Measurement Error and Model Specification in Determining How Duration of Tasks Affects Level of Occupational Exposure

IGOR BURSTYN*

Community and Occupational Medicine Program, Department of Medicine, Faculty of Medicine and Dentistry, University of Alberta, Edmonton, Alberta, Canada, T6G 2G3

Received 27 October 2008; in final form 4 January 2009; published online 2 February 2009

Statistical modeling of determinants of exposure ascertained in large-scale surveys is an increasingly popular approach to both (i) identifying effective occupational exposure controls that arise in ‘natural experiments’ and (ii) predicting how altering some working conditions may impact exposure levels. This paper sheds light on two underappreciated methodological challenges of such studies. First, I examine the impact of measurement error in the observed determinant of exposure on an investigator’s ability to correctly rank the determinants of exposure in terms of their exposure rate (one aspect of how important a given determinant is). Simultaneously, I consider the issue of whether empirical models fitted for the sake of statistical convenience actually reflect the physical reality that is being modeled and how this may affect the answer to the question about ranking determinants of exposure. These general issues are examined in the context of the ‘time per task’ determinant of exposure and true exposure model that states that exposure is equal to product of exposure rate and duration of a task. Simulation studies were conducted and their conclusions applied in re-examining the data on the impact of duration of some key task on exposure levels to flour dust among bakers. The simulation study demonstrated that bias due to measurement error in observed effects can be either positive or negative. The main conclusion is that the correct ranking of exposure rates can be obtained from both true and poorly specified exposure models, but can be severely distorted by errors in estimates of the duration of tasks performed.

Keywords: determinants of exposure; linear regression; mismeasured covariate; physical model; simulation; statistical model

INTRODUCTION

It is important to accurately identify determinants of exposure that point to efficient interventions and effective occupational exposure control measures. This is often accomplished through relating observed exposure levels to contextual variables by a statistical model. Time spent on a task is one of the potentially critical factors that determines the extent of occupational exposure and is a work characteristic for which occupational hygienists developed a variety of ascertainment methods, including self-report at the end of the workshift, task diaries and direct observations (Burstyn and Teschke, 1999). It is notoriously difficult to obtain accurate measurements of how a person’s working day is split among different tasks (Cardinaels and Labro, 2008); see also literature review of Barrero et al. (2009). A recent experiment on 24 volunteers recalling the duration of manual handling tasks within a 2-h session showed that self-reported duration of tasks can differ by −22% to +38% relative to the true duration, with the error increasing with dullness and discontinuity of the tasks (Barrero et al., 2009). When a duration of task is observed by an occupational hygienist, errors are also inevitable, because key tasks can occur in rapid succession, are typically performed by several subjects observed ‘simultaneously’ (i.e. sequentially in reality with the observer ‘scanning’ workers whose exposures and determinants are measured) and, lastly, start and stop times for tasks are not always easy to define. For example, in the motivating example for this paper of the study of determinants of exposure to dust...
among bakery employees (Burstyn et al., 1997), the precision of the estimate of time spent per task was determined, in large part by the ‘observation time window’ used to ascertain the presence of a task. Specifically, if a task was performed, but an observer missed it in a given 15-min interval, then the total duration of task would be underestimated by 15 min and conversely, if the task was erroneously noted on the observation form, then its duration would be overestimated by the same 15 min—the length of the ‘time window’ during which tasks were observed. Yet in all regression analyses in the occupational hygiene literature to date that relate task duration to observed exposure level, duration of tasks is implicitly assumed to be perfectly measured. This practice violates the fundamental assumption of regression analysis that the independent variable is observed without error, or in some practical sense with negligible error (Zar, 1974). This is a familiar concept to occupational hygienists who deal with exposure misclassification and/or measurement error in an epidemiological context, and the typical consequence of a violation of this assumption is believed to be the attenuation in the estimated regression coefficients (Armstrong, 1998), although bias away from the null is also possible (Jurek et al., 2005). In the specific application considered here, if such a bias is severe, conclusions about which tasks are most influential in conferring elevated occupational exposure can be misleading.

The main goal of this manuscript is to examine the impact of uncertainty in the estimated duration of a task on the magnitude of observed influence on personal exposure. The secondary goal is to consider whether purely empirical regression models of time-per-task determinants of exposure yield meaningful coefficients and whether an alternative/complementary modeling approach is warranted. The conclusions of this theoretical exploration (through simulation studies) are applied to a reanalysis of part of the data that arose in exposure survey of Canadian bakery workers (Burstyn et al., 1997).

**METHODS**

**Exposure model**

Let us assume that a worker performed $N$ mutually exclusive tasks $i \in \{1, 2, \ldots, N\}$ during a working day and that each $i$th task is associated with distinct rate of exposure $Q_i$ and lasts a fixed (but unknown precisely to the investigator) duration $T_i$. The total exposure during such a working day would be

$$X = \sum_{i=1}^{N} Q_i \times T_i. \quad (1a)$$

Note: $Q_i$ is not a random variable, but a parameter in the model and is a feature of $i$th task; there would be only one such value per task.

Let us also assume that there were $J$ workers in an exposure survey, $j \in \{1, 2, \ldots, J\}$, with identical tasks of somewhat varying duration

$$T_{ij} = T_i \times v_{ij}, \quad \text{where } \log(v_{ij}) \sim N(0, \sigma_v^2), \quad (1b)$$

and one can measure workers’ full-shift exposure as $X_j$. NB: the observed exposure for each worker will reflect their individual variation in the tasks’ durations ($v_{ij}$) as well as some random variation from worker to worker, $u_j$. Here, I postulate $\log(u_j) \sim N(0, \sigma_u^2)$; $v_{ij}$’s and $u_j$’s are independent. Log-normality of within- and between-worker distributions is commonly observed in occupational hygiene (Kromhout et al., 1993; Kromhout and Vermeulen, 2001). Thus, the true model of the observed exposure is

$$X_j = u_j \sum_{i=1}^{N} Q_i \times T_i \times v_{ij}. \quad (2)$$

This model is analogous to the one proposed by Nicas and Spear (1993), with a restriction that all $Q_i$’s [= short-term time-weighted averages of Nicas and Spear (1993)] and $v_{ij}$’s are independent (e.g. no autocorrelation in exposure rate due to successive tasks).

If I let $T_{ij}$ denote the observed duration of $i$th task of $j$th worker, then the rate of exposure for each task can be estimated from fitting the following linear regression model to data $(X_j, T_{1j}, T_{2j}, \ldots, T_{Nj})$ and solving it for $Q_i$’s:

$$X_j = \sum_{i=1}^{N} Q_i^* \times T_{ij}^* + \epsilon_j, \quad (3)$$

where $\epsilon_j$ is the residual from fitting the regression model, related to random between-worker variability and the variation in duration of tasks. NB: $Q_i$’s and $Q_i^*$’s are not necessarily equal if $T_{ij}$’s are observed with error (i.e. not equal to $T_{ij}$’s). The measured occupational exposures, $X_j$’s, typically (and by design in this paper) follow a frequency disruption that is well approximated by log-normal. As a result, the regression model fitted to the data imagined above would most likely take the form of

$$\log(X_j/X_0) = \sum_{i=1}^{N} \beta_i \times T^*_i + \epsilon_j, \quad (4)$$

where $\epsilon_j \sim N(0, \sigma_e^2)$ and $X_0$ is a constant (with units of measurement of $X$). Note that equation (4) deviates strongly from equation (3) unless the values $X_j$ are generally greater than the constant $X_0$. Also observe that the mathematical relationships among equations (2), (3) and (4) cannot be readily derived, in part because $\log(a + b) \neq \log(a) + \log(b)$ [i.e. $Q_i^*$ in equation (3) is not equal to $\beta_i$ in equation (4) when $k = l$]. Therefore, equation (4) is a misspecified or poorly specified model in a sense that it does not
represent the process generating exposure that we presume to be true (or know to be true by simulation). Furthermore, in forcing the regression model through the origin in equation (4), it is tacitly assumed that exposure in absence of studied tasks \((T_{ij} = 0 \text{ for all } i)\) is equal to one unit of the scale of measurement of \(X_i\). Lastly, in equation (4) one can ‘acquires an intercept’ by changing units of measurement of \(X\): if for a constant \(c\) we have \(\log(cX_i/X_0)\) instead of \(\log(X_i/X_0)\), then the right-hand side of equation (4) will contain \(-\log(c)\) terms which is an intercept.

Lastly, I need to specify the relationship between the true and observed times devoted to tasks. I will consider the classical additive measurement error model such that

\[ T_{ij} = T_{ij}^* + \tau_{ij} \quad (5) \]

and the measurement error \(\tau_{ij} \sim \text{Uniform}(-t, t)\) for \(t \geq 0\), with the restriction that \(t < \min(T_{ij})\) to ensure that \(T_{ij}^* > 0\); \(\tau_{ij}\)'s are independent of \(T_{ij}\); by properties of uniform distribution: \(E(\tau_{ij}) = 0\) and \(V(\tau_{ij}) = t^2/3\). The exact form of the measurement error model is of course unknown in practice, but it seems sensible to assume that error in the estimate of the duration of each task is less than the duration of the shortest task, is centered on zero and is not concentrated on any specific values over its plausible range.

**Simulation studies**

I consider inhalation exposure and assume that all workers are monitored for the same amount of time using the same sampling rate so that the total volume of air into which an agent is emitted is constant, i.e. ‘m³ of air’ in \(Q_i\)'s is a constant that can be ignored. Let us assume that a typical worker performed four main mutually exclusive tasks during an 8-h working day, each associated with distinct rate of exposure \((Q = \{Q_1, \ldots, Q_4\}; \text{ e.g. in } \mu g \text{ min}^{-1} \text{ m}^{-3})\) and lasting a fixed (but unknown) duration \((T = \{T_1, \ldots, T_4\}; \text{ e.g. in minutes})\). For the purpose of simulation I fix \(Q = \{0.1, 0.2, 1, 2\}\) and \(T = \{140, 200, 100, 40\}\). I deliberately set up the task associated with the highest exposure rate to be of the shortest duration to mimic a realistic situation in which ‘high’ exposures can be conferred by the task with the shortest duration within a workday (rare events). Since exposure rates \(Q\) are the target of inference, they were fixed in the simulations. However, duration of true duration of task for each worker is allowed to vary around the expected value for a typical worker (elements of \(T\)): the observed exposures were generated according to equations (1b) and (2) with \(\sigma^2 = \sigma^2 = 0.01\), noting that examining the effect of these error terms is not my goal at present. I further imagine that the durations of tasks were observed either exactly or with measurement error determined by equation (5) with values of \(t \in \{1, 5, 15, 30\}\), spanning the errors in duration of tasks that are between 1 and 30 min per task. Please recall that \(t\) refers to the maximum absolute value of the variation about zero. These errors span a plausible range: Barrero et al. (2009) observed absolute error in the duration of a 40-min task to be generally within \(\pm 20\) min. Each task is assumed to be measured with the same degree of uncertainty. Finally, I fit both equations (3) and (4) to the data in linear regression with the intercept forced through the origin (Eisenhauer, 2003), using SAS PROC REG (SAS Institute, Cary, NC, USA). In fitting these equations to the data I focus on the magnitude of the effect estimates rather than their formal statistical significance, i.e. variance of the estimate in finite/realistic samples. Therefore, to examine this ‘large sample performance’ of different models under varying degrees of measurement error, I simulated a very large exposure survey with 100 000 workers \((= J)\) for each value of measurement error. Simulation studies were completed before reanalyzing data from the survey of bakeries.

**Application: characterizing tasks associated with exposure to flour dust in bakeries**

The two modeling approaches explored in the simulations are illustrated using data from a study of determinants of flour dust exposure among bakery employees (Burstyn et al., 1997). In that survey, one measurement of full-shift time-weighted average exposure was collected and related to a number of determinants of exposure, including duration of 14 distinct tasks. The duration of tasks was estimated by observing the presence of tasks in non-overlapping 15-min periods: if three tasks were observed in such a period, each was assigned 5-min duration; if only one task was observed in a 15-min window, it was assigned the full 15 min, etc.; total time spent on a task was the sum of durations estimated from the 15-min observation windows. It is clear that such method of estimating duration of tasks is imperfect and errors in the order of 15 min or greater are possible. The reanalysis focuses on five tasks found to be associated with elevated exposures in previously published work. These tasks were

(a) ‘pouring’ of flour (manual or automated), including dusting, ‘forming’ (manipulation of dough’s shape) in
(b) a fully automated process,
(c) with reversible sheeter or
(d) with dough breaker and
(e) ‘determining the weight’ of ingredients in powder form via scale,

with all other tasks combined into one rubric (see the original publication for details).

The data from five subjects who used ‘divider oil’ to eliminate dust was excluded because it proved to
be an extremely effective control measure (28-fold reduction in full-shift time-weighted average exposure to inhalable dust) and its effect is of no interest to current analysis. This data set was chosen because within it I have strong *a priori* reasons to believe that one task in particular, forming with dough breaker, is associated with the highest true exposure rate and it is instructive to observe what happens when the duration of tasks is observed imprecisely and then used in regression analysis to determine the impact of this task on overall exposure relative to others.

**RESULTS**

The simulation study indicates that as the difference between true and observed duration of tasks increases, the estimated exposure rate deviates from its true value to an ever greater extent (Fig. 1). All estimates of exposure rate seem to converge on the same value, which means that the estimates of true ‘high’ exposure rates (Tasks 3 and 4) are increasingly attenuated, while the estimates of true ‘low’ exposure rates (Tasks 1 and 2) are exaggerated. The rank order of tasks in terms of exposure rate is distorted once uncertainty in duration of each task approaches the duration of the shortest task; note that the largest bias is observed for Task 4, which has the highest error of ±15 (= t) min relative to the expected duration of the task (40 min). The pattern of results was not altered when the logarithm of exposure level was modeled in relation to duration of tasks (Fig. 2). Note again that in these analyses the regression coefficients cannot be interpreted as exposure rates. When intercept is considered in fitting equation (4) to simulated data (another form of model misspecification), the same pattern in regression coefficients was observed as in Fig. 2 and the value of intercept increased as the measurement error grew, suggesting that more and more of the between-task differences in exposure was represented by the intercept (function of

![Fig. 1. The impact of measurement error in duration of each of the four tasks (uniformly distributed between −t and t) on bias in estimates of the estimated exposure rate (Q’s in equation (3)); true values appear above t = 0 (no measurement error, true effect observed); each line represents one task.](https://academic.oup.com/annweh/article-abstract/53/3/265/173618)

![Fig. 2. The impact of measurement error on estimation of duration of each of four tasks (uniformly distributed between −t and t) on the estimated regression coefficient with misspecified model [log-exposure is modeled; β’s (min⁻¹) in equation (4)]; true unbiased values appear above t = 0 (no measurement error); each line represents one task.](https://academic.oup.com/annweh/article-abstract/53/3/265/173618)
exposure level common to all tasks) as the information contained in times-per-task variables declined due to measurement error.

In all simulations, owing to the large size of the data set, all coefficients were associated with very small $P$-values for the test of null hypotheses ($Q_i = 0$ and $\beta_i = 0$). However, it must be noted that the test of null hypothesis is of little interest to characterizing the impact of duration of tasks on level of exposure: the inferential target is only the magnitude of the exposure rate. The test of null hypothesis is also not meaningful in the context of estimates of $Q_i$’s because the statistical model used in the evaluation does not have a homoscedastic error term, rendering ‘significance’ tests unreliable (see Hayes and Cai, 2007 for an extensive discussion of heteroscedasticity and heteroscedastic linear regression that can be implemented).

In revisiting the study of determinants of exposure among bakers, it was observed that in the analysis of a subset of the original data, all specific tasks examined were associated, in a statistical sense, with different personal exposures to flour (inhaled) dust ($\hat{P}_i$ in Table 1). Estimated exposure rates were all greater than zero and suggested the following ranking of strength of sources of emission during different tasks: forming with reversible sheeter > forming with dough breaker > pouring/dusting > automated forming > weighing ($\hat{Q}_i$ in Table 1). It must be noted that the standard errors of the estimated emission rates are large relative to the parameter estimates in the case of automated forming and weighing, casting uncertainty on the reliability of the above ranking at lower exposure rates. The two top-ranked tasks are of special note in the context of measurement error in the duration of task. Forming with dough breaker was associated with an 8-fold higher time-weighted average full-shift exposure than forming with the reversible sheeter in the previous analysis in which exposure levels were simply stratified by the presence or absence of this task, a categorization that is free of error in this case (Table 4 in Burstyn et al., 1997). The ranking of forming with reversible sheeter above that of dough breaker is counter-intuitive to anyone who actually observed these machines in operation: dough breaker throws flour dust directly at the worker’s face, while the reversible sheeter moves dough and flour away from the personal breathing zone (see Figs 1 and 2 in Burstyn et al., 1998). Thus, it is tempting to ascribe the reversal in ranking to measurement error in duration of task, of the nature that is suggested by Fig. 1. The same phenomenon may be invoked in explaining the lack of a priori expected ranking in $\hat{P}_i$’s: dough breaker is now ranked behind all tasks except automated forming, which is a conclusion that lacks face validity.

### CONCLUSIONS

This paper illustrates how uncertainty in observed duration of tasks can influence estimates of the impact of duration of within-workday tasks on time-weighted average exposure. Only a limited set of circumstances is explored to draw readers’ attention to the problem without trying to explore all possible consequences of measurement error in duration of tasks. The simulation study demonstrated that bias due to measurement error in observed effects can be either positive or negative. However, it is hoped that the results presented here, especially those showing that ranking of tasks in terms of their ‘importance’ can be affected by measurement error, will motivate occupational hygienists to evaluate the extent to precision of their measures of task duration, so that the impact of this source of error can be either (i) evaluated in simulations or (ii) used to adjust regression results for known extent of error using existing statistical methods (Carroll et al., 1995; Gustafson, 2003).

I considered the issue of model misspecification in conjunction with that of measurement error. One of the apparent drawbacks of the ‘additive’ model used to estimate exposure rates ($Q$) is that correct estimates of statistical significance are not possible when observed exposure levels are not normally distributed (usually leading to violation of the assumption of homoscedasticity of residuals in the linear regression

<table>
<thead>
<tr>
<th>Task (i)</th>
<th>Parameter estimate $\hat{Q}_i$</th>
<th>Standard error of $\hat{Q}_i$</th>
<th>Parameter estimate $\hat{\beta}_i$</th>
<th>Standard error of $\hat{\beta}_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pouring/dusting</td>
<td>0.097</td>
<td>0.033</td>
<td>0.014*a</td>
<td>0.002</td>
</tr>
<tr>
<td>Automated forming</td>
<td>0.025</td>
<td>0.020</td>
<td>0.006*a</td>
<td>0.001</td>
</tr>
<tr>
<td>Forming with reversible sheeter</td>
<td>0.204</td>
<td>0.087</td>
<td>0.014*a</td>
<td>0.006</td>
</tr>
<tr>
<td>Forming with dough breaker</td>
<td>0.168</td>
<td>0.036</td>
<td>0.011*a</td>
<td>0.002</td>
</tr>
<tr>
<td>Weighting</td>
<td>0.011</td>
<td>0.073</td>
<td>0.015*a</td>
<td>0.005</td>
</tr>
<tr>
<td>Other tasks</td>
<td>0.002</td>
<td>0.005</td>
<td>-0.001*a</td>
<td>0.0003</td>
</tr>
</tbody>
</table>

*a with $P < 0.05$
model of the determinants of exposure). Formal testing of statistical significance in the models may be of less interest if a mechanistically meaningful physical model can be fitted to the data instead. This is clearly the case when the impact of task duration on overall exposure is investigated and the rate of exposure is the primary parameter of interest. [See also evaluation of single-cell exposure model among construction painters using survey data (Burstyn and Kromhout, 2002) for an example of empirical modeling where formal tests of the null hypothesis are meaningless.] In more complex situation than that explored in the simulation, it may well be advantageous to first test statistical significance of an association through a purely empirical model [e.g. equation (4)] and only subsequently fit a more interpretable ‘physical’ model to the data using selected determinants of exposure/tasks [e.g. equation (3)]. Measurement error appears to produce similar trends in bias in the correct and misspecified model under the studied conditions. However, as the example of exposure to flour dust in bakeries illustrates, correctly ranking task and thereby setting priorities for intervention should be done with great caution when determinants of exposure are not observed precisely.

If the presence of tasks is dichotomized based on whether it was observed or not, then this dichotomization will certainly inherit some misclassification error from imperfect observations of work practices. However, I hazard to guess that error in ascertaining presence of a task is smaller than that involved in estimation of its duration and hence measurement error bias is likely to be smaller. But in such analysis one can only estimate contribution of a task to total (e.g. full-shift) exposure, not exposure rate associated with a task. The consequence of this is that one can judge the impact of removal of task on overall exposure, but not the impact of reducing the duration or other features of the task that may affect exposure rate.

The main conclusion is that the correct ranking of exposure rates can be obtained from both true and poorly specified (misspecified) exposure models, but can be severely distorted by errors in estimates of the duration of tasks performed.

FUNDING

National Sciences and Engineering Research Council Discovery; Population Health Investigator salary award from the Alberta Heritage Foundation for Medical Research.

Acknowledgements—Dr Irina Dinu provided valuable statistical advice. Michele Hamm proofread and edited the final manuscript.

REFERENCES


