1998) Design and analysis of group-randomized trials. New York: Oxford University Press. ISBN 0 195 12036 1.

Murray DM, Wolfinger RD. (1994) Analysis issues in the evaluation of community trials: progress toward solutions in SAS/STAT MIXED. J Community Psychol; CSAP Special Issue: 140–54.

Murray DM, Hannan PJ, Baker WL. (1996) A Monte Carlo study of alternative responses to intraclass correlation in community trials: is it ever possible to avoid Cornfield's penalties? Eval Rev; 20: 313–37.

National Institute for Occupational Safety and Health. (1988) National occupational exposure survey analysis of management interview responses. DHHS publication no. 89-10. Cincinnati, OH: DHHS (NIOSH).

Ockene JK, McBride P, Sallis J, Bonollo D, Ockene IS. (1997) Synthesis of lessons learned from cardiopulmonary preventive interventions in healthcare practice settings. Ann Epidemiol; 7 (suppl. 7): 32–45.

Parkinson DK, Bromet EJ, Dew MA, Dunn LO, Barkman M, Wright M. (1989) Effectiveness of the United Steel Workers of America coke oven intervention program. J Occup Med; 31: 464–72.

Rappaport SM, Lyles RH, Kupper LL. (1995) An exposure-assessment strategy accounting for within- and between-worker sources of variability. Ann Occup Hyg; 39: 469–95.

Rappaport SM, Weaver M, Taylor D, Kupper L, Susi P. (1999) Application of mixed models to assess exposures monitored by construction workers during hot processes. Ann Occup Hyg; 43: 457–69.

Resnicow K, Robinson TN. (1997) School-based cardiovascular disease prevention studies: review and synthesis. Ann Epidemiol; 7 (suppl. 7): S14–S31.

Rosenstock L, Olenec C, Wagner GR. (1998) The national occupational research agenda: a model of broad stakeholder input into priority setting. Am J Publ Health; 88: 353–6.

SAS Institute. (1997) SAS/STAT software: changes and enhancements through release 6.12. Cary, NC: SAS Institute. ISBN 1 55544 873 9.

Schooler C, Farquhar JW, Fortmann SP, Flora JA. (1997) Synthesis of findings and issues from community prevention trials. Ann Epidemiol; 7 (suppl. 7): S54–S68.

Searle SR. (1977) Linear models. New York: John Wiley & Sons. ISBN 0 471 18499 3.

Sorensen G, Stoddard A, Hung MK, Hebert JR, Ockene JK, Avrunin JS, Himmelstein J, Hammond SK. (1998) The effects of a health promotion–health protection intervention on behavior change: the WellWorks Study. Am J Publ Health; 88: 1685–90.

Stiratelli R, Laird N, Ware JH. (1984) Random-effects models for serial observations with binary response. Biometrics; 40: 961–71.

Tornero-Velez R, Symanski E, Kromhout H, Yu RC, Rappaport SM. (1997a) Compliance versus risk in assessing occupational exposures. Risk Anal; 17: 279–92.

Tornero-Velez R, Symanski E, Kromhout H, Yu RC, Rappaport SM. (1997b) Errata. Risk Anal; 17: 657.

Ware JH. (1985) Linear models for the analysis of longitudinal studies. Am Statistician; 39 (2): 95–101.

Zucker DM. (1990) An analysis of variance pitfall: the fixed effects analysis in a nested design. Educ Psychol Meas; 50: 731–8.

Zwerling C, Daltroy LH, Fine LJ, Johnston JJ, Melius J, Silverstein BA. (1997) Design and conduct of occupational injury intervention studies: a review of evaluation strategies. Am J Ind Med; 32: 164–79.