Web Appendix 1: Computer programs

A. SAS code to account for outcome misclassification in a Poisson model using direct maximum likelihood and known values of sensitivity and specificity

The SAS code below is applied to the dataset with the 120,010 person-years summarized into 4183 strata of distinct covariate patterns. In the code below, lam = λ_j, mu = μ_j, wj = w_j, dj = d_j, and nj = n_j. In this example, sensitivity is assumed to be 85% and specificity is assumed to be 95%.

```
title "rate ratio accounting for outcome misclassification (se=&se, sp=&sp)";
proc nlmixed data=tabled gconv=1e-15 fconv=1e-15;
        parms b0=-5 b1=.5 b2=0 b3=0  b4=0  a0=-5 a1=0 a2=0 a3=0  a4=0;
        se=0.85; sp=0.95;
        lam=exp(b0+b1*asbestos+b2*sex+b3*log(age)+b4*year);
        mu=exp(a0+a1*asbestos+a2*sex+a3*age+a4*year);
        lik=(lam*se+mu*(1-sp))**(wj)*( lam*(1-se)+mu*sp)**(dj-wj)*exp(-(lam*se+mu*(1-
sp)+ lam*(1-se)+mu*sp)*nj);
        model nj~general(log(lik));
run;
```

Web Appendix 2: Simulation Study

We used simulation to explore the finite sample properties of using the modified maximum likelihood estimates to account for outcome misclassification. Simulations were performed with sensitivity and specificity assumed to be known. The simulations were intended to mimic the data from the cohort of textile workers exposed to asbestos in South Carolina. Let *i* index simulated participants in each stratum of distinct covariate patterns (i =1, … ,n_j), where n_j is the number of participants in stratum *j,* and *X* represent exposure ranging from 0 to 500 (mean = 45, standard deviation = 28). The time to death due to lung cancer (R) and time to death due to other causes (S) followed exponential distributions with means determined by the exposure value. In expectation, a 100-unit increase in exposure decreased the time to lung cancer (R) by one-half and the time to non-lung cancer death (S) by one-third. The total time (T) contributed by each record was the minimum of R and S.

Cause of death was represented by δ. If death due to lung cancer occurred before death due to other causes would have occurred ($R < S$) then δ was set to 1. Otherwise, if death due to other causes occurred before death due to lung cancer ($S < R$), then δ was set to 2. Simulated participants were censored after 5 years; for participants with $T > 5$, δ was set to 0.

Error-prone cause of death indicator δ^* was generated based on δ and values of sensitivity and specificity. We simulated five possible scenarios with varying degrees of outcome misclassification: 1) both sensitivity and specificity set to 1; 2) specificity set to 0.95 and sensitivity set to 0.9; 3) specificity set to 0.95 and sensitivity set to 0.6; 4) both sensitivity and specificity set to 0.9; and 5) specificity set to 0.9 and sensitivity set to 0.6. In

all scenarios, outcome misclassification was nondifferential with respect to exposure and other measured covariates. For each scenario, δ^* was sampled from a Bernoulli distribution with probability determined by sensitivity and specificity. Where δ = 1, the probability that δ^*= 1 was equal to the value of sensitivity; where δ = 2, the probability that δ^*= 1 was equal to 1 – specificity. If $\delta^*$was not drawn to be 1 (lung cancer death), $\delta^*$was set to 2 (other death). If T > 5 years, then δ^* was set to 0.

Each scenario was simulated 10,000 times. For each simulated cohort, we summarized the data into J strata of distinct covariate patterns following the same categorization used for the actual data and calculated two counts for each stratum: y_j, the sum of all actual lung cancer deaths in each stratum, $\sum_{i=1}^{n_j} I(\delta_i = 1)$, and w_j, the sum of all reported lung cancer deaths in each stratum, $\sum_{i=1}^{n_j} I(\delta_i^* = 1)$. We used Poisson regression to estimate the rate ratio of the lung cancer death per 100-unit increase in exposure. We estimated the true rate ratio (using y_j as the count of lung cancer deaths) and the standard rate ratio (using w_j as the count of lung cancer deaths) with standard methods. We then compared these results to results using the method described above using modified maximum likelihood to account for outcome misclassification by setting values of sensitivity and specificity.

We evaluated the performance of this method to account for outcome misclassification by comparing bias and 95% confidence interval coverage between the standard analysis using w_j as the count of lung cancer deaths and the analysis using modified maximum likelihood to set values of sensitivity and specificity. Bias was defined as 100 times the difference between the average estimated log rate ratio and true log rate ratio, and confidence interval coverage was calculated as the proportion of simulations in

which the estimated Wald-type confidence limits included the true value. The bias-precision tradeoff was considered through examination of the mean-squared error, which was the sum of the square of the bias and the square of the standard deviation of the bias.

In Web Table 1, we compare the average standard errors of the natural log of the rate ratios with the standard deviations of the natural log of the rate ratios across the 10,000 simulations to evaluate the performance of the variance estimator. The similarity between the mean standard error and the standard deviation of the natural log of the rate ratios indicates that our variance estimator is appropriate in this setting.

Web table 1. Comparison of average standard errors and standard deviations of point estimates from 10,000 simulated cohorts [a]

Scenario	Method	Mean β	Mean standard error	Standard deviation of β
1. Specificity = 1, Sensitivity = 1	Truth	0.695	0.087	0.088
2. Specificity = 0.95, Sensitivity = 0.9	Standard ML	0.639	0.091	0.092
	Modified ML	0.691	0.094	0.095
3. Specificity = 0.95, Sensitivity = 0.6	Standard ML	0.612	0.110	0.114
	Modified ML	0.698	0.110	0.114
4. Specificity = 0.9, Sensitivity = 0.9	Standard ML	0.588	0.089	0.091
	Modified ML	0.696	0.097	0.098
5. Specificity = 0.9, Sensitivity = 0.6	Standard ML	0.542	0.108	0.112
	Modified ML	0.700	0.115	0.121

[a] The models accounting for imperfect sensitivity and specificity did not converge in 6, 7, 9, and 5 simulated cohorts for scenarios 2,3, 4, and 5, respectively.