The International Journal of Biostatistics

Volume 4, *Issue* 1 2008 *Article* 8

Instrumental Variables vs. Grouping Approach for Reducing Bias Due to Measurement Error

Evridiki Batistatou, *University of Manchester*
Roseanne McNamee, *University of Manchester*

Recommended Citation:
Batistatou, Evridiki and McNamee, Roseanne (2008) "Instrumental Variables vs. Grouping Approach for Reducing Bias Due to Measurement Error," *The International Journal of Biostatistics*: Vol. 4: Iss. 1, Article 8.
DOI: 10.2202/1557-4679.1087

Instrumental Variables vs. Grouping Approach for Reducing Bias Due to Measurement Error

Evridiki Batistatou and Roseanne McNamee

Abstract

Attenuation of the exposure-response relationship due to exposure measurement error is often encountered in epidemiology. Given that error cannot be totally eliminated, bias correction methods of analysis are needed. Many methods require more than one exposure measurement per person to be made, but the `group mean OLS method,' in which subjects are grouped into several a priori defined groups followed by ordinary least squares (OLS) regression on the group means, can be applied with one measurement. An alternative approach is to use an instrumental variable (IV) method in which both the single error-prone measure and an IV are used in IV analysis. In this paper we show that the `group mean OLS' estimator is equal to an IV estimator with the group mean used as IV, but that the variance estimators for the two methods are different. We derive a simple expression for the bias in the common estimator which is a simple function of group size, reliability and contrast of exposure between groups, and show that the bias can be very small when group size is large. We compare this method with a new proposal (group mean ranking method), also applicable with a single exposure measurement, in which the IV is the rank of the group means. When there are two independent exposure measurements per subject, we propose a new IV method (EVROS IV) and compare it with Carroll and Stefanski's (CS IV) proposal in which the second measure is used as an IV; the new IV estimator combines aspects of the `group mean' and `CS' strategies. All methods are evaluated in terms of bias, precision and root mean square error via simulations and a dataset from occupational epidemiology. The `group mean ranking method' does not offer much improvement over the `group mean method.' Compared with the `CS' method, the `EVROS' method is less affected by low reliability of exposure. We conclude that the group IV methods we propose may provide a useful way to handle mismeasured exposures in epidemiology with or without replicate measurements. Our finding may also have implications for the use of aggregate variables in epidemiology to control for unmeasured confounding.

KEYWORDS: exposure measurement error, grouping approach, aggregate variables, instrumental variables

Author Notes: The authors owe special thanks to Dr. Martie van Tongeren (Institute of Occupational Medicine, Edinburgh, UK) who provided us with the data from the carbon black study. We also thank the two referees for their helpful comments.

1. INTRODUCTION

Predictor variable measurement error is a common source of error in many epidemiological studies which seek to relate potential risk factors to health outcomes. In nutritional epidemiology, for example, the average daily dietary intake of specific nutrients may be measured with great imprecision [33]. Occupational or environmental exposure measurements may also have considerable error [22]. Although exposure measurement error may have a great effect on study results [3, 16, 27], often epidemiologists do not account for error, as indicated by a survey of 57 epidemiological studies published recently [18]. Failure to account for measurement error in the analysis tends to result in bias, often attenuation, when estimating regression coefficients, and can hide the true effect of exposure on health [11, 16].

This paper is concerned with the problem of estimating the regression of response Y on true exposure X when X is observed through error-prone W only, by using bias correction methods. In recent years, an extensive range of methods has appeared [10, 13], many of which depend on information about the measurement error variance obtained through *independent replicate measurements*. The method of regression calibration [10] for example, involves an approximate replacement of the true exposure variable by its expected value, conditional on the observed error-prone exposure W.

Attenuation due to mismeasured variables could also be handled by using instrumental variables (IV). An elementary introduction to instrumental variable methodology can be found in [15, 21]. In the IV method, an instrumental variable which is uncorrelated with the error in W, but correlated (ideally highly) with the predictor variable is used *together with W,* in an *IV analysis*. If the key IV assumptions, discussed later, are satisfied, the IV method provides a consistent estimate of the predictor-response relation; otherwise the IV estimator is likely to be biased. Here we will refer to a 'true' IV as one which satisfies the key assumptions. Carroll and Stefanski [9], following an idea first proposed by Fuller [13], suggested that an independent replicate measurement could be treated as an instrumental variable (IV). We will call this approach the *CS* (Carroll and Stefanski) *IV method*.

Other methods, *which do not require replicate measurements*, involve grouping or ranking observations. In one such method first described by Wald [32], one groups the data into two equal sized groups on the basis of the ordered observations W; the Wald estimator is then a function of the group means of Y and W. In an extension of this method, Bartlett argued [4] that by forming three groups instead of two, then excluding the middle one and using the same estimator as Wald, one could achieve a higher precision. Durbin [12] showed that the Wald estimator is the same as an IV estimator when the instrumental variable takes the values -1 (first group) and +1 (second group); he also suggested another IV approach in which the ranks of the Ws is used as the IV. However, he noted that, so far as bias was concerned, the grouping in the Wald method and the ranking in his method should be "relatively unaffected by the errors" in X. It was left unclear how a relationship between measurement error E and rank of W can be avoided. The Wald estimator is only a true IV if the grouping is independent of W [7].

Another grouping approach, applicable when there is just one measurement per subject, is often applied in occupational epidemiology. We will refer to it as the *group mean OLS (ordinary least squares) method*. Suppose that, prior to measuring worker exposure, hygienists group people on the basis of similar work characteristics. The underlying principle of the grouping is that workers within each group experience similar exposure levels. Exposure W is then measured, the group mean of W is assigned to each subject within a group *in place of their individual measure W,* and an *ordinary least squares* (OLS) regression of Y on group means is carried out [22, 23]. Tielemans et al. [28] present formulae for the estimated exposure-response relation and its precision under this method. Some authors [3, 27] have justified the method on the grounds that a Berkson error model applies to the means, but Kim et al. [19] note that a true Berkson error structure never applies to observational data. The Berkson model was originally proposed [5] for experimental data in which there is a fixed target 'dose' for a number of experimental units but where, because of imprecise control, the true dose is randomly distributed around the target value. With true Berkson error, there is no attenuation in the estimated predictor-response relation [5, 12].

There are similarities between the problem of unmeasured confounders when estimating exposure or treatment effects and the measurement error problem; IV methods have also been proposed for the former [1, 2]. In an application to compare two treatments in the presence of unmeasured confounding, Johnston et al [17] applied a method identical in structure to the *group mean OLS* approach, replacing actual treatment by a grouped (by hospital) treatment variable in their logistic regression. They briefly justified their method on the grounds that use of grouped variables in place of individual level variables in regression was equivalent to IV methodology but there does not appear to be any formal proof in the literature.

The *group mean OLS method* is potentially useful for measurement error problems because it does not require replicate measurements, although it does require other, 'external' information (e.g. 'hygienist judgement') to provide an independent grouping. The concept links with the use of 'ecological' variables, for example dietary fat per capita measured for countries rather than individuals, in regression. Therefore its properties need to be properly established. In this paper we provide a theoretical link between the *group mean OLS method* and instrumental variable theory. In particular, we prove that the *group mean OLS* estimator of the regression coefficient of Y on W is equal to the estimator from a *group mean IV method* in which group means are used as the instrumental variable in an *IV analysis*. We go on to show that this common estimator is biased - although the bias can be very small - and derive a simple expression for the bias. However we prove that the variance estimators from the *group mean IV method* and the *group mean OLS method* are *not* identical, and use simulation studies to explore which variance estimator is closer to the truth.

We also propose and evaluate two new IV methods which build on the concept of using group means: in the first, referred to as the *group mean ranking IV method,* the ranking of group means rather than the means themselves, is used as an instrumental variable. The second proposed method, applicable when an independent replicate measurement is also available, combines the *CS IV* approach and the *group mean IV* approach: this method will be called the *EVROS IV method*. The rationale for these new methods is explained further later. The performance of all the methods is evaluated in terms of attenuation, precision and root mean square error (RMSE), which combines bias and precision, through simulation studies. The methods are illustrated using an application from occupational epidemiology.

The structure is as follows: Section 2 introduces the carbon black study which motivated the work and Section 3 the basic notation adopted in the paper. In Section 4, the theoretical link between the *group mean OLS approach* and *group mean IV methods* is shown and mathematical expressions are derived for judging the effect of invalid IV assumptions on bias and the strength of a grouping scheme. In Section 5, the simulation methods and software are discussed. In Sections 6 and 7, we evaluate and compare the methods proposed in this work via simulated studies and the carbon black study. Conclusions are addressed in Section 8.

2. MOTIVATING EXAMPLE

Typically in occupational epidemiology, measures of cumulative or average exposure over a period of time are needed for each worker. For reasons of practicality, measuring devices often operate only over a short period of time, yet the exposure of a worker can vary greatly within and between days. Hence, mismeasurement of exposure is common and can lead to failure to identify the true effect of exposure on the health outcome of interest. Given that occupational epidemiology aims to prevent morbidity/mortality due to work, methods for overcoming exposure measurement error should be used.

This paper was motivated by a study investigating the acute effect of exposure to carbon black on respiratory morbidity in the European carbon black manufacturing industry [14, 30]. The study was carried out over three cross sectional phases between 1987 and 1995, among several factories in seven European countries. Here, only data from the third phase (1994-1995) are used. During this period of time, several daily measurements for the inhalable exposure were taken from each of 990 workers. Since inhalable dust exposure may vary greatly within and between days, considerable measurement error may occur. As a measure of worker's lung function, FEV_1 (Forced Expiratory Volume in 1 second) was determined together with information on age, height and cumulative smoking. Further technical details are given elsewhere [29, 30].

3. NOTATION

The basic notation is shown below. Other notations are defined later.

X true exposure
W error-prone exposure
E exposure measurement error
λ reliability coefficient of exposure W
Y response
T instrumental variable

4. LINK BETWEEN '*GROUP MEAN OLS*' APPROACH AND IV METHODS

Consider a simple linear model without covariates relating a continuous exposure variable X and a continuous health outcome Y:

$$Y=\alpha+\beta X+\delta, \quad (1)$$

where $\mathrm{cov}(\delta,X)=0$ and the variance of X is V(X). Suppose that, due to random measurement error, the true exposure X is not observed; instead an error-prone measure W which follows the classical (additive) error model is available:

$$W=X+E, \quad (2)$$

where $E \sim N(0, V(E))$. The measurement error E is assumed to be independent of X and non-differential, that is, W is independent of the outcome Y, given the true measurement, X.

DOI: 10.2202/1557-4679.1087

Suppose that data on Y and W are available on n subjects. It is well documented [16] that the regression coefficient, $\hat{\beta}_1$ say, estimated from the ordinary least squares regression (OLS) of Y on W satisfies the equation

$$E(\hat{\beta}_1)=\lambda\beta, \tag{3}$$

where λ, the 'reliability' coefficient of W, is defined by

$$\lambda=\frac{\mathrm{V(X)}}{\mathrm{V(X)+V(E)}}. \tag{4}$$

Suppose now the n subjects can be grouped into g *a priori* defined groups, according to an expectation of similar exposure characteristics, with exposure measurements W_{ij}, j=1,…,k_i and group mean $\overline{\mathrm{W}}_i=k_i^{-1}\sum_{j=1}^{k_i}\mathrm{W_{ij}}$ in the ith group, i=1,…,g. In the *group mean OLS* method, the group mean estimate, $\hat{\beta}_\mathrm{G}$ say, is obtained by replacing W_{ij} for each subject by their group mean and then using ordinary OLS to regress Y on these values. An intuitive rationale for the method is that, if there was no within-group exposure variation in X, then the k_i measurements for group i could be viewed as replicate measurements; since the mean of k_i replicates has higher reliability, $\frac{\mathrm{V(X)}}{\mathrm{V(X)+V(E)}/k_i}$, than a single measurement alone, $\hat{\beta}_\mathrm{G}$ would tend to be less attenuated than $\hat{\beta}_1$.

We next provide a theoretical link between the *group mean OLS* method and instrumental variable theory. In the general IV approach to measurement error using any instrumental variable T, the IV estimate of β is given [13] by:

$$\hat{\beta}_\mathrm{IV}=\frac{\sum_j(\mathrm{Y}_j-\overline{\mathrm{Y}})(\mathrm{T}_j-\overline{\mathrm{T}})}{\sum_j(\mathrm{W}_j-\overline{\mathrm{W}})(\mathrm{T}_j-\overline{\mathrm{T}})}=\frac{\hat{\beta}_\mathrm{YT}}{\hat{\beta}_\mathrm{WT}}, \tag{5}$$

where $\hat{\beta}_\mathrm{YT}$ and $\hat{\beta}_\mathrm{WT}$ are estimated regression coefficients in OLS regressions of Y on T and W on T, respectively. The variance of $\hat{\beta}_\mathrm{IV}$ is estimated [13] by:

$$\hat{\mathrm{V}}(\hat{\beta}_\mathrm{IV})=\frac{s^2(\hat{\beta}_\mathrm{IV},\mathrm{W})\hat{\mathrm{V}}(\mathrm{T})}{(n-1)\hat{\mathrm{cov}}^2(\mathrm{W,T})}, \tag{6}$$

where $\hat{\mathrm{V}}(\mathrm{T})$, $\hat{\mathrm{cov}}(\mathrm{W,T})$ are estimates of the variance of T and its covariance with W respectively, and $s^2(\hat{\beta}_\mathrm{IV},\mathrm{W})$ is a residual sum of squares:

$s^2(\hat{\beta}_{\mathrm{IV}}, \mathrm{W}) = (n-2)^{-1}\sum_i\sum_j[\mathrm{Y}_{ij} - \overline{\mathrm{Y}} - \hat{\beta}_{\mathrm{IV}}(\mathrm{W}_{ij} - \overline{\mathrm{W}})]^2$. It follows from (6) that the stronger the correlation between T and W, the more precise the estimator β_{IV}.

Consider now an instrumental variable defined by:

$$\mathrm{T}_{ij} = \overline{\mathrm{W}}_i, \tag{7}$$

where i=1,…,g, j=1,…,k_i and the corresponding IV estimator, $\hat{\beta}_{\mathrm{IV}=\overline{\mathrm{W}}}$, say. The numerator of (5) is now $\hat{\beta}_{\mathrm{Y}\overline{\mathrm{W}}}$, which is equal to $\hat{\beta}_{\mathrm{G}}$, while the denominator is

$$\hat{\beta}_{\mathrm{W}\overline{\mathrm{W}}} = \frac{\hat{\mathrm{cov}}(\mathrm{W},\overline{\mathrm{W}})}{\hat{\mathrm{V}}(\overline{\mathrm{W}})} = \frac{\sum_i\sum_j(\mathrm{W}_{ij} - \overline{\mathrm{W}}_i + \overline{\mathrm{W}}_i - \overline{\mathrm{W}})(\overline{\mathrm{W}}_i - \overline{\mathrm{W}})}{\sum_i k_i(\overline{\mathrm{W}}_i - \overline{\mathrm{W}})^2} = \frac{0 + \sum_i k_i(\overline{\mathrm{W}}_i - \overline{\mathrm{W}})^2}{\sum_i k_i(\overline{\mathrm{W}}_i - \overline{\mathrm{W}})^2} = 1. \tag{8}$$

Hence, from (5) and (8), we see that $\hat{\beta}_{\mathrm{IV}=\overline{\mathrm{W}}} = \hat{\beta}_{\mathrm{G}}$. Also from (8), $\hat{\mathrm{cov}}(\mathrm{W},\overline{\mathrm{W}}) = \hat{\mathrm{V}}(\overline{\mathrm{W}})$. Thus, expression (6) simplifies to

$$\hat{\mathrm{V}}(\hat{\beta}_{\mathrm{IV}=\overline{\mathrm{W}}}) = \frac{s^2(\hat{\beta}_{\mathrm{IV}=\overline{\mathrm{W}}}, \mathrm{W})}{(n-1)\hat{\mathrm{V}}(\overline{\mathrm{W}})}. \tag{9}$$

This is *not* the same as the variance estimator obtained from a conventional OLS regression of Y on the group means; the latter is

$$\hat{\mathrm{V}}(\hat{\beta}_{\mathrm{G}}) = \frac{s^2(\hat{\beta}_{\mathrm{G}}, \overline{\mathrm{W}})}{(n-1)\hat{\mathrm{V}}(\overline{\mathrm{W}})} = \frac{s^2(\hat{\beta}_{\mathrm{IV}=\overline{\mathrm{W}}}, \overline{\mathrm{W}})}{(n-1)\hat{\mathrm{V}}(\overline{\mathrm{W}})}. \tag{10}$$

Hence the *group mean OLS* method and the *group mean IV* method give rise to the same estimate of β but estimate its variance differently. It can be shown that the difference in variances is given by:

$$\frac{\hat{\mathrm{V}}(\hat{\beta}_{\mathrm{IV}=\overline{\mathrm{W}}}) - \hat{\mathrm{V}}(\hat{\beta}_{\mathrm{G}})}{\hat{\beta}_{\mathrm{G}}^2} = \frac{1}{(n-2)}\left[1 - \frac{1}{\varepsilon(\overline{\mathrm{W}})}\cdot\left(\frac{2\hat{\beta}_1}{\hat{\beta}_{\mathrm{G}}} - 1\right)\right]. \tag{11}$$

where $\varepsilon(\overline{\mathrm{W}}) = \hat{\mathrm{V}}(\overline{\mathrm{W}})/\hat{\mathrm{V}}(\mathrm{W})$, and both variances are across individuals. The difference in (11) can be positive or negative. The question of which method is more accurate is addressed later.

DOI: 10.2202/1557-4679.1087

In general, instrumental variable methods provide consistent estimators only if they satisfy certain assumptions. The following formula given by Carroll et al. [10] is useful for judging the impact of invalid assumptions:

$$\hat{\beta}_{\mathrm{IV}} \rightarrow \frac{\mathrm{cov}(\mathrm{T},\mathrm{X})\beta + \mathrm{cov}(\mathrm{T},\delta)}{\mathrm{cov}(\mathrm{T},\mathrm{X}) + \mathrm{cov}(\mathrm{T},\mathrm{E})}, \tag{12}$$

where cov(T,X), cov(T,δ) and cov(T,E) are covariances. Thus we see that, for a consistent estimator, an instrumental variable should have:
(a) a nonzero correlation with X;
(b) a zero correlation with the measurement error E=W-X and
(c) a zero correlation with Y, after accounting for X.

For the instrumental variable (7), assumption (c) is satisfied if the model in (1) and (2) is valid. However assumption (b) is not satisfied since:

$$\mathrm{cov}(\mathrm{T},\mathrm{E}) = \mathrm{cov}(\overline{\mathrm{W}},\mathrm{E}) = \mathrm{cov}(\overline{\mathrm{X}}+\overline{\mathrm{E}},\mathrm{E}) = \mathrm{cov}(\overline{\mathrm{E}},\mathrm{E}) = \mathrm{V}(\overline{\mathrm{E}})\,. \tag{13}$$

Hence the group mean is not a ‘true’ IV. However, the covariance (13) can be made arbitrarily small by increasing group sizes; for example, when $k_i=k$ for each i, we have $\mathrm{cov}(\mathrm{T},\mathrm{E}) = \mathrm{V}(\mathrm{E})/k$.

The covariance corresponding to (a) is:

$$\mathrm{cov}(\mathrm{T},\mathrm{X}) = \mathrm{cov}(\overline{\mathrm{W}},\mathrm{X}) = \mathrm{cov}(\overline{\mathrm{X}}+\overline{\mathrm{E}},\mathrm{X}) = \mathrm{cov}(\overline{\mathrm{X}},\mathrm{X}) = \mathrm{V}(\overline{\mathrm{X}}), \tag{14}$$

where $\mathrm{V}(\overline{\mathrm{X}})$ is the between-group exposure variance. Hence, in the case where $k_i=k$ for each i, expression (12) becomes

$$\hat{\beta}_{\mathrm{IV}=\overline{\mathrm{W}}} \rightarrow \frac{\mathrm{V}(\overline{\mathrm{X}})\beta}{\mathrm{V}(\overline{\mathrm{X}})+\mathrm{V}(\mathrm{E})/k} = \frac{\varepsilon(\overline{\mathrm{X}})}{\varepsilon(\overline{\mathrm{X}})+{(1-\lambda)}/{\lambda k}}\beta = \lambda_{\overline{\mathrm{W}}}\beta, \text{ say.} \tag{15}$$

where $\varepsilon(\overline{\mathrm{X}}) = \mathrm{V}(\overline{\mathrm{X}})/\mathrm{V}(\mathrm{X})$. The measure ε will be referred to as the ‘*contrast of exposure*’. If grouping of subjects was entirely random, then $\mathrm{V}(\overline{\mathrm{X}}) = \mathrm{V}(\mathrm{X})/k$ i.e. $\varepsilon(\overline{\mathrm{X}})=1/k$, so that $\lambda_{\overline{\mathrm{W}}} = \lambda$; in this case, nothing is gained by grouping. Hence, we need $k\varepsilon(\overline{\mathrm{X}}) > 1$ to achieve any bias reduction. From (15), it can also be shown that, in choosing between two grouping schemes with characteristics $\varepsilon_1(\overline{\mathrm{X}})$ and k_1 in scheme 1 and $\varepsilon_2(\overline{\mathrm{X}})$ and k_2 in scheme 2, scheme 2 produces less attenuation if $k_1\varepsilon_1(\overline{\mathrm{X}}) < k_2\varepsilon_2(\overline{\mathrm{X}})$. Thus the index $k \cdot \varepsilon(\overline{\mathrm{X}})$ is the key measure for judging the ‘strength’ of a grouping scheme.

In practice we cannot observe $\varepsilon(\overline{\mathrm{X}})$ but $\varepsilon(\overline{\mathrm{W}})$ can be estimated from data. However it is easy to show that

$$\varepsilon(\overline{\mathrm{W}}) = \lambda\varepsilon(\overline{\mathrm{X}}) + (1\text{-}\lambda)/k \; . \tag{16}$$

Substituting for $\varepsilon(\overline{\mathrm{X}})$ in (15) then gives the following remarkably simple expression for $\lambda_{\overline{\mathrm{W}}}$:

$$\lambda_{\overline{\mathrm{W}}} = \frac{k\varepsilon(\overline{\mathrm{W}}) - (1\text{-}\lambda)}{k\varepsilon(\overline{\mathrm{W}})} \, . \tag{17}$$

We can also show that $k_1\varepsilon_1(\overline{\mathrm{X}}) < k_2\varepsilon_2(\overline{\mathrm{X}}) \Leftrightarrow k_1\varepsilon_1(\overline{\mathrm{W}}) < k_2\varepsilon_2(\overline{\mathrm{W}})$ so that the choice between schemes is a simple function of observables.

The non-zero covariance in (13) arises because the instrument, being a function of W, is correlated with the error in W. If one could, independently of W, rank subjects in terms of X, or put them into groups which could then be ranked, the covariance (13) based on the ranking variable would then be zero, the IV assumptions (a)-(c) satisfied and a consistent IV estimator achieved. But, in practice, it may be difficult to achieve a ranking which has sufficient correlation with X, or to convince oneself that there is a reasonable correlation given that X is unobserved. It is tempting instead, to consider as an IV, a ranking variable which is derived from the group means. Compared to using the means as IV, one might expect this to produce a smaller covariance (13) – thus reducing bias – but also a smaller covariance between the instrument and X – thus increasing variance. Below we investigate whether there is any overall gain – in terms of RMSE – by using this new approach which we call the *group mean ranking IV method.* We note that Durbin's [12] suggestion that the ranking be based on W itself is in fact a special case of this method with a group size, *k*, equal to one, i.e. no grouping.

The *CS IV approach*, in which there are two independent measurements per subject, one of which is used as an IV, is a true IV method since the covariance (13) is expected to be zero and all the IV assumptions are satisfied. This is also true of our proposed *EVROS IV method,* also applicable with two measurements, which combines the group means and CS strategies: in this method, the replicate measurement is replaced by its group mean and used in IV analysis as an IV for the first measurement. Our interest in the new method was to see if it could outperform the *CS IV method* in terms of variance.

DOI: 10.2202/1557-4679.1087

5. SIMULATION METHODS

We carried out simulations to compare the performance of the *group mean OLS method*, the *group mean IV method* and the *group mean ranking IV method* for single measurement scenarios, and to compare the *EVROS IV* and the *CS IV methods* when there are two measurements. In all simulations, we assumed that true exposure, X, is Normally distributed: X~N(5,1) and the relation between X and the continuous Y is given by the model Y=4-0.1X+δ, where δ~N(0,0.15). Thus, in (1), the true value of β is -0.1.

We assumed that exposure is observed through an error-prone value W which follows the additive error model in (2). The error variances V(E) were chosen to give three different values for the reliability λ of W: 0.2, 0.5 and 0.8. For the *EVROS* and *CS IV methods,* two independent measurements, W_1 and W_2 were assumed to be available. The same simulated data were used to evaluate the single measurement methods; in this case, a single measure of exposure for each individual – equal to the mean of W_1 and W_2 and with corresponding reliability $\lambda_2=2\lambda/(1+\lambda)$ – was used as if it were a single observation of reliability 033, 0.67 or 0.89. The reason will be more apparent later. The sample size in all simulations was fixed at 1,000 subjects.

To apply the methods based on group means and group mean ranks, we had to first simulate the process of independent grouping of subjects according to similar exposure characteristics, using 'external' criteria not correlated with measurement error. We first simulated a grouping variable Z=X+v, where v~N(0,V(v)), assuming reliabilities for Z of 0.2, 0.5 and 0.8; equivalently, Z may be though of as an exogenous variable with correlation to X of $\sqrt{0.2}$, $\sqrt{0.5}$ and $\sqrt{0.8}$ respectively. The observed exposures (the mean of W_1 and W_2 for the *group mean OLS*, *group mean IV* and *group mean ranking method,* and W_2 for the *EVROS method*) are then ranked according to Z and split into several groups. The number of groups considered here are 2, 20 and 100 with corresponding group sizes *k* of 500, 50 and 10.

To apply the group mean methods for single measurement scenarios, the group mean (i.e. the means of the individual means) exposure was calculated and assigned to each subject within the group. As discussed in Section 4, the success of the grouping method in reducing attenuation depends on the index $k \cdot \varepsilon(\overline{\mathrm{X}})$, where ε is the '*contrast of exposure*'. To estimate $\varepsilon(\overline{\mathrm{W}})$ for each simulation, the ratio of the variance of the group mean exposure to the individual exposure (the mean of W_1 and W_2) was calculated. For *group mean OLS* estimation of β, conventional least squares regression was used to regress Y on the group means. For the *group mean IV method,* the set of group means was treated as an instrumental variable and used together with the individual mean in an IV analysis (see below). A similar approach, but with the ranks of the group means as the instrument, was applied under the *group mean ranking IV method*.

To apply the *EVROS IV method*, W_1 was used as the individual exposure, with the group mean of W_2 as its instrument in an IV analysis. To apply the *CS IV method*, W_1 was used as the individual exposure in an IV analysis with W_2 as its instrument.

In total, 20,000 independent datasets were generated at each combination of corr(Z,X), reliability λ (or λ_2) and *k*. For each set, the mean estimate $\bar{\hat{\beta}}$, the standard deviation of the estimates $\hat{\beta}$, denoted SE($\hat{\beta}$), and the Root Mean Square Error (RMSE) of $\hat{\beta}$, which combines both the bias in $\hat{\beta}$ and SE($\hat{\beta}$), were derived. The means of the estimated variances of $\hat{\beta}$ using the *group mean OLS method* and the *group mean IV method* – see (10) and (9) respectively – relative to the true variance of $\hat{\beta}$ were also compiled. The number of simulations was chosen so that the estimate of the above mean variance ratios (see later) would have a 95% CI of maximum width 0.01.

In practice, IV analysis is often implemented using Two Stage Least Squares (2SLS) regression. For the IV analyses, the *ivreg* STATA command was used (STATA Release 9 [26]).

6. SIMULATION RESULTS

We consider first the group mean methods for use with a single exposure measurement with reliability λ_2. (Here we may ignore the fact that the 'single' measure was formed from the mean of two independent W values each with reliability λ). Table 1 shows the mean value across simulations of the contrast of exposure $\varepsilon(\overline{\mathrm{W}})$. The contrast increases with corr(Z,X) and λ_2 but decreases with *k*. The correlation between the grouping factor and true exposure, corr(Z,X), has the greatest effect on $\varepsilon(\overline{\mathrm{W}})$. For example, with λ_2 =0.89, and *k*=10, the mean of $\varepsilon(\overline{\mathrm{W}})$ increases from 0.26 to 0.74 as corr(Z,X) increases from $\sqrt{0.2}$ to $\sqrt{0.8}$. It can be shown from (16) that $\varepsilon(\overline{\mathrm{X}}) > \varepsilon(\overline{\mathrm{W}})$ when $k \cdot \varepsilon(\overline{\mathrm{W}}) > 1$. Hence the exposure contrast in X are higher than in Table 1; for example with λ_2=0.67, corr(Z,X)=$\sqrt{0.5}$ and *k*=500 and using the average $\varepsilon(\overline{\mathrm{W}})$ shown of 0.21, (16) suggests that $\varepsilon(\overline{\mathrm{X}}) = 0.31$.

Table 1 Mean, across 20,000 simulations, of observed contrast of exposure between groups as a function of the reliability λ_2 of a 'single' measure $W=(W_1+W_2)/2$, group size *k* and the correlation between the grouping factor Z and true exposure X.

	contrast $\varepsilon(\overline{W})$								
λ_2	*k*=500			*k*=50			*k*=10		
	corr(Z,X)			corr(Z,X)			corr(Z,X)		
	$\sqrt{0.8}$	$\sqrt{0.5}$	$\sqrt{0.2}$	$\sqrt{0.8}$	$\sqrt{0.5}$	$\sqrt{0.2}$	$\sqrt{0.8}$	$\sqrt{0.5}$	$\sqrt{0.2}$
0.89	0.45	0.28	0.11	0.70	0.45	0.19	0.74	0.50	0.26
0.67	0.34	0.21	0.09	0.53	0.34	0.15	0.58	0.40	0.22
0.33	0.17	0.11	0.04	0.28	0.18	0.08	0.34	0.25	0.16

As proven in Section 4, the estimators under the *group mean OLS method* and the *group mean IV method* are identical but biased. Formula (17) together with the results in Table 1, already predict how bias in the *group mean OLS/IV methods* depends on corr(Z,X) and on λ_2: it should decrease both as corr(Z,X) and λ_2 increase. The dependence on *k* is a little more complicated as, in (17), *k* affects bias directly and through $\varepsilon(\overline{W})$. However, when *k* is reasonably large (as in Table 1), the direct effect should dominate and bias should decrease with *k*.

These are the patterns we see in Table 2 (non-parenthesised results) which summarises the behaviour of the estimator $\hat{\beta}_{IV=\overline{W}} = \hat{\beta}_G$ in the simulations of the *group mean* OLS/IV *methods.* The bias accords well with the prediction from (17); for example with *k*=10, corr(Z,X)=$\sqrt{0.2}$, λ_2=0.67 and taking $\varepsilon(\overline{W})$=0.22 from Table 1, the predicted attenuation is 0.85; the average value shown in Table 2 is also 0.85. We also see that bias decreases as *k* increases but the improvement in bias is accompanied by a loss of precision. For example, with reliability λ_2=0.67 and corr(Z,X)=$\sqrt{0.2}$, the attenuation factors for β are {0.85, 0.96, 1.00} for *k* equal to 10, 50 and 500, respectively, while at the same time the SEs are {0.0218, 0.0268, 0.0358}. The RMSEs, which represent a trade-off between bias and precision, are {0.0265, 0.0271, 0.0358}; on this criterion, a grouping with *k*=10 is preferred.

If we adopt low RMSE as the criterion for choosing the best *k*, then we see that, regardless of the correlation, forming groups of 50 subjects gave the lowest RMSE at the lower reliabilities (λ_2=0.33 and 0.67), except when λ_2=0.67 and corr(Z,X)=$\sqrt{0.2}$ in which case, forming groups of 10 subjects gave the lowest RMSE. When λ_2=0.89, groups of 10 subjects gave the lowest RMSE. The results in Table 2 suggest that RMSE is most strongly affected by corr(Z,X), followed by *k* and then the reliability λ_2.

Table 2 Evaluation of *group mean IV* and *group mean ranking IV method* (in parentheses) for bias correction when a single exposure measurement of reliability λ_2 is available and true β=-0.1. The 1st, 2nd and 3rd rows give the mean of the estimates, $\hat{\beta}$, their SD and the RMSE respectively, across 20,000 simulations.

		group mean IV method	(*group mean ranking IV method*)			
λ_2	corr.	*k*=500	*k*=50		*k*=10	
	$\sqrt{0.8}$	-0.100	-0.100	(-0.100)	-0.098	(-0.098)
		0.0173	0.0138	(0.0141)	0.0135	(0.0138)
		0.0173	0.0138	(0.0141)	0.0136	(0.0139)
	$\sqrt{0.5}$	-0.100	-0.100	(-0.100)	-0.098	(-0.098)
0.89		0.0220	0.0174	(0.0178)	0.0165	(0.0169)
		0.0220	0.0174	(0.0178)	0.0166	(0.0170)
	$\sqrt{0.2}$	-0.100	-0.099	(-0.099)	-0.096	(-0.096)
		0.0352	0.0269	(0.0277)	0.0229	(0.0235)
		0.0352	0.0269	(0.0277)	0.0232	(0.0238)
	$\sqrt{0.8}$	-0.100	-0.099	(-0.099)	-0.094	(-0.094)
		0.0175	0.0139	(0.0142)	0.0133	(0.0136)
		0.0175	0.0139	(0.0142)	0.0146	(0.0149)
	$\sqrt{0.5}$	-0.100	-0.098	(-0.099)	-0.092	(-0.092)
0.67		0.0223	0.0175	(0.0180)	0.0162	(0.0166)
		0.0223	0.0176	(0.0180)	0.0181	(0.0184)
	$\sqrt{0.2}$	-0.100	-0.096	(-0.096)	-0.085	(-0.085)
		0.0358	0.0268	(0.0276)	0.0218	(0.0224)
		0.0358	0.0271	(0.0279)	0.0265	(0.0270)
	$\sqrt{0.8}$	-0.100	-0.096	(-0.096)	-0.081	(-0.081)
		0.0185	0.0143	(0.0146)	0.0126	(0.0129)
		0.0185	0.0148	(0.0151)	0.0228	(0.0230)
	$\sqrt{0.5}$	-0.100	-0.094	(-0.094)	-0.074	(-0.074)
0.33		0.0236	0.0178	(0.0183)	0.0149	(0.0152)
		0.0236	0.0188	(0.0193)	0.0300	(0.0301)
	$\sqrt{0.2}$	-0.101	-0.086	(-0.086)	-0.058	(-0.058)
		0.0386	0.0262	(0.0271)	0.0185	(0.0190)
		0.0386	0.0297	(0.0305)	0.0459	(0.0461)

As noted earlier, the group mean method with two groups (here *k*=500) is the same as an IV method with IV equal to ranks -1 and +1 [12]. Hence in this case there is no distinction between the *group mean IV* and *group mean IV ranking methods*. For the other values of *k*, the bias with the *group mean ranking IV method* (Table 2, parenthesised results) was always slightly less than with the *group mean method*, but the improvement was so small as to be unnoticeable in means shown to 3 decimal places. As also predicted, the variance of the *group mean ranking* estimator was increased, but to a greater extent; overall there was an increase in RMSE. The dependence of RMSE on corr(Z,X), *k* and λ_2 followed the same pattern found for the *group mean IV method*.

The true variance, $V(\hat{\beta})$ of the *group mean OLS* and *group mean IV* estimators is given by $SD^2(\hat{\beta})$ in Table 2. But, as shown by (11), the variance estimates under the two methods will not generally be the same. To investigate which method is better, the mean variance ratio was estimated for each method (Table 3). The variance ratio is the ratio of the variance estimate in each simulation to the true variance. A mean ratio closer to 1 could be taken as an indication that a method is better, with perhaps the additional proviso that it should not underestimate the true variance by too much.

Table 3 Mean variance ratios across 20,000 simulations for the *group mean OLS* (first row) and *group mean IV* (second row) methods. The variance ratio is the ratio of the estimate of the variance of $\hat{\beta}$ divided by the true variance.

		Mean Variance Ratios		
λ_2	corr.	*k*=500	*k*=50	*k*=10
0.89	$\sqrt{0.8}$	1.02 1.00	1.01 1.01	1.01 1.01
	$\sqrt{0.5}$	1.03 0.99	1.03 1.00	1.02 1.00
	$\sqrt{0.2}$	1.04 0.99	1.04 1.00	1.04 1.01
0.67	$\sqrt{0.8}$	1.00 1.00	0.99 1.01	1.00 1.01
	$\sqrt{0.5}$	1.00 0.99	1.01 1.01	1.00 1.00
	$\sqrt{0.2}$	1.01 1.00	1.02 1.01	1.02 1.00
0.33	$\sqrt{0.8}$	0.91 1.01	0.92 1.02	0.96 1.02
	$\sqrt{0.5}$	0.91 1.00	0.93 1.01	0.96 1.00
	$\sqrt{0.2}$	0.91 1.02	0.97 1.02	0.99 1.00

In 19 of the 27 scenarios shown, the *group mean IV method* gave a mean ratio closer to one and the lowest value of the ratio from this method was 0.99. In five scenarios – all when λ_2=0.67-0.89 and corr(Z,X)=$\sqrt{0.5}$-$\sqrt{0.8}$ – both methods performed equally well. In one scenario (λ_2=0.67, corr(Z,X)=$\sqrt{0.8}$, *k*=50), both methods are only slightly biased, though in different directions. In the remaining two scenarios – both when λ_2=0.67 and corr(Z,X)$\geq\sqrt{0.5}$ – the *group mean OLS method* performed slightly better. With low reliability (λ_2=0.33), the *group mean IV method* performed better throughout, although with a slight overestimation of variance of up to 2% in some cases. On the other hand the *group mean OLS method* for this λ_2 tended always towards underestimation, especially for large *k* when it was under by 9%.

In the true IV methods – *EVROS IV* and *CS IV* – applicable when two independent measurements are available – bias should in theory be eliminated. The results using these methods (Table 4) show no bias when λ=0.5 or 0.8, but evidence of slight overestimation in both when λ=0.2. The bias is +2% in the *CS IV method*; for the *EVROS IV method,* it varies with corr(Z,X), reaching a maximum of +4% when corr(Z,X) is equal to $\sqrt{0.2}$. Bias is not affected by group size *k* under the *EVROS IV method* but it does affect precision. With the highest reliability, λ=0.8, groups of 10 gave the best precision, while for λ=0.2 and 0.5, the highest precision was found for *k*=50 (except when λ=0.5, corr(Z,X)=$\sqrt{0.2}$, when *k*=10 was best). The impact of *k* on RMSE was similar to that for precision. The limited values of *k* shown do not allow us to show the optimal *k*, but additional simulations for other values (not shown) suggested that optimal *k* increased as λ decreased: it was around *k*=100 when λ=0.2 for example.

The SE of the *CS IV estimator* increased strongly as reliability of exposure decreases, almost doubling, for example, when λ decreases from 0.5 to 0.2. Interestingly, the effect of λ on the SE for the *EVROS IV method* is much weaker: for example when corr(Z,X)=$\sqrt{0.8}$, it increased on average (across the three values of *k*) by 23% for the same change in λ. Low correlation (i.e. poor grouping) had a stronger effect on SE than reliability of exposure for the latter method. When the grouping factor was highly correlated with true exposure (corr(Z,X)=$\sqrt{0.8}$), the *EVROS IV method* performed equally well or better than the *CS IV method* in terms of SE and RMSE for all *k* and λ except when *k*=500 and λ=0.8. The *EVROS IV method* also performed better when corr(Z,X)=$\sqrt{0.5}$ and λ=0.2. Since Table 4 does not necessarily show the optimum value of *k*, it is possible that *EVROS IV* could perform better for other combinations of λ and corr(Z,X).

Table 4 Evaluation of *EVROS IV* and *CS IV* methods for bias correction when two independent exposure measurements, each of reliability λ, are available and the true value β=-0.1. The first row is the mean of the estimates $\hat{\beta}$, across 20,000 simulations, with the SD in parenthesis. The second row is the RMSE.

		EVROS IV method			*CS IV method*
λ	corr.	*k*=500	*k*=50	*k*=10	
0.8	$\sqrt{0.8}$	-0.100 (0.0174) 0.0174	-0.100 (0.0140) 0.0140	-0.100 (0.0139) 0.0139	-0.100 (0.0140) 0.0140
	$\sqrt{0.5}$	-0.100 (0.0222) 0.0222	-0.100 (0.0176) 0.0176	-0.100 (0.0172) 0.0172	
	$\sqrt{0.2}$	-0.100 (0.0354) 0.0354	-0.100 (0.0275) 0.0274	-0.100 (0.0247) 0.0247	
0.5	$\sqrt{0.8}$	-0.100 (0.0179) 0.0179	-0.100 (0.0145) 0.0145	-0.100 (0.0148) 0.0148	-0.100 (0.0183) 0.0183
	$\sqrt{0.5}$	-0.100 (0.0228) 0.0228	-0.100 (0.0184) 0.0184	-0.100 (0.0189) 0.0189	
	$\sqrt{0.2}$	-0.100 (0.0368) 0.0368	-0.100 (0.0293) 0.0293	-0.100 (0.0286) 0.0286	
0.2	$\sqrt{0.8}$	-0.101 (0.0198) 0.0198	-0.101 (0.0165) 0.0165	-0.101 (0.0189) 0.0189	-0.102 (0.0332) 0.0333
	$\sqrt{0.5}$	-0.101 (0.0256) 0.0256	-0.101 (0.0215) 0.0215	-0.101 (0.0259) 0.0259	
	$\sqrt{0.2}$	-0.104 (0.0431) 0.0433	-0.103 (0.0381) 0.0382	-0.104 (0.0470) 0.0472	

An alternative approach to analysing data with two measurements per subject is to first create individual means and then treat these as a single measurement in a 'single measure' method such as the *group mean OLS method* described earlier. As the results in Table 2 were in fact created by averaging two measurements with the same λs as in Table 4, results can be compared directly. The RMSEs for the *group mean OLS method* (Table 2) are lower than those for the *EVROS IV method* in all but two scenarios, both with λ=0.2, *k*=10. (As pointed out by a referee, the use of one measurement as an instrument entails weaker assumptions than treating the two measurements as identically distributed and hence a cost in terms of variance might be expected). However bias could be severe in the *group mean OLS method* while the *EVROS method* is virtually bias-free. Thus, if an investigator prefers an estimator which balances bias and precision, the group mean *OLS method* (with pre-averaging) is, in general, preferred but the *EVROS IV* and *CS IV* methods are better for bias control.

7. APPLICATION OF METHODS IN THE CARBON BLACK STUDY

The aim of the carbon black study was to estimate the effect of an increase in inhalable dust exposure of 1 $mg.m^{-3}$ on worker's lung function, here measured by FEV_1. The exposure of workers was thought to vary by factory - because of differences in type of equipment, emissions, and ventilation - and type of job. Hence two schemes for grouping workers were considered by the original investigators: by job category and by a combination of factory and job category. The number of exposure measures per worker varied but here we restrict analysis to two measurements, W_1, W_2, for 990 workers from 16 factories. For the *group mean OLS,* the *group mean IV* and *group mean ranking IV methods*, the mean of W_1 and W_2 is treated as a 'single' exposure measure per worker. The *EVROS IV method* was applied with W_1 used as the individual exposure and the group mean of W_2 as its instrument, while in the *CS IV method*, W_2 was the instrument for W_1.

Here, in contrast to the simulation study, other covariates – age, height and cumulative smoking category – were included in the regression models and assumed to have a linear relationship with FEV_1. As before, 2SLS regression was used to apply the IV methods. Note that, in the presence of additional covariates measured without error, the first stage of a 2SLS analysis regresses W on the instrumental variable and the other covariates [34] to give predicted values $\hat{W}$; in the second stage, Y is regressed on $\hat{W}$ and the covariates. An adjustment is made to the standard errors in the second regression.

The estimated reliability of a single W was 0.32 and for the mean of W_1 and W_2, λ_2=0.49. Grouping workers by both factory and job gave a higher contrast of exposure $\varepsilon(\overline{W})$ compared to grouping by job alone (Table 5). However the values of $\bar{k} \cdot \varepsilon(\overline{W})$ suggest that attenuation might actually be less in a scheme based on grouping by job alone. Application of formula (17), with λ_2=0.49 and the values $\bar{k} \cdot \varepsilon(\overline{W})$ in Table 5, yields expected attenuation factors equal to 0.97 and 0.79 for job and job and factory combination respectively.

DOI: 10.2202/1557-4679.1087

Table 5 Impact of two grouping schemes on exposure contrasts, using a 'single' $W=(W_1+W_2)/2$, in the carbon black study.

Grouping scheme	$\bar{k}$ +	*No of groups*	$\varepsilon(\overline{\mathrm{W}})$	$\bar{k}\,\varepsilon(\overline{\mathrm{W}})$
Job Category	135	8	0.17	22.88
Factory x Job Category	11	118	0.29	3.17

+ *mean number of workers per group* (rounded to nearest whole number).

The estimated changes in FEV_1 (litres) for 1 $mg.m^{-3}$ change in inhalable dust are shown in Table 6. We compare first the three methods for a single exposure measurement. In contrast to the simulations when there were no additional covariates, the *group mean OLS* and the *group mean IV* estimates are no longer equivalent: the former estimates of the slope were slightly smaller. The difference may be due to the way that covariates are handled in the 2SLS (IV) analysis – see above. Also, the difference in estimated standard errors – which we expect regardless of covariates – is more apparent here: the estimated standard errors are higher under the *group mean IV* analyses for both grouping schemes compared to the *group mean OLS method*.

Table 6 Estimate of the slope of the relationship between FEV_1 and carbon black concentration and (in parenthesis) its standard error, under two grouping schemes and several bias correction methods of analysis.

	OLS analysis	**IV analysis**			
	group mean	*group mean*	*group mean ranking*	*EVROS*	*CS*
Job Category	-0.134 (0.0438)	-0.137 (0.0453)	-0.139 (0.0466)	-0.148 (0.0525)	-0.132 (0.0432)
Factory x Job category	-0.102 (0.0284)	-0.103 (0.0286)	-0.118 (0.0359)	-0.143 (0.0389)	

Comparing the two grouping schemes we see that, when workers are grouped by both factory and job, the relationship between FEV_1 and carbon black concentration appears weaker than when grouped by job alone. As the true value of the slope is not known, we cannot say which estimate is nearer the truth but our earlier estimate of attenuation using (17) suggested that the grouping based on job alone might be less biased. (Also estimates from the latter are closer to those from the true IV methods, *CS IV and EVROS IV*). The *group mean ranking IV method* gave slightly stronger estimates of association, and higher estimated standard errors than the other two methods.

Although the *CS IV* and *EVROS IV* methods are unbiased, in this single application the estimates of association are different, with the *EVROS IV method* giving stronger (i.e. further from zero) associations; their relative precisions depends on the grouping scheme. The estimated standard error from the three grouping methods which first average the two replicates are always smaller than for the EVROS method, and smaller than for the *CS IV* scheme when grouping is based on both factory and job but not for job alone.

8. DISCUSSION

In this paper, we have shown a theoretical equivalence between the *group mean OLS* estimator of a linear regression coefficient and the *group mean IV estimator* when there are no other covariates; however the estimated variances are not the same. If underestimation of variance is to be avoided then the *group mean IV method* is probably to be preferred but neither method was unanimously better. The carbon black application suggested that the relationships between these methods might be less simple when other covariates are present; further work is needed for investigating the properties of the two estimators in this scenario.

The *group mean* is not a true IV – in the sense of satisfying all the assumptions – but the bias in the method can be slight. Therefore, when only a single exposure measurement is available, it seems worth further consideration. By making a link with IV theory, we were able to use it to derive a simple expression (17) which can be used in practice to judge the extent of bias given an estimate of λ. A new insight is that, in choosing between alternative grouping schemes, it is the *product* of group size and contrast of exposure which matters for bias. This has not been recognised by other authors [20, 31] who have argued in terms of contrast of exposure alone.

The link with instrumental variable theory is also important because it clarifies rules by which groups should be formed. Not only should the grouping be strongly correlated with true exposure, but it should not be independently predictive of the health outcome. Thus, if there were important differences in groups, in terms of other risk factors for Y, and these risk factors were not accounted for in the regression analyses, group methods could introduce a new source of bias. Moreover IV theory, and by extension the *group mean OLS approach*, does not require a Berkson error model assumption, as sometimes assumed by occupational hygienists.

A true IV for an error-prone W should be independent of the errors in W. Durbin [12] appeared to recognise this principle, but at the same time suggested the rank of W as an IV if the ranking was “relatively unaffected by the errors”, a situation which cannot be true unless the error is very small. A truly independent ranking sometimes occurs in retrospective occupational studies when there is no measured exposure: hygienists may try to rank or categorise (Low, Medium or High) subjects by likely exposure, taking account of job tasks and their knowledge of industrial processes. This suggests the idea that, in prospective studies also, occupational hygienists could be asked, prior to exposure data collection, to rank subjects. The problem however is that if the ranking is poorly correlated with exposure, IV methods - and hence the *group mean OLS method* - can perform badly [6, 25, 34]. In much of the literature in the field of occupational epidemiology successful grouping of subjects is presumed and thus there is no investigation of the effect of poor grouping. For example, Kim et al. [19] and Burstyn et al. [8] who investigate bias in Cox proportional-hazards and logistic models when using grouping, assume evenly spaced group means as a given. Here we developed a predictive grouping index and investigated its impact on bias and precision.

In this work, the *group mean OLS/IV method* can be seen as an attempt to trade some degree of bias – due to the fact that the IV is correlated with measurement error – for an instrument where there is a stronger expectation of good correlation with the true exposure. Our proposed *group mean ranking IV* method was an attempt to reduce the correlation with measurement error compared to the *group mean IV method*. However, as simulations showed, the attenuation was only slightly decreased, and at the expense of precision. Therefore we do not recommend this or other 'internal' ranking methods. By extension, Durbin's method - equivalent to our *group mean ranking IV* method with k=1 - would be expected to perform even worse since k is too small.

In the new *EVROS IV method*, the value of k does not affect bias but the optimal k for precision appeared to depend on reliability. Further work is needed to derive this. There was no clear winner between this method and the *CS IV method* in terms of precision for the limited values of k shown, but *EVROS IV* appear to be favoured both when corr(Z,X) is strong and when reliability is poor. In occupational and environmental studies, where researchers are often trying to capture long-term exposures, low reliability of the observed exposures may be the norm, so this method is worth consideration. However these IV methods still need to be compared with other non-IV approaches [10] for bias correction using multiple measurements.

Many authors have noted the strong assumptions that IV theory require in order to give consistent estimators and the difficulty of finding suitable variables to act as instruments. However, knowledge and application of instrumental variable theory can, as shown here, provide new insight into methods which may not be 'true' IVs but are useful nevertheless. Our results may also be relevant to the related problem of using aggregate variables to try to control for unmeasured confounders and studies where only aggregate ('ecological') exposure measurements are available [24].

REFERENCES

1. Angrist, J. D. & Imbens, G. W. 1995, "2-Stage Least-Squares Estimation of Average Causal Effects in Models with Variable Treatment Intensity", *Journal of the American Statistical Association*, vol. 90, no. 430, pp. 431-442.
2. Angrist, J. D., Imbens, G. W., & Rubin, D. B. 1996, "Identification of causal effects using instrumental variables", *Journal of the American Statistical Association*, vol. 91, no. 434, pp. 444-455.
3. Armstrong, B. G. 1990, "The effects of measurement errors on relative risk regressions", *Am.J.Epidemiol.*, vol. 132, no. 6, pp. 1176-1184.
4. Bartlett, M. S. 1949, "Fitting A Straight Line When Both Variables Are Subject to Error", *Biometrics*, vol. 5, no. 3, pp. 207-212.
5. Berkson, J. 1950, "Are There 2 Regressions", *Journal of the American Statistical Association*, vol. 45, no. 250, pp. 164-180.
6. Bound, J., Jaeger, D. A., & Baker, R. M. 1995, "Problems with Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogenous Explanatory Variable Is Weak", *Journal of the American Statistical Association*, vol. 90, no. 430, pp. 443-450.
7. Bowden, R. J. & Turkington, D. A. 1990, *Instrumental variables*, 1st paperback ed. edn, Cambridge; New York: Cambridge University Press.
8. Burstyn, I., Kim, H. M., Cherry, N., & Yasui, Y. 2006, "Metamodels of bias in Cox proportional-hazards and logistic regressions with heteroscedastic measurement error under group-level exposure assessment", *Ann.Occup.Hyg.*, vol. 50, no. 3, pp. 271-279.
9. Carroll, R. J. & Stefanski, L. A. 1994, "Measurement error, instrumental variables and corrections for attenuation with applications to meta-analyses", *Stat.Med.*, vol. 13, no. 12, pp. 1265-1282.
10. Carroll, R. J., Ruppert, D., & Stefanski, L. A. 1998, *Measurement error in nonlinear models* Boca Raton, Fla: Chapman & Hall/CRC.
11. Dunn, G. 2004, *Statistical evaluation of measurement errors: design and analysis of reliability studies*, 2nd ed edn, London: Arnold.
12. Durbin, J. 1954, "Errors in Variables", *Review of the International Statistical Institute*, vol. 22, no.1/3, pp.23-32.
13. Fuller, W. A. 1987, *Measurement error models* New York; Chichester: John Wiley & Sons.

14. Gardiner, K., van, T. M., & Harrington, M. 2001, "Respiratory health effects from exposure to carbon black: results of the phase 2 and 3 cross sectional studies in the European carbon black manufacturing industry", *Occup.Environ.Med.*, vol. 58, no. 8, pp. 496-503.
15. Greenland, S. 2000, "An introduction To instrumental variables for epidemiologists", *Int.J.Epidemiol.*, vol. 29, no. 6, p. 1102.
16. Gustafson, P. 2004, *Measurement error and misclassificaion in statistics and epidemiology: impacts and Bayesian adjustments* Boca Raton, Fla.; London: Chapman & Hall/CRC.
17. Johnston, S. C., Henneman, T., McCulloch, C. E., & van der, L. M. 2002, "Modeling treatment effects on binary outcomes with grouped-treatment variables and individual covariates", *Am.J.Epidemiol.*, vol. 156, no. 8, pp. 753-760.
18. Jurek, A. M., Maldonado, G., Greenland, S., & Church, T. R. 2006, "Exposure-measurement error is frequently ignored when interpreting epidemiologic study results", *Eur.J.Epidemiol.*, vol. 21, no. 12, pp. 871-876.
19. Kim, H. M., Yasui, Y., & Burstyn, I. 2006, "Attenuation in risk estimates in logistic and Cox proportional-hazards models due to group-based exposure assessment strategy", *Ann.Occup.Hyg.*, vol. 50, no. 6, pp. 623-635.
20. Kromhout, H. & Heederik, D. 1995, "Occupational epidemiology in the rubber industry: implications of exposure variability", *Am.J.Ind.Med.*, vol. 27, no. 2, pp. 171-185.
21. Martens, E. P., Pestman, W. R., de, B. A., Belitser, S. V., & Klungel, O. H. 2006, "Instrumental variables: application and limitations", *Epidemiology*, vol. 17, no. 3, pp. 260-267.
22. Nieuwenhuijsen, M. J. 1997, "Exposure assessment in occupational epidemiology: measuring present exposures with an example of a study of occupational asthma", *Int.Arch.Occup.Environ.Health*, vol. 70, no. 5, pp. 295-308.
23. Nieuwenhuijsen, M. J. 2003, *Exposure assessment in occupational and environmental epidemiology* Oxford: Oxford University Press.
24. Sheppard, L., Prentice, R. L., & Rossing, M. A. 1996, "Design considerations for estimation of exposure effects on disease risk, using aggregate data studies", *Statistics in Medicine*, vol. 15, no. 17-18, pp. 1849-1858.
25. Staiger, D. & Stock, J. H. 1997, "Instrumental variables regression with weak instruments", *Econometrica*, vol. 65, no. 3, pp. 557-586.
26. StataCorp, L. P. 2005, *Stata programming: reference manual: release 9* College Station, Tex.: Stata Press.

27. Thomas, D., Stram, D., & Dwyer, J. 1993, "Exposure measurement error: influence on exposure-disease. Relationships and methods of correction", *Annu.Rev.Public Health*, vol. 14, pp. 69-93.
28. Tielemans, E., Kupper, L. L., Kromhout, H., Heederik, D., & Houba, R. 1998, "Individual-based and group-based occupational exposure assessment: some equations to evaluate different strategies", *Ann.Occup.Hyg.*, vol. 42, no. 2, pp. 115-119.
29. van Tongeren, M. J. & Gardiner, K. 2001, "Determinants of inhalable dust exposure in the European carbon black manufacturing industry", *Appl.Occup.Environ.Hyg.*, vol. 16, no. 2, pp. 237-245.
30. van Tongeren, M. J., Kromhout, H., Gardiner, K., Calvert, I. A., & Harrington, J. M. 1999, "Assessment of the sensitivity of the relation between current exposure to carbon black and lung function parameters when using different grouping schemes", *Am.J.Ind.Med.*, vol. 36, no. 5, pp. 548-556.
31. van, T. M., Gardiner, K., Calvert, I., Kromhout, H., & Harrington, J. M. 1997, "Efficiency of different grouping schemes for dust exposure in the European carbon black respiratory morbidity study", *Occup.Environ.Med.*, vol. 54, no. 10, pp. 714-719.
32. Wald, A. 1940, "The Fitting of Straight Lines if Both Variables Are Subject to Error", *The annals of Mathematical Statistics*, vol. 11, no. 3, pp. 284-300.
33. Willett, W. C., Sampson, L., Stampfer, M. J., Rosner, B., Bain, C., Witschi, J., Hennekens, C. H., & Speizer, F. E. 1985, "Reproducibility and validity of a semiquantitative food frequency questionnaire", *Am.J.Epidemiol.*, vol. 122, no. 1, pp. 51-65.
34. Wooldridge, J. M. 2006, *Introductory econometrics: a modern approach*, 3rd ed., International student ed. edn, Mason, Ohio: Thomson/South-Western.

DOI: 10.2202/1557-4679.1087