Effects of Data Limitations When Modeling Fatal Occupational Injury Rates

James F. Bena,[1]* A. John Bailer,[1,2] Dana Loomis,[3] David Richardson,[3] and Steve Marshall[3]

Background *Occupational fatal injury rate studies are often based upon uncertain and variable data. The numerator in rate calculations is often obtained from surveillance systems that can understate the true number of deaths. Worker-years, the denominator in many occupational rate calculations, are frequently estimated from sources that exhibit different amounts of variability.*
Methods *Effects of these data limitations on analyses of trends in occupational fatal injuries were studied using computer simulation. Fatality counts were generated assuming an undercount. Employment estimates were produced using two different strategies, reflecting either frequent but variable measurements or infrequent, precise estimates with interpolated estimates for intervening years. Poisson regression models were fit to the generated data. A range of empirically motivated fatality rate and employment parameters were studied.*
Results *Undercounting fatalities resulted in biased estimation of the intercept in the Poisson regression model. Relative bias in the trend estimate was near zero for most situations, but increased when a change in fatality undercounting over time was present. Biases for both the intercept and trend were larger when small employment populations were present. Denominator options resulted in similar rate and trend estimates, except where the interpolated method did not capture true trends in employment.*
Conclusions *Data quality issues such as consistency of conditions throughout the study period and the size of population being studied affect the size of the bias in parameter estimation.* Am. J. Ind. Med. 46:271–283, 2004. *Published 2004 Wiley-Liss, Inc.*[†]

KEY WORDS: fatality undercount; uncertainty; variability; Poisson regression; computer simulation

INTRODUCTION

The emergence of surveillance systems for fatal occupational injuries has facilitated research on which workers face the greatest risk from severe injuries. The National Traumatic Occupational Fatality (NTOF) surveillance system, which is maintained by the National Institute for Occupational Safety and Health (NIOSH), was one of the first data sources to monitor occupational fatal injuries for the entire United States. The NTOF surveillance system uses death certificates marked as "injury at work," and has tracked fatal occupational injuries since 1980. The NTOF system is frequently used to provide numerator data for fatality rate studies. Employment figures from either the Bureau of Labor Statistics' Current Population Survey (CPS) or the Census Bureau's decennial census are typically used to obtain denominators for these rate studies. These fatality rates are often shown descriptively for subgroups of the population [Marsh and Layne, 2001; Bailer et al., 2003], and can be used to determine trends over time.

[1]National Institute for Occupational Safety and Health, Cincinnati, Ohio
[2]Department of Mathematics and Statistics, Miami University, Oxford, Ohio
[3]Department of Epidemiology, School of Public Health, University of North Carolina, Chapel Hill, North California
James F. Bena's present address is Department of Biostatistics and Epidemiology, The Cleveland Clinic Foundation, Cleveland, Ohio 44195.
*Correspondence to: James F. Bena, Department of Biostatistics and Epidemiology/Wb4, 9500 Euclid Avenue, Cleveland, OH 44195. E-mail: jbena@bio.ri.ccf.org

Accepted 4 June 2004
DOI 10.1002/ajim.20070. Published online in Wiley InterScience (www.interscience.wiley.com)

Generalized linear models [McCullagh and Nelder, 1989] are a common method for assessing trends in these rates over time. Bailer et al. [1997] employed Poisson regression models to examine trends in occupational fatal injury rates, and later reported an overall annual decrease in fatality rates of 3.4% for the years 1983–1992 [Bailer et al., 1998]. These publications used NTOF data and employment information from the CPS when constructing rates. More recently, Loomis et al. [2003] found a 3.3% annual decrease in fatality rates for 1980–1996, also using the NTOF surveillance system, but with employment data from the decennial census, coupled with linear interpolation and extrapolation to determine annual employment estimates for non-census years.

The results obtained from these analyses are subject to limitations in the data used. A commonly recognized shortcoming associated with death certificate based occupational fatal injury surveillance systems, such as NTOF, is that they undercount injuries. Stout and Bell [1991] found that an average of 81% of fatal occupational injuries were correctly identified using death certificates alone, based upon results from several studies. Others have noted similar undercounts in other death certificate based systems and for certain subgroups of the population [Murphy et al., 1990; Russell and Conroy, 1991; Peek-Asa et al., 1997].

Estimates of employment used as denominators in fatality rates also have certain features that warrant examination. The CPS is a monthly sample of workers that can be aggregated to form yearly estimates [U.S. Bureau of the Census, 1978]. This survey has the advantage of giving individual yearly estimates, and correctly reflecting seasonal changes in the workforce. However, the relatively small sample in this survey can produce imprecise estimates of employment counts, leading to the potentially unstable rates if a fine level of stratification by variables such as industry or occupation is desired. The decennial census is much larger than the CPS and yields precise estimates even at fine stratification levels. However, since the census is taken only once every 10 years, interpolation or extrapolation is necessary to obtain estimates for non-census years. Also, because the census is taken in April, it does not account for seasonal workers who may not be employed at the time of the census.

The impact of undercounting and denominator precision when using a Poisson regression framework for estimating annual trends in occupational injury rates was examined using a computer simulation experiment. Levels of fatality rates, workforce size, and fatality undercounting were varied in an effort to quantify the impact of errors and uncertainties in the systems currently used to measure changes in occupational fatal injury risk. Poisson regression models applied to rates with undercounted numerators and uncertain and variable denominators were compared to Poisson regression models fit using the "true" numerator and denominator data.

MATERIALS AND METHODS

Regression Model and Parameter Interpretation

A Poisson regression model, a generalized linear model with a log link and a Poisson response distribution, was used to estimate the trend in fatality rates. Observed fatality rates were constructed as O/N, where "O" is number of observed fatal injuries in a subgroup of the population that has "N" person-years of employment. The natural log of fatality rates was modeled as a function of year. The model can be expressed as:

$$\log(\text{fatality rate}) = \beta_0 + \beta_1(\text{year}).$$

To interpret the values of these parameters in terms of fatal injury rates and annual changes in fatality rates, exponentiation is necessary. From the model, e^{β_0} is the estimated fatality rate when year equals zero, and the quantity $(1 - e^{\beta_1})$ is the annual proportion of decline in the fatality rate (when $\beta_1 < 0$). For ease of description, the value of e^{β_0} is referred to as the intercept, and the result of $(1 - e^{\beta_1})$ is called the trend. The biases observed in the estimates were expressed in terms of these exponentiated values.

Simulation Conditions

Data were simulated and models were fit using S-Plus 6.0 (2001). A flowchart of the simulation process and its inputs is given in Figure 1. To perform the simulation, parameters were specified to define population or true values for the fatality rate intercept and trend, as were parameters that characterized initial employment levels and trends in employment over the study period.

Fatality Rate and Trend Inputs

Parameters for fatality rate and trend were chosen to reflect a range of observed results from published studies [Bailer et al., 1998; Loomis et al., 2003]. Three levels of intercept (e^{β_1}) and trend $(1 - e^{\beta_1})$ for the true fatality rate trend model were used to produce nine hypothetical fatality rate situations. The intercept levels reflected initial fatality rates of 1, 5.8, and 20 per 100,000 workers, and the trends evaluated represented annual decreases of 1, 3.4, and 6.4%. Corresponding increasing trends were also evaluated (results not shown), and provided similar results to the decreasing trends.

Employment and Employment Change Inputs

Employment parameters were selected by evaluating employment levels and patterns of change observed for

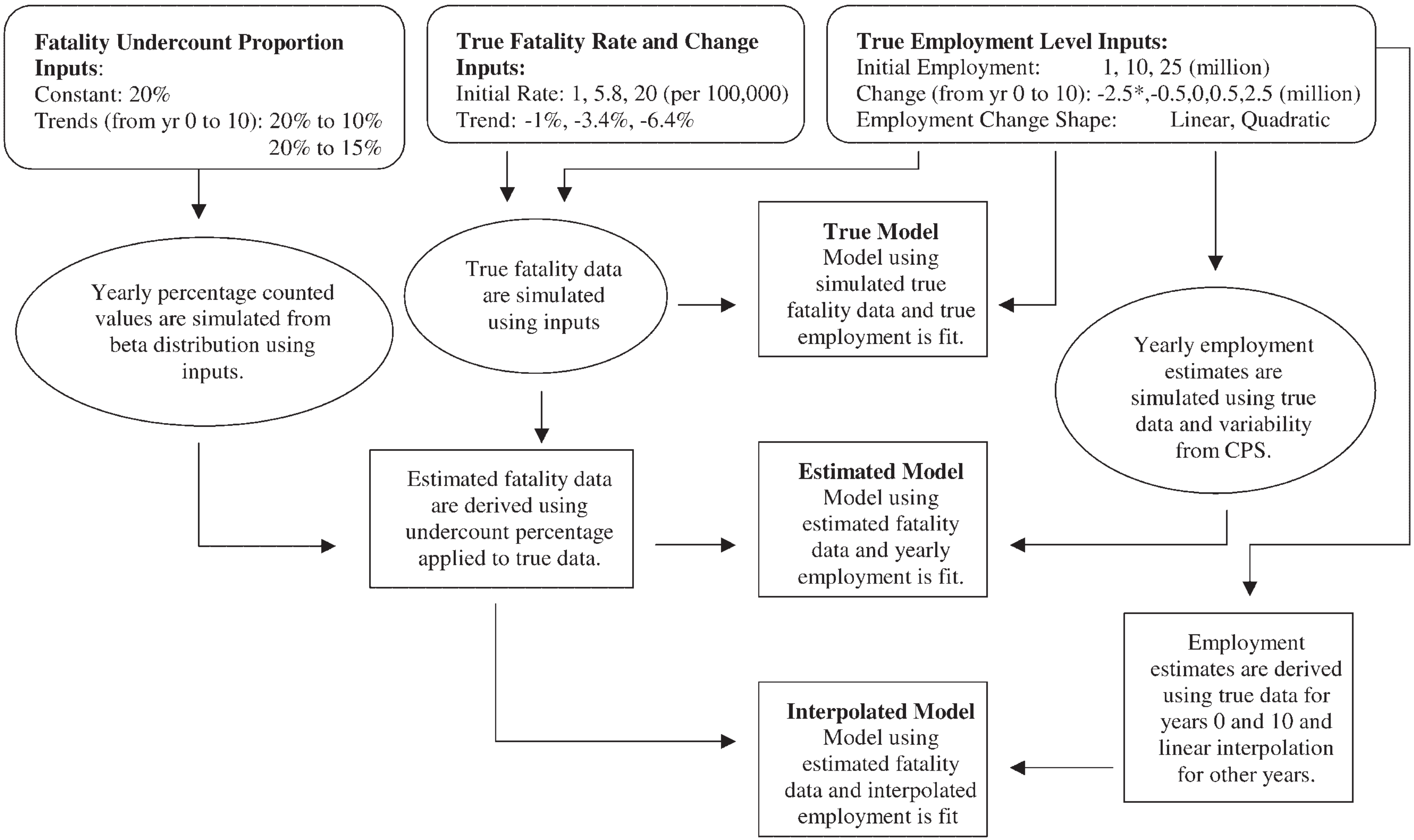

FIGURE 1. Flowchart of inputs to simulation, and process followed.

*-5 million employment decrease was not simulated when the initial employment was 1 million.

subgroups of the population in the CPS dataset between the years 1983–1994. Three levels of initial employment (N = 1,000,000, 10,000,000, and 25,000,000) were considered. Parameter values for linear and quadratic changes in employment over time, representing an increasing and decreasing workforce, were specified to mirror employment trends observed in different subgroups of the population. Parameters reflecting constant employment throughout the study were also specified. For both the linear and quadratic employment change shapes, two sizes of employment change depicting small (500,000 fewer or additional workers employed over an 11-year period) and large (2.5 million fewer or additional workers employed over an 11-year period) shifts in employment in both the positive and negative directions were studied, as was a model with no employment change over the 11 year period. The model with an initial employed population of 1 million and a large negative population change was not used to simulate employment data. In the quadratic employment change models, two sets of parameter values were specified at each combination of initial employment and employment change to show small (25% of the total change in employment in first 5 years) and large (75% of the total change in first 5 years) changes early in the time period studied.

The Simulation Process

Generating fatal injury counts and the undercount process

Using the employment parameters, annual values of true employment for an 11-year period were derived. The number of fatalities was simulated by sampling from a Poisson distribution having a mean defined by the true employment size and fatality rate parameters. Poisson regression models were fit to these data. The years were coded 0 through 10 to allow for interpretation of the exponentiated intercept as the fatality rate in the first year of the study. Year "0" could be thought of as the first year of a trend study, e.g., 1980, and year "10" could correspond to 1990. These results were used as the "true" model results.

To estimate undercounting of fatalities, a value representing the proportion of all occupational fatalities observed was simulated from a beta distribution [Hogg and Craig, 1995]. Initially, the simulation for all years was made from a beta distribution with a mean of 0.8 and standard deviation (SD) of 0.047, closely matching the proportion counted (0.81) suggested by literature [Stout and Bell, 1991]. This provided a probability distribution where 95% of the estimated undercount proportions fell between 0.70 and 0.90. A separate value for the proportion counted was simulated for each year of the study, and these values were multiplied by the annual "true" number of fatalities to determine the annual undercounted fatality total. To examine the effect of changes in the undercount of fatalities, a subset of models was run that evaluated two trends in proportion counted that allowed for the proportion to change from 0.80 (SD = 0.024) to 0.90 (SD = 0.020), and from 0.80 (SD = 0.025) to 0.85 (SD = 0.025) over the 11-year period.

Generating employment estimates

Annual employment was estimated using different methods to simulate the CPS and census derived estimates. For the CPS-based models, the employment was derived by sampling from a normal distribution with the "true" employment as the mean, and a standard error derived using methods described in the CPS survey [U.S. Bureau of Labor Statistics, 1999]. A separate standard error was generated for each year based upon the total employment values using an adjustment for yearly averages, as given in Tables 1-F and 1-H in the source above.

For the census based model, years 0 and 10 were assumed to be years that the census was taken. Standard errors for the estimation of employment in years 0 and 10 were computed assuming the estimates were derived from the 5% public use microdata sets (PUMS). These standard error computations assumed that the population reflected in the employment figures represented 10% of the total population, and did not use a standard error adjustment factor. These standard errors derivations are explained in the technical documentation for the PUMS [U.S. Bureau of the Census, 1983]. For years 1 through 9, linear interpolation was used to estimate employment. Models using the annually estimated employment values, like the CPS, are referred to as CPS-estimated models, while models using linear interpolation to form employment estimates, as is necessary when the census is used, are called Census-interpolated models.

Relative Bias Calculation

During each simulation, a Poisson regression model based on the "true" data was fit, as were a regression model with the undercounted fatality totals and CPS-based employment data, and a regression model with the same undercounted fatality values and Census-based employment totals. Two thousand replications of each simulation condition were run, and the average estimated intercept and trend, along with average standard errors for each estimated parameter, were obtained from each model. Parameter estimates for the CPS-estimated and Census-interpolated models were evaluated against those from the "true" model that uses the fatality counts observed without undercounting and actual employment value in the Poisson regression model. The relative bias of the intercept was defined as the percentage change of the estimated fatality rates in year 0, and was estimated by

$$\text{Bias}\ (\widehat{\beta}_0) = [(e^{\bar{b}_{0A}} - e^{\bar{b}_{0T}})/e^{\bar{b}_{0T}}] * 100,$$

where $\bar{b}_{0A}$ is the average intercept parameter value from an CPS-estimated or Census-interpolated model, and $\bar{b}_{0T}$ is the average intercept parameter value from the ''true'' model.

The difference in the trend was estimated using the relative change in the proportion of decline, defined as:

$$\text{Bias }(\widehat{\beta}_1)\{[(1-e^{\bar{b}_{1A}})-(1-e^{\bar{b}_{1T}})]/[1-e^{\bar{b}_{1T}}]\} * 100,$$

where $\bar{b}_{1A}$ and $\bar{b}_{1T}$ are average trend estimates from the same models described previously. These definitions allow for interpretation of the biases in terms of change in fatality rates.

To assess differences in standard errors between the models, the percentage difference in the average standard errors for the CPS-estimated or Census-interpolated models relative to the true models was examined and the bias was defined for both the intercept and trend estimates as:

$$\text{Bias.}[se(\widehat{\beta}_i)] = \{[se(\bar{b}_{iA}) - se(\bar{b}_{iT})]/se(\bar{b}_{iT})\} * 100,$$

for $i = 0$ (intercept), 1 (trend), where $se(\bar{b}_{iA})$ is the average standard error estimate for parameter i for the CPS-estimated or Census-interpolated model and $se(\bar{b}_{iT})$ is the average standard error estimate for parameter i from the ''true'' model. An empirical check was performed to ensure that the average standard error for each parameter estimate closely matched the SD of the parameter estimate across all simulations. There was very little difference between these two quantities, so the average standard error was used in bias calculations.

RESULTS

Observing the results from all of the simulations run when the undercount parameter was constant at 20% over time, both the CPS-estimated models and Census-interpolated models gave similar results. The negative bias in the intercept estimate was near 21% for both models, correctly reflecting the fact that these models were based upon data that captured only 80% of the ''true'' number of fatalities on average. Both types of models showed small positive biases in the trend estimate, indicating that the models predict a slightly steeper rate of decline than the ''true'' model predicts (Table I, lines 1 & 2). Both intercept and trend estimates had inflated standard error estimates that were nearly 13% larger than those obtained from the model using actual fatal injury counts and ''true'' employment values.

Employment and Fatality Rate Parameter Effect

The size of the population being studied and its general risk of injury both affected the accuracy of the results obtained. By viewing Table I, it is apparent that initial size of the population and initial fatality rate both influenced the intercept bias. Models based on either a small population (1 million) or a group with a very small initial fatality rate (1 per 100,000) had negative intercept biases that were nearly 2% larger than models with larger values for these parameters. Smaller changes in employment also were associated with slightly larger intercept biases. Standard errors of the intercept for these subgroups with elevated negative biases were also slightly larger than those for other subgroups. Results for the initial employment of 25 million were very similar to those obtained with 10 million as a beginning employment, and are not presented.

To further examine how different conditions impacted the estimation of the intercept, the average bias by different levels of employment size and initial fatality rate was also explored. Additional models with more levels of both initial employment and initial fatality rate were fit, holding all other simulation conditions constant. ''True'' employment was assumed to be constant across all years, although the estimate was subject to variability, an annual fatality rate decrease of 3.4% was assumed, and the proportion of deaths observed was sampled for all years from a distribution with a mean 0.8. As can be observed in Figure 2a, the largest intercept bias is apparent when both employment size and initial fatality rate are small. Increases in either parameter greatly reduced this negative bias. It appears that if employment is at least 7.5 million, the intercept bias does not change much for different fatality rate levels, and it remains constant near the amount of undercount present. Figure 2b shows the bias change in the trend estimate. The trend bias behaves similarly to the intercept bias, with reductions in bias seen as the initial employment grows to 5 million, but very little change in bias is seen with employment increases beyond that level. Unlike the intercept, the bias in the trend estimate shrinks to zero with larger worker populations.

Bias in the trend estimate varies slightly across most of the simulation conditions. The most striking differences occurs when employment changes in opposite directions are compared (Table I). Models with increasing employment trends tend to underpredict the rate of decline in the fatality rate, while models that are based on decreasing employment overpredict the fatality rate decline.

Since it is possible that these results are affected by other employment parameters, we explored bias for different levels of initial employment size, employment trend size, and employment trend direction simultaneously (Table II). Initial employment size again impacted the bias greatly; especially for the trend parameter. At the 1 million employment size, the magnitude of biases in both the trend and the standard errors are greater than when employment was 10 million. Estimates of trend overpredicted the decline by nearly 20% when employment is small and decreasing, while the biases in standard errors for the trend in these groups were also larger. Biases in all parameters and standard errors were much smaller and more consistent across different sizes of employment change when the initial employment was 10 million.

TABLE I. Comparison of Simulation Parameter Condition Effects on Intercept and Trend Estimates in CPS-Estimated and Census-Interpolated Models

Simulation condition	Condition value	Model type[a]	Intercept bias	Trend bias	Intercept SE bias	Trend SE bias
All		CPS-estimated	−21.0	1.4	12.6	12.6
		Census-interpolated	−20.9	0.6	12.6	12.7
Initial employment	1 million	CPS-estimated	−21.9	2.7	13.3	13.6
		Census-interpolated	−21.7	0.7	13.4	13.8
	10 million	CPS-estimated	−20.2	0.4	12.0	11.8
		Census-interpolated	−20.2	0.4	12.0	11.9
Employment trend size	No change	CPS-estimated	−21.2	1.3	12.6	13.0
		Census-interpolated	−21.2	1.2	12.6	13.0
	Small change	CPS-estimated	−21.1	3.9	12.8	12.9
		Census-interpolated	−21.1	4.1	12.8	13.0
	Large change	CPS-estimated	−20.7	−1.9	12.3	12.2
		Census-interpolated	−20.5	−4.3	12.4	12.3
Employment trend direction	Increasing	CPS-estimated	−21.1	−2.3	12.5	12.5
		Census-interpolated	−20.9	−4.1	12.6	12.5
	Decreasing	CPS-estimated	−20.8	6.3	12.6	12.7
		Census-interpolated	−20.8	6.7	12.6	12.9
Initial fatality rate	1 per 100,000	CPS-estimated	−22.3	3.3	13.6	14.0
		Census-interpolated	−22.2	2.3	13.7	14.0
	5.5 per 100,000	CPS-estimated	−20.4	0.7	12.1	12.3
		Census-interpolated	−20.3	−0.1	12.2	12.4
	20 per 100,000	CPS-estimated	−20.1	0.2	11.9	11.6
		Census-interpolated	−20.1	−0.5	12.0	11.8
Fatality rate trend	1% annual decrease	CPS-estimated	−21.0	1.1	12.5	12.5
		Census-interpolated	−20.9	−0.9	12.6	12.6
	3% annual decrease	CPS-estimated	−21.0	1.4	12.6	12.6
		Census-interpolated	−20.9	1.0	12.6	12.7
	6% annual decrease	CPS-estimated	−21.0	1.4	12.6	12.6
		Census-interpolated	−20.9	0.6	12.6	12.7

[a]"CPS-estimated" refers to models that used annually estimated employment values to create fatality rates. "Census-interpolated" refers to models that used a linear interpolation of decennial employment data to produce fatality rates.

Employment Estimate Choice

Results for the yearly CPS-estimated models and Census-interpolated models were nearly identical when the "true" employment change over time was linear. In this situation, biases for the CPS-estimated models versus Census-interpolated models were nearly equal for both intercept and trend estimates and their standard errors with linear changes in "true" employment (Table III). However, differences in the two methods were present when "true" employment change was nonlinear (quadratic) over time. The patterns in the CPS-estimated employment models remained similar to those seen with linear employment trends, although the biases in the trend estimate tended to be slightly larger in models with nonlinear trends.

The differences observed between the CPS-estimated and Census-interpolated models with quadratic employment trends depended on the value of other simulation conditions. The models produced similar estimates of intercept bias under most conditions, and standard error biases for the intercept were also comparable. The trend bias appeared to be more susceptible to employment estimate type differences when the "true" employment trend was quadratic. Census-interpolated models tended to underestimate the annual decline relative to the CPS-estimated model when initial employment was small, the employment change size was large, or the annual fatality rate decrease was small. The Census-interpolated model assumed a linear trend, thus when fit to a true quadratic pattern of employment change, this corresponded to model misspecification. Standard error estimates for the trend were also elevated under these conditions for the Census-interpolated model.

To illustrate the differences observed between the two model types when the "true" employment changes quadratically, Figure 3 shows how the model fits are altered when nonlinearity increases. A "true" model with a 5.8 per

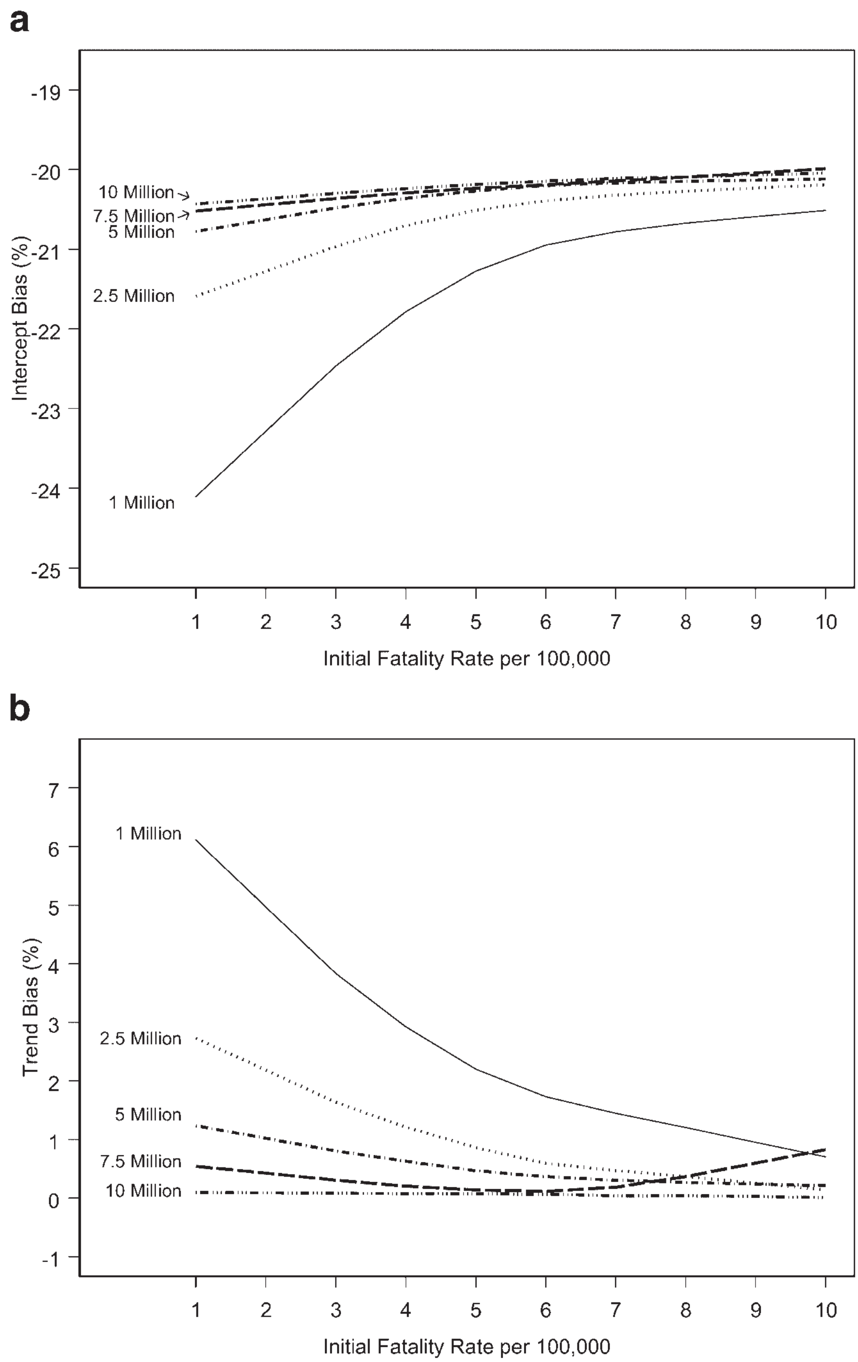

FIGURE 2. Intercept (**a**: top figure) and trend (**b**: bottom figure) bias plots of initial employment by initial fatality rate. True employment was constant across all years, and an annual fatality rate decrease of 3.4% was assumed for all models.

100,000 initial fatality rate, a fatality rate decline of 3.4%, and an initial employment of 1 million was fit. Figure 3a shows a small quadratic increase in employment, and as a result, the differences observed in the initial fatality rate and the trend are small. As seen in Figure 3b, if the nonlinearity grows, the Census-interpolated model fit changes. Underestimation of employment early in the study period causes overestimation of the fatality rate in the Census-interpolated model, and leads to an intercept estimate that is very close to the "true" model intercept, despite having an undercount in fatalities. Later in the time frame, as the rate of employment change decreases, the fatality rates in the CPS-estimated and Census-interpolated models become similar. Thus, the trend estimate for the Census-interpolated model reflects the

TABLE II. Estimates of Bias by Initial Employment, Employment Trend, and Employment Trend Direction

Initial employment	Employment trend	Employment trend direction	Model type[a]	Intercept bias	Trend bias	Intercept SE bias	Trend SE bias
1 million	0.5 million total change	Increasing	CPS-estimated	−22.1	−2.4	13.3	13.4
			Census-interpolated	−22.1	−2.7	13.3	13.4
		Decreasing	CPS-estimated	−21.8	17.5	13.7	14.8
			Census-interpolated	−21.8	18.5	13.7	14.6
	2.5 million total change	Increasing	CPS-estimated	−21.8	−7.0	13.4	13.2
			Census-interpolated	−21.2	−14.1	12.9	12.8
10 million	0.5 million total change	Increasing	CPS-estimated	−20.2	0.4	12.0	11.8
			Census-interpolated	−20.2	0.4	12.0	11.8
		Decreasing	CPS-estimated	−20.2	0.3	12.0	11.8
			Census-interpolated	−20.2	0.3	12.0	11.8
	2.5 million total change	Increasing	CPS-estimated	−20.2	0.0	12.0	11.7
			Census-interpolated	−20.2	−0.1	12.0	11.9
		Decreasing	CPS-estimated	−20.2	1.2	12.1	12.2
			Census-interpolated	−20.2	1.2	12.1	11.9

[a]"CPS-estimated" refers to models that used annually estimated employment values to create fatality rates. "Census-interpolated" refers to models that used a linear interpolation of decennial employment data to produce fatality rates.

change in fatality rate estimates caused by error in the employment estimation, as well as the actual decline in fatality rates, which leads to error in its fit.

Effect of Undercounting Fatal Injuries

To assess the reliance on the assumption that the proportion of deaths undercounted is constant over time, a subset of models were fit where the average undercount was constant, but differed from year to year due to sampling variability. These models were chosen to reflect a variety of fatality rate and employment input situations. Since the complete set of conditions was not performed, the relative bias in the CPS-estimated and Census-interpolated models differed slightly for both the intercept and trend estimates, but this does not impact comparisons of bias size across different levels of undercounting change over time.

Models with no change in undercounting over time were compared to models run under the same fatality rate and employment conditions where the proportion of deaths not counted decreases by 5 and 10% over the entire time frame. As Table IV displays, the intercept bias became more negative and standard error biases decreased as the undercounting estimate over time became smaller. While both initial employment and initial fatality rate were predictors of differences in intercept bias, neither condition enhanced nor diminished the differences seen in intercept biases as the undercounting gets smaller.

Trend estimates were substantially affected by the change in the proportion counted over time. The annual decline estimate decreases by more than 20% when the proportion of deaths counted grows by 5% over the time span (from 80 to 85% counted). This grows to close to 70% when the change in proportion changes by 10% (from 80 to 90% counted). The difference in bias between the estimated and interpolated models grows slightly as the change in proportion counted grows. Interpolated models showed a 3–5% larger bias in the trend than estimated model. Larger differences were observed when initial employment or initial fatality rate was small.

As a visual reference, Figure 4 shows the true model fit along with estimated model fits when the undercounting is constant, when the proportion not counted decreases by 5% (from 20 to 15% not counted) and when the proportion not counted decreases by 10% (from 20 to 10% not counted). The model displayed had constant 10 million employment, sampled with variability, an initial fatality rate of 5.8 per 100,000, and an annual fatality rate decrease of 3.4%. While the intercepts for all of the estimated models are about 20% lower than the "true" model, the decreasing trend becomes less steep as the change in undercounting proportion becomes larger.

DISCUSSION

The situations studied here for both employment and fatality rate trends were based on real data situations encountered in previous research. The choice to study only log-linear trends in fatality rates follows previous data analysis methods. The inputs chosen reflect results for subgroups of the population that have been produced elsewhere [Bailer et al., 1998; Loomis et al., 2003]. It is unlikely that every

TABLE III. Estimates of Bias for Intercept and Trend Parameters by Employment Trend Shape. Bias Estimates are Presented Overall, by Employment Size, Employment Change Size, and Fatality Rate Trend Size

Simulation parameters	Employment trend shape	Model type[a]	Intercept bias	Trend bias	Intercept SE bias	Trend SE bias
All simulations	Linear	CPS-estimated	−21.0	1.3	12.6	12.6
		Census-interpolated	−21.0	1.2	12.6	12.6
	Quadratic	CPS-estimated	−20.9	1.5	12.5	12.6
		Census-interpolated	−20.8	0.2	12.6	12.8
Initial employment 1 million	Linear	CPS-estimated	−21.9	2.5	13.3	13.7
		Census-interpolated	−22.0	2.3	13.3	13.7
	Quadratic	CPS-estimated	−21.9	2.8	13.3	13.6
		Census-interpolated	−21.6	−0.3	13.5	13.9
Initial employment 10 million	Linear	CPS-estimated	−20.2	0.3	12.0	11.7
		Census-interpolated	−20.2	0.3	12.0	11.7
	Quadratic	CPS-estimated	−20.2	0.5	12.0	11.9
		Census-interpolated	−20.2	0.5	12.0	12.0
Employment trend 0.5 million total change	Linear	CPS-estimated	−21.1	3.8	12.8	12.9
		Census-interpolated	−21.1	3.7	12.8	12.9
	Quadratic	CPS-estimated	−21.1	4.0	12.7	12.9
		Census-interpolated	−21.1	4.3	12.8	13.0
Employment trend 2.5 million total change	Linear	CPS-estimated	−20.7	−2.1	12.3	12.0
		Census-interpolated	−20.7	−2.2	12.3	12.0
	Quadratic	CPS-estimated	−20.7	−1.9	12.3	12.3
		Census-interpolated	−20.4	−5.4	12.4	12.5
Fatality rate trend 1% annual decrease	Linear	CPS-estimated	−21.0	0.8	12.6	12.6
		Census-interpolated	−21.0	0.4	12.6	12.6
	Quadratic	CPS-estimated	−20.9	1.3	12.5	12.4
		Census-interpolated	−20.8	−1.7	12.6	12.6
Fatality rate trend 3% annual decrease	Linear	CPS-estimated	−21.0	1.4	12.6	12.6
		Census-interpolated	−21.0	1.4	12.6	12.6
	Quadratic	CPS-estimated	−20.9	1.4	12.5	12.7
		Census-interpolated	−20.8	0.7	12.6	12.8
Fatality rate trend 6% annual decrease	Linear	CPS-estimated	−21.0	1.7	12.6	12.7
		Census-interpolated	−21.0	1.7	12.6	12.7
	Quadratic	CPS-estimated	−20.9	1.7	12.6	12.8
		Census-interpolated	−20.8	1.5	12.7	13.1

[a]"CPS-estimated" refers to models that used annually estimated employment values to create fatality rates. "Census-interpolated" refers to models that used a linear interpolation of decennial employment data to produce fatality rates.

subgroup of the population conforms to these model forms, but such models do give easily interpreted results, and accurately reflect the changes over the entire range of the data, despite possibly being inappropriate for certain subintervals of the time span. Changes in fatality rates over time, such as those caused by regulation changes or improvements in technology, may follow a log-linear pattern as studied here. However, such changes could also result in a single adjustment in the fatality rate, which may not be appropriately represented by the log-linear model. Misspecification of the fatality rate model form could also introduce additional bias.

Employment change within every subgroup cannot be expected to be monotonic, and more extreme patterns of employment change can alter the accuracy of analyses that rely on assumptions about linear changes in employment over time. However, linear interpolation of employment appears to be a parsimonious choice that is appropriate when some of the actual yearly employment estimates are unknown. Since the interpolated employment estimates do not react to large yearly changes in either direction for non-census years, it is as likely to underestimate yearly employment as it is to overestimate that quantity. Use of linear extrapolation to estimate employment beyond the last known

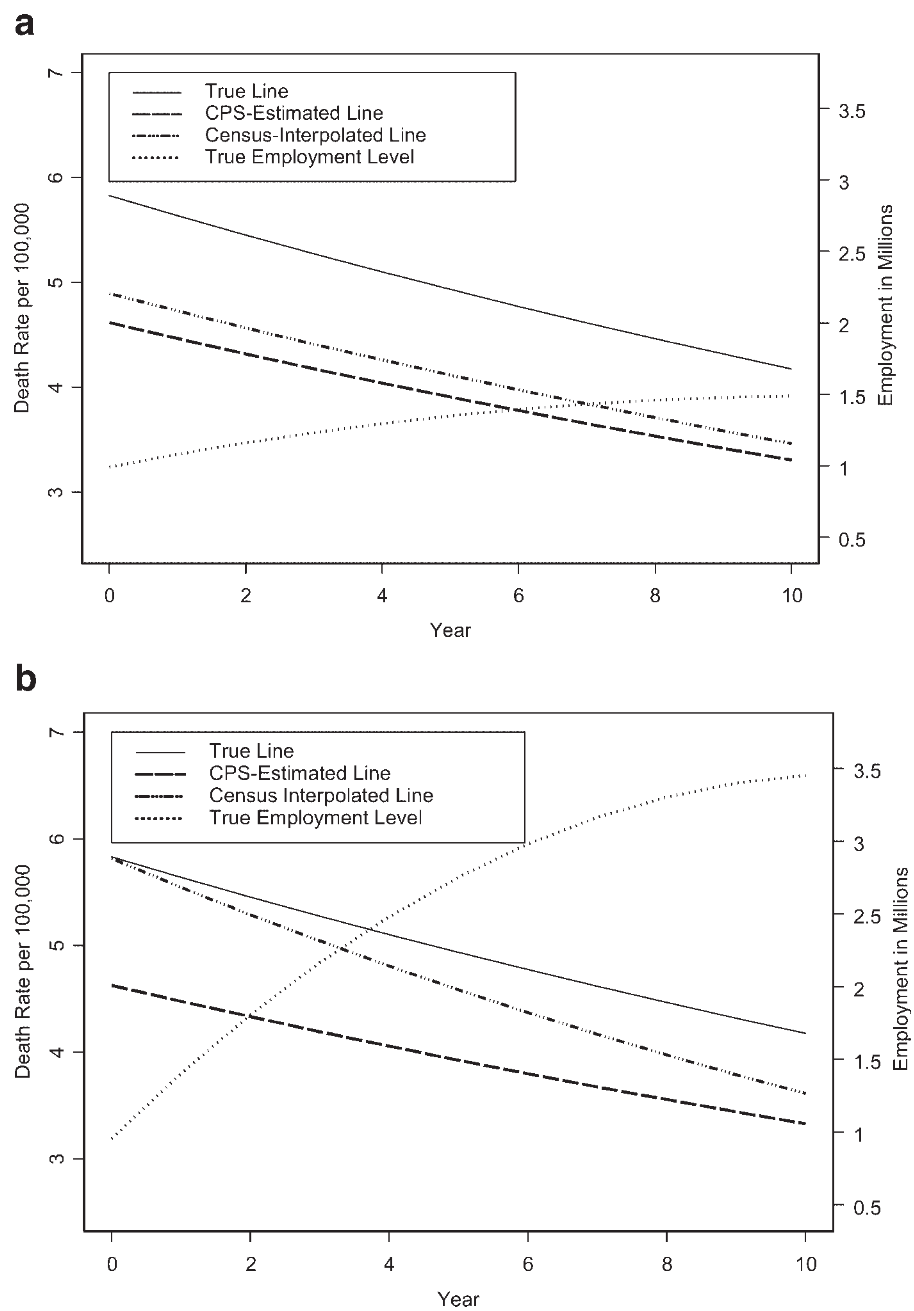

FIGURE 3. Comparison of interpolated and estimated models with nonlinear trends in employment. **a**: (top) corresponds to a small degree of nonlinear employment change, while **b** (bottom) corresponds to a larger degree of nonlinearity. Refer to the right vertical axis for interpreting employment changes over time and to the left vertical axis to examine fatality rate changes over time. All models assumed a 5.8 per 100,000 initial fatality rate, an annual fatality rate decline of 3.4%, and initial employment of 1 million.

data value, though not studied here, has similar disadvantages to interpolation. However, the errors could be larger than those seen using interpolation because changes in employment after the last known data point are not directly measured.

Interpolated estimates from the census fail to incorporate seasonal changes in the workforce because employment information is taken in April of the census year. While this may cause employment in certain sectors to be underestimated, the effect would likely be similar to that observed by the undercounted fatalities. If the proportion of seasonal workers not counted remains constant over time, the intercept would be affected, but estimates of trend should remain unchanged.

TABLE IV. Comparison of Simulation Parameter Condition Effects on Intercept and Trend Estimates in CPS-Estimated and Census-Interpolated Models Under Varying Undercount Conditions and a Subset of Simulation Conditions

Simulation condition	Condition value	Undercount change	Model type[a]	Intercept bias	Trend bias	Intercept SE bias	Trend SE bias
All simulations		Constant	CPS-estimated	−21.0	2.4	12.6	12.7
			Census-interpolated	−18.3	6.7	12.1	12.1
		5% Decrease	CPS-estimated	−21.1	−27.3	12.0	11.2
			Census-interpolated	−18.3	−22.6	11.5	10.5
		10% Decrease	CPS-estimated	−21.6	−71.5	11.2	8.9
			Census-interpolated	−18.9	−66.5	10.7	8.2
Initial employment	1 Million	Constant	CPS-estimated	−21.8	4.7	13.3	13.6
			Census-interpolated	−16.9	13.0	12.4	12.5
		5% decrease	CPS-estimated	−21.9	−25.3	12.6	11.9
			Census-interpolated	−16.9	−16.3	11.7	10.8
		10% decrease	CPS-estimated	−22.4	−69.1	11.8	10.0
			Census-interpolated	−17.4	−59.4	10.9	8.8
	10 Million	Constant	CPS-estimated	−20.2	0.2	12.0	11.8
			Census-interpolated	−19.7	0.4	11.8	11.6
		5% decrease	CPS-estimated	−20.3	−29.3	11.3	10.5
			Census-interpolated	−19.8	−28.9	11.2	10.2
		10% decrease	CPS-estimated	−20.9	−74.0	10.7	7.9
			Census-interpolated	−20.3	−73.5	10.6	7.6
Initial fatality rate	1 per 100,000	Constant	CPS-estimated	−21.0	2.8	12.6	12.7
			Census-interpolated	−18.2	12.9	12.1	12.1
		5% decrease	CPS-estimated	−21.1	−58.8	11.9	10.9
			Census-interpolated	−18.2	−47.9	11.4	10.3
		10% decrease	CPS-estimated	−21.7	−151.0	11.2	8.7
			Census-interpolated	−18.8	−139.5	10.7	8.1
	5.5 per 100,000	Constant	CPS-estimated	−21.0	2.2	12.6	12.5
			Census-interpolated	−18.3	4.3	12.1	11.9
		5% decrease	CPS-estimated	−21.1	−15.9	12.0	11.2
			Census-interpolated	−18.3	−13.4	11.5	10.5
		10% decrease	CPS-estimated	−21.7	−42.5	11.2	9.0
			Census-interpolated	−18.9	−39.7	10.7	8.1
	20 per 100,000	Constant	CPS-estimated	−21.0	2.3	12.6	13.0
			Census-interpolated	−18.4	2.9	12.1	12.3
		5% decrease	CPS-estimated	−21.1	−7.2	12.0	11.5
			Census-interpolated	−18.3	−6.5	11.5	10.8
		10% decrease	CPS-estimated	−21.6	−21.1	11.3	9.2
			Census-interpolated	−18.9	−20.2	10.8	8.5

[a]"CPS-estimated" refers to models that used annually estimated employment values to create fatality rates. "Census-interpolated" refers to models that used a linear interpolation of decennial employment data to produce fatality rates.

The effects of seasonal workforces were studied in a companion paper, which used empirical employment data from the CPS and decennial census along with fatality data from NTOF [Richardson et al., 2004]. Fatality rates differed by as much as 10% for industries such as Agriculture/Forestry/Fishing and Construction that have varied employment levels throughout the year. The magnitude of the trend estimates also showed some differences in these industries, but the direction of the trend was unaffected, except in cases where the observed trend was very small.

Underreporting of fatalities is only one drawback to using death certificate based surveillance systems. Limitations common to these systems, including NTOF, are described elsewhere [Jenkins et al., 1993; Marsh and Layne, 2001]. Changes to data collection methodology within the study period that alters the proportion of all fatalities that are observed, such as the introduction of standardized guidelines

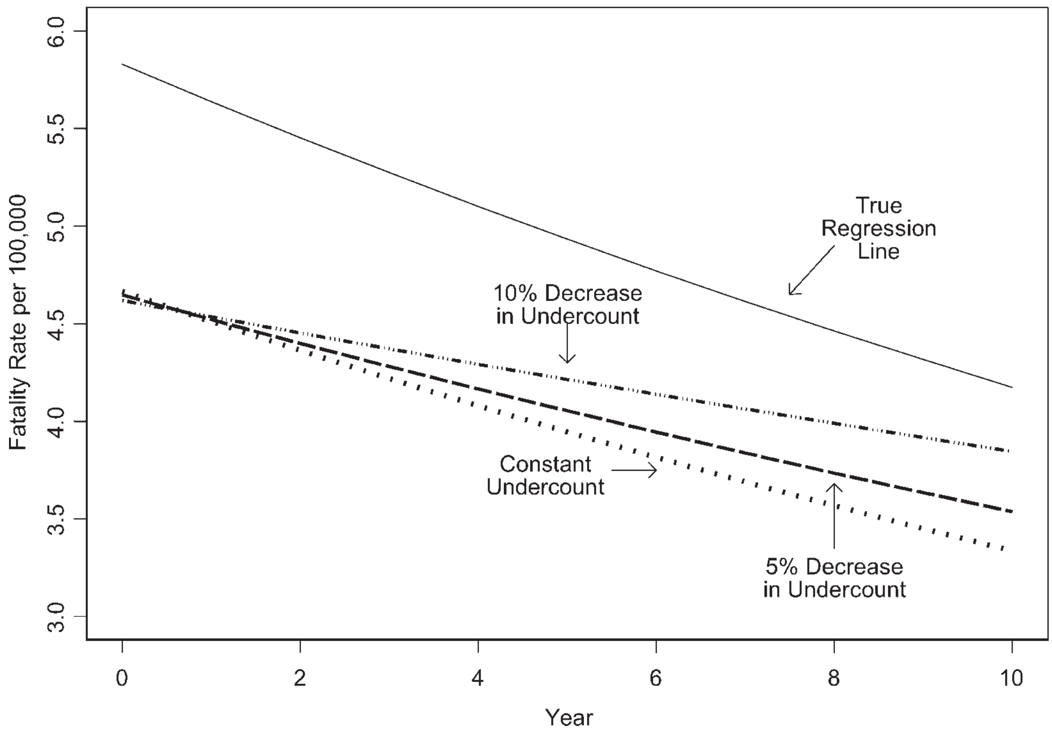

FIGURE 4. Comparison of estimated lines under different assumptions about the proportion undercounted over time. True fatality rate lines and CPS-estimated fatality rate lines with different levels of undercounting are presented. All models presented have constant 10 million employment, an initial fatality rate of 5.8 per 100,000, and an annual fatality rate decline of 3.4%.

for coding the injury at work item on death certificates in 1992, could affect results. A decreasing trend in underreporting can lead to underestimation of both the intercept and trend estimate for fatality rates. Thus, the quality of results obtained using death certificate based systems is strongly linked to issues regarding data collection.

Multiple source fatal occupational injury systems, such as the Bureau of Labor Statistics' Census of Fatal Occupational Injuries (CFOI) that was started in 1992, are alternatives to death certificate based systems. A description of the CFOI system is given in Austin [1995]. Recent comparisons of the CFOI and NTOF systems have shown that NTOF captured an average of 84% of the number of deaths in the CFOI system for the time period 1992–1994 [Biddle and Marsh, 2002]. However, since the CFOI system is a more recent source, suitable amounts of data to perform fatality rate trend analysis are only now becoming available.

The study of the effects of undercounting fatalities and uncertainties in the employment estimates allows conclusions to be drawn regarding the direction and magnitude of potential errors in estimating occupational fatality rates and trends. The undercount of fatalities (or employment) affects mainly the estimated intercept, if the undercount is constant over time.

Models using estimated employment values and interpolated employment values are very similar when employment changes follow a linear, or nearly linear, trend. As the change in employment becomes more nonlinear, bias in the estimation of the trend parameter grows for the Census-interpolated model. Results for the trend estimate bias seemed to be close to zero for most of the situations studied, and appear to unrelated to undercount present, if the effect of the undercount is consistent over time. Changes in the undercounting of fatal injuries over time (i.e., indicating a change in the proportion of fatal injuries observed over time) caused increases in the trend estimate bias.

These findings indicate the importance of data quality for reliable results when assessing trend in fatal injury rates. Estimates of the trend are accurate under many varying conditions, as long as those conditions remain consistent throughout the study period. Changes in data quality, such as an increase in the proportion of fatalities observed, can affect results, and should be documented whenever possible. Most models with small employment and fatality counts, which might be found in finely stratified analyses based on subgroups of the workforce, tended to have larger biases than those seen in models with larger employment and fatality counts. Researchers should be aware that biases are increased when fine levels of stratification are performed, and that more accurate results may be obtained when larger populations are studied.

ACKNOWLEDGMENTS

We thank Suzanne Marsh, John Preisser, Bob Park, and Lorann Stallones for comments and suggestions that have improved the quality and clarity of presentation. We also thank Matt Wheeler for development of an early form of the computer simulation, on which the simulations used in this project were based.

REFERENCES

Austin C. 1995. An evaluation of the census of fatal occupational injuries as a system for surveillance. Comp Work May:51–54. See http://www.bls.gov/iif/oshcfoiarchive.htm or http://www.bls.gov/iif.oshwc/c far0007.pdf for more information.

Bailer AJ, Reed LD, Stayner LT. 1997. Modeling fatal injury rates using Poisson regression: A case study of workers in agriculture, forestry, and fishing. J Safety Res 28:177–186.

Bailer AJ, Stayner LT, Stout NA, Reed LD, Gilbert SJ. 1998. Trends in rates of occupational fatal injuries in the United States (1983–92). Occup Environ Med 55:485–489.

Bailer AJ, Bena JF, Stayner LT, Halperin WE, Park RM. 2003. External cause-specific summaries of occupational fatal injuries. Part I: An analysis of rates. Am J Ind Med 43:237–250.

Biddle EA, Marsh SM. 2002. Comparison of two fatal occupational injury surveillance systems in the United States. J Safety Res 33:337–354.

Hogg RV, Craig AT. 1995. Introduction to mathematical statistics. 5th edn. New Jersey: Prentice Hall.

Jenkins EL, Kisner SM, Fosbroke DE, Layne LA, Stout NA, Castillo DN, Cutlip PM, Cianfranco R. 1993. Fatal injuries to workers in the United States, 1980–1989: A decade of surveillance—National and state profiles. DHHS (NIOSH) Number 93-108S, Morgantown, WV.

Loomis D, Bena JF, Bailer AJ. 2003. Diversity of trends in occupational injury mortality in the United States, 1980–1996. Inj Prev 9: 9–14.

Marsh SM, Layne LA. 2001. Fatal injuries to civilian workers in the United States, 1980–1995. DHHS (NIOSH) Number 2001-129S, Cincinnati, OH.

McCullagh P, Nelder J. 1989. Generalized linear models. 2nd edn. New York: Chapman & Hall.

Murphy DJ, Seltzer BL, Yesalis CE. 1990. Comparison of two methodologies to measure agricultural occupational fatalities. Am J Public Health 80:198–200.

Peek-Asa C, McArthur DL, Kraus JF. 1997. Determining injury at work on the California death certificate. Am J Public Health 87:998–1002.

Richardson D, Loomis D, Bailer AJ, Bena J. 2004. The effect of rate denominator source on US fatal occupational injury rate estimates. Am J Ind Med 46:261–270 (this issue).

Russell J, Conroy C. 1991. Representativeness of deaths identified through the injury-at-work item on the death certificate: Implications of surveillance. Am J Public Health 81:1613–1628.

S-PLUS 6 for Windows Programmer's Guide. 2001. Insightful Corporation, Seattle, WA.

Stout N, Bell C. 1991. Effectiveness of source documents for identifying fatal occupational injuries: A synthesis of studies. Am J Pub Hlth 81: 725–728.

U.S. Bureau of Labor Statistics. 1999. Explanatory notes and estimates of error. Employment Earnings 46(3):162–179.

U.S. Bureau of the Census. 1978. The current population survey: Design and methodology. Technical Paper 40.

U.S. Bureau of the Census. 1983. Census of Population and Housing, 1980: Public-use Microdata Samples Technical Documentation, U.S. Department of Commerce, Bureau of the Census, Washington, DC.