Ann. Occup. Hyg., Vol. 50, No. 6, pp. 573–582, 2006

doi:10.1093/annhyg/mel018

Modeling, Estimation and Validation of Cotton Dust and Endotoxin Exposures in Chinese Textile Operations

GEORGE ASTRAKIANAKIS[1]*, NOAH S. SEIXAS[1], JANICE E. CAMP[1], DAVID C. CHRISTIANI[2], ZIDING FENG[3], DAVID B. THOMAS[3,4] and HARVEY CHECKOWAY[1,4]

[1]Department of Environmental and Occupational Health Sciences, University of Washington, Seattle, WA, USA; [2]Harvard School of Public Health, Harvard University, Boston, MA, USA; [3]Program in Epidemiology, Fred Hutchinson Cancer Research Center, Seattle, WA, USA; [4]Department of Epidemiology, University of Washington, Seattle, WA, USA

Received 4 September 2005; in final form 14 February 2006; published online 21 April 2006

In support of an epidemiological study of cancer incidence among cotton textile workers in Shanghai, PRC, historical estimates of exposure to cotton dust and endotoxin were developed for subjects drawn from a cohort of 267 400 female textile workers. A large dataset of historical cotton dust measurements were available from 56 of the study factories. Using these data, a series of models were developed to estimate cotton dust exposure by year, factory and process. Model estimates were validated by comparing with independently collected measurements gathered over a 15 year period and indicated a low relative bias (<2%) and relative accuracy (±61%). Endotoxin exposures were estimated using the endotoxin content of cotton dust by major processes obtained in five separate surveys in these factories. The validated exposures were assigned to the 7242 jobs held by the 3812 study subjects. Among the exposed workers, the mean cumulative exposure levels were 113.8 mg m^{-3}*years for cotton dust and 6707.7 EU m^{-3}*years for endotoxin, respectively. The overall correlation among cotton dust and endotoxin exposures for these subjects was $r = 0.58$. Using an unusually rich set of historical cotton dust measurements, along with independently collected exposure measurements for validation and conversion to endotoxin levels, validated estimates of cumulative exposure were constructed for this large case-cohort study in the textile industry.

Keywords: cotton dust; endotoxin; modeling; retrospective exposure; validation

INTRODUCTION

Over the last 30 years, significant progress has been made to advance the quality of exposure assessments in occupational studies. The increased availability of measurement data (Blair and Stewart, 1992; Blair *et al.*, 1999), an appreciation of importance of average exposures over time (Rappaport, 1991a; Rappaport and Smith, 1991) and the adoption of rigorous statistical techniques have increased the confidence with which such data can be incorporated into epidemiological research (Hornung, 1991; Kromhout *et al.*, 1991; Rappaport, 1991b; Seixas and Sheppard, 1996). Regardless of the amount of exposure data available and the manner in which they are organized, it is always desirable to attempt exposure assessment validation (Bouyer and Hemon, 1993; Stewart, 1999). In a recent review of their assessment methods for acrylonitrile exposure, Stewart *et al.* emphasize the necessity of such a step, pointing out that the lack of a validity evaluation reduces the usefulness of the study and opens the investigators to criticism (Stewart *et al.*, 2003); and more to the point, may undermine the inferences drawn from the study.

Past epidemiological studies investigating lung cancer among textile workers in China have been population-based case–control studies and relied on job title as a surrogate of exposure (Levin *et al.*, 1987, 1988; Wu-Williams *et al.*, 1993; Wang *et al.*, 1995).

*Author to whom correspondence should be addressed. Tel: +1-778-328-8017; fax: +1-778-328-8001; e-mail: george@ohsah.bc.ca

Similarly, exposure assessments for occupational studies of cancer incidence and mortality from all causes among textile workers outside China have also been limited to surrogate dose metrics, such as duration of employment (Berry and Molyneux, 1981; Merchant and Ortmeyer, 1981; Kuzmickiene *et al.*, 2004). This paper describes the development of quantitative estimates of exposure to cotton dust and endotoxin for members of a cohort of workers in the textile industry in Shanghai. The estimates were evaluated against independently collected cotton dust measurements from two of the factories.

METHODS

Data development and modeling

This study is based on a cohort of workers from 526 textile factories in the Shanghai Textile Industry Bureau (STIB) who enrolled in 1989–1991 (Thomas *et al.*, 2002). Employment records were maintained at the factories, or in the case of those that had closed down since study inception, at the offices of the STIB or at local offices of the Shanghai Municipal Center for Disease Control (SMCDC), formerly known as the Shanghai Anti-Epidemic Bureau. Industrial hygienists employed by the STIB collected factory information from participating factories using a standardized factory profile form (FPF). Along with general information on factory age, size and production processes, the industrial hygienists used the FPFs to record exposure measurement data from factory records and from government inspection reports.

Approximately 4700 fixed-location dust measurement records were abstracted from records in 93 of the 503 factories in which women in this study had worked. The measurements were originally collected during the course of regular surveys required by the SMCDC for compliance with occupational health regulations. They included measurements of cotton, wool, synthetic fiber and silica dust among other exposures. The standard method of dust collection was the high volume Chinese dust sampler (CDS) described previously (Astrakianakis *et al.*, 2006). This high-volume 'total' dust sampler operates at 20 l.p.m. using open-faced 40 mm PVC filters.

The exposure assessment for the case–cohort study included several steps and is described in detail below. In summary, the historical data were reviewed for inconsistencies and obvious errors. The predictive model for cotton dust exposure was developed followed by the validation of the estimates for two study factories. Based on a series of decision rules, exposure levels for cotton dust were assigned to all jobs associated with cotton manufacturing in these factories. Exposure levels for endotoxin were then assigned based on the product of cotton dust concentrations and endotoxin content of the airborne dust. Finally, the cumulative exposure to cotton dust and endotoxin were calculated for each subject according to all jobs held.

The STIB hygienists coded the abstracted major and specific processes corresponding to each historical measurement according to standardized process and job dictionary developed for this study (Wernli *et al.*, 2005). The dictionary contained codes for 18 major and 166 specific manufacturing processes in the textile industry. The recorded exposure measurements included 69 data points representing very high (> 25 mg m^{-3}) dust concentrations. Among these, two transcription errors and five miscoding errors (silica measurements entered as cotton dust measurements) were identified and corrected, but no errors were found for the remaining 62 entries. Concentrations of cotton dust >25 mg $m^{-3}$are possible in dust collection areas or in other special operations, but are unlikely to reflect the levels experienced by workers. All 62 of these high measurements data points were removed. The mean and median values for the removed measurements were 100 and 49.5 mg m^{-3}, respectively, and ranged from 25.1 to 849 mg m^{-3}. An additional 22 measurements were removed from the database because identical measurements were listed for two different dates; the measurements for the earlier date were maintained.

Thirteen measurements were listed as below the detection limit (<DL), and were replaced with $DL/\sqrt{2}$ mg m^{-3}. The DL was set at 0.1 mg m^{-3} (the lowest measurement provided by the balance used by the STIB laboratory (Sartorius, 1997) and a total volume of 0.4 m^3 (20 min at 20 l min^{-1}).

All descriptive statistics, correlations and model development steps were carried out using Stata 8.2 for Windows (Intercooled Stata for Windows, 2004). The measurements were positively skewed and therefore log-transformed for the purpose of description and modeling. Seven variables were examined for inclusion in a linear model for cotton dust concentration:

- Factory (FID)—Categorical variable for each factory
- Major Process (MP)—Categorical variable for 18 major processes, all unexposed administrative processes were reassigned to a single category
- Specific Process (SP)—Categorical variable for 166 specific tasks
- Sample Year (SY)—Continuous variable for the year in which the measurement was collected. The range of values includes 1975 to 1999.
- Season (S)—Categorical, based on the month in which samples were collected as follows
  Winter: December through March
  Spring/Fall: October, November, April and May
  Summer: June through September

- Year Factory Opened (FO)—Categorical; based on the first year of production for each factory and coded as <1930, 1930 to ⩽1950 and >1950.
- Number of workers at Peak Year (NW)—Categorical according to quintiles of the size of peak employment workforce, coded as <1425, 1426–5200, 5201–7250, 7251–7865 and >7865.

All seven variables were added to the linear model sequentially in a forward modeling procedure. Variables were kept in the model if their coefficients were statistically significantly different from zero ($P < 0.05$) and if the addition of the variable significantly increased the proportion of variance responded to by the model (r^2). Interaction terms were considered but were not included for the sake of parsimony.

Model validation

For external validation of the model, the predicted concentrations were compared with a set of independently collected fixed-position cotton dust measurements obtained over 15 years at two of the factories in the current study (Olenchock *et al.*, 1983; Kennedy *et al.*, 1987; Christiani *et al.*, 1993, 1999). However, because the predicted cotton dust exposures were estimated in terms of the CDS concentrations and the data used for validation were measured using a vertical elutriator (VE), the CDS estimates were first converted to a VE-equivalent concentration using a conversion model described previously (Astrakianakis *et al.*, 2006). Briefly, the CDS is a high-volume open-faced total dust sampler operated at 20 l.p.m. and the VE is an inhalable dust sampler with a designed cut-point of 15 μm aerodynamic diameter operated at 7.4 l.p.m. Side-by-side sampling was performed at three Shanghai textile mills using the CDS and VE sampling devices in November 2002. The ratios of CDS to VE measurements were between 2 and 10 depending on process. A linear model was developed based on the sampling results to convert CDS measurements into VE equivalents and incorporated the specific processes where sampling took place.

$$\ln(\mathrm{VE}) = \beta_0 + \beta_{\mathrm{CDS}} \ln(\mathrm{CDS}) + \sum \beta_{\mathrm{SP}_i} \mathrm{SP}_i + \varepsilon, \tag{1}$$

where: ln (VE), log of dust concentration as measured by the VE; ln (CDS), log of dust concentration as measured by the CDS; and SP (Specific Process), specific production process.

Direct comparisons of model estimates and the previously collected exposure data were possible for six specific processes at two textile mills collected during four surveys, or potentially 48 exposure categories. However, in one factory the process of spinning was not sampled in all 4 surveys, and in another factory the process of combing was not sampled for 1 survey. Therefore, for each of the 43 exposure categories the model estimate was compared with the mean on the available measurements. The differences between the predicted and measured dust concentrations were used to determine the bias (‘*b*’, mean of the difference), precision (‘*p*’, standard deviation of the difference) and accuracy $(b^2 + p^2)^{1/2}$ of the exposure estimates (Hornung, 1991). Comparisons were made for specific processes within each mill and are reported as relative bias, relative precision and relative accuracy, where relative statistics were calculated as the bias (or precision or accuracy) divided by the measurement mean multiplied by 100. A smaller value of each of these statistics indicates better performance of the model.

Calculation of cumulative exposure for study subjects

Estimated cotton dust exposures from the model were assigned to each subject’s work history using a set of decision rules. For the lung cancer case–cohort study, the work histories for 3812 subjects were collected from 462 of the 503 factories. The subjects held 7242 jobs. Jobs identified in factories where cotton was not used ($n = 2661$) were assigned zero exposure. For 1803 jobs, the primary model was used to estimate the process, factory and time-specific exposure level. For the remaining 2778 jobs where some of the predictors were not available, a series of secondary predictive models were developed using a restricted set of predictors. The order of application of the secondary models was guided by the strength of the model (r^2) and by inspection, noting which predictor was omitted after each step.

There were no measurements of endotoxin in the historical data obtained from the factories in this study. Information regarding the source of the cotton used by each factory was requested but details regarding raw cotton purchases were discarded after 2 years. Consequently, estimates of exposure for the current study were approximated based on measurements of the endotoxin content of airborne cotton dust collected from several Shanghai textile factories over 15 years (Olenchock *et al.*, 1983; Kennedy *et al.*, 1987; Christiani *et al.*, 1993, 1999) and endotoxin measurements obtained for this study (Astrakianakis *et al.*, 2006). Average endotoxin content of airborne dust for seven specific processes was derived from a total of 765 samples from five factories obtained during these five surveys. The geometric mean endotoxin content (EU mg^{-1} cotton dust), by major process, was multiplied by the predicted cotton dust levels. Maintenance jobs were assessed at specific endotoxin levels according to the production areas where they were assigned, such as machine maintenance in roving. All maintenance jobs assigned to non-production areas were considered unexposed to cotton dust and endotoxin. All administrative jobs in non-production processes and jobs in non-cotton factories were considered unexposed to endotoxin.

RESULTS

Dust concentration data were available for 2475 cotton dust samples collected from 56 of 244 cotton factories of the total 503 study factories. The cotton dust samples were collected between 1975 and 1999. An additional 29 samples that were labeled 'mixed fiber', but had come from cotton processes, were added to the cotton dust data, bringing the total number of measurements to 2504. After review and correction of errors and removal of measurements collected outside usual work environments, the historical data used to develop the predictive model included 2413 results from 56 cotton factories, with the number of samples per factory ranging from 1 to 329.

Exposure concentrations as measured by the CDS by major processes are given in Table 1. The GM was 2.0 mg m^{-3} and the GSD 2.7. Among major processes with at least 10 measurements, the GM ranged from 1.6 to 4.8 mg m^{-3}. The measurements represented 10 of 13 major processes and 39 of 166 specific processes. There were no measurements for the Warehouse and for Cutting, Sewing and Garment Manufacturing (major processes 1 and 10).

The historical measurements of cotton dust were collected from factories representing 19% of the cotton factories in this study. The source of data was generally evenly distributed among factories in terms of the size of the work force, with somewhat fewer measurements obtained from larger factories. The distribution ranges from 22% of the measurements from factories with the smallest peak workforce (<1000 workers) to 15% from the largest factories (~10 000 workers). About 75% of the samples were from factories opened before 1930, 17% from factories opened between 1930 and 1950, and the remaining 8% were from operations opened since 1950.

The available CDS measurements were collected from 1975 to 1999. Geometric mean dust concentration levels fall from 4.0 to 1.1 mg m^{-3} over the 24 year period (Fig. 1). The data were collected from factories where the main fiber used in manufacturing is cotton (99%). In 22 of 56 factories, cotton was used exclusively and in 23 factories synthetic fibers were used as a secondary fiber in addition to cotton. The remaining factories utilized fibers other than synthetic as secondary fibers. Although some non-cotton fibers were used in some processes, the air samples were designated 'cotton dust' and are therefore believed to be valid cotton measurements.

Each of the variables, Major Process (MP), Specific Process (SP), Factory ID (FID), Sample Year (SY), Season (S), Year Opened (FO), Peak Number of Workers (NW), was considered for inclusion in a model predicting the natural log of the cotton dust concentration. Of these, FID, SP, MP and SY were included in the final predictive model. The addition of season, although statistically significant, was not practical because it was not available from the subjects' work history data. None of the other variables considered improved the model's fit. The proportion of the variance explained by the model is $r^2_{adj} = 0.32$. An equation representing the primary predictive model is given below.

$$\ln(\mathrm{CDS})_p = \beta_0 + \beta_{SY}\mathrm{SY}_i + \sum \beta_{FID_i}\mathrm{FID}_i + \sum \beta_{MP_i}\mathrm{MP}_i + \sum \beta_{SP_i}\mathrm{SP}_i + \varepsilon, \quad (2)$$

where: ln $(\mathrm{CDS})_P$, log-dust concentration as measured by the CDS; SY (Sample Year), 1975–1999; FID (Factory ID), factories 1–56; MP (Major Process), major production processes 1–10; SP (Specific Process), specific production process 1–39.

Table 1. Historical cotton dust measurements by major process[a,b]

Dictionary code	Major process	*n*	GM (mg m^{-3})	GSD (mg m^{-3})	Min (mg m^{-3})	Max (mg m^{-3})
2	Material handling	504	2.5	2.6	0.1	24.7
3	Fiber processing	159	2.7	2.3	0.2	23.0
4	Spinning	840	1.6	2.4	0.1	25.0
5	Bleaching	69	4.8	2.6	0.1	20.0
6	Dyeing	6	7.9	2.1	3.3	20.0
7	Textile finishing	102	4.5	2.2	0.6	23.6
8	Weaving	689	1.7	2.8	0.1	25.0
9	Printing	18	3.4	2.5	0.5	12.9
11	Finishing	18	1.9	2.2	0.6	7.6
12	Quality control	8	1.5	2.6	0.4	7.2
	Total	2413	2.0	2.7	0.1	25.0

[a]All historical measurements were collected using the Chinese dust sampler (CDS).
[b]Summary based on historical data used to develop the predictive model.

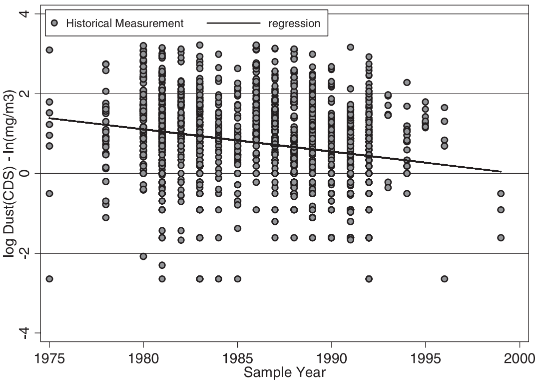

Fig. 1. Dust concentration versus Sample Year (n = 2413 samples).

For illustration purposes, the coefficients for 10 of the 13 major manufacturing processes, for seven of the 39 specific processes and for the two factories used in the validation of the model are shown in Table 2.

The predicted cotton dust exposures were compared with measurement data collected by Christiani *et al*. (1993, 1999) (Olenchock *et al*., 1983; Kennedy *et al*., 1987). The 797 individual measurements collected over four surveys were collapsed into 43 mean exposures by SY, FID, SP and MP; 19 for Factory 1 and 24 for Factory 2. The measurement means represented six specific processes and three major processes. For comparison the measurement means were compared with the corresponding predictions from the models. The CDS-equivalent predictions were converted to VE-equivalent levels using equation (1) in order to perform the comparison.

The comparison of model predictions with the independent dataset is given in Table 3, and plotted in Fig. 2. The overall relative accuracy of the model was 61% and was influenced largely by a lack of precision in the estimates. The model slightly underestimated exposure, with bias of −2%, and was moderately correlated with the measurements ($r_P = 0.59$, $P < 0.0001$; $r_{Spearman} = 0.61$).

The model results in estimates that overestimate exposure for the specific process roving (Rel. Acc. = 121%). The scatter plot demonstrates this result where the slope of the regression line is much shallower than expected. At measured concentrations <ln (1 mg m^{-3}), the predictive models will overestimate exposure, whereas the converse is true at higher measurements. The relative imprecision of the estimates in roving was low (29%) compared with the imprecision overall of 61% for the model. Apart from roving, estimates are predicted most accurately in opening (48%) and least accurately in spinning (69%). Notably, the predictions for spinning are closer to the point where the regression line would ideally pass.

Table 2. Predictive model for cotton dust (ln mg m^{-3}, as measured by the CDS)[a]

Variables	Df	MS error	F-test	P-value
Model	100	8.2	12.5	<0.0001
Sample Year	1	25.4	38.8	<0.0001
Factory ID	55	6.1	9.4	<0.0001
Specific Process	36	2.8	4.4	<0.0001
Major Process	8	1.8	2.7	0.0054
Residual	2312	0.65		

	Coefficients	Standard error
Constant	65.0	10.3
Sample Year	−0.03	0.01
Major Process		
Material Handling	0.44	0.23
Fiber Processing	0.50	0.29
Spinning	0.12	0.20
Bleaching	0.07	0.24
Dyeing	0.49	0.46
Textile Finishing	0.23	0.20
Printing	−0.42	0.26
Finishing	−0.59	0.26
Quality Control	1	
Weaving	reference	
Specific Process[b]		
Opening	−0.05	0.24
Carding	0.08	0.23
Drawing	−0.12	0.40
Combing	−0.17	0.31
Roving	0.14	0.29
Spinning	−0.02	0.20
Weaving	reference	—
Factory ID[c]		
FID-344	0.21	0.21
FID-345	0.80	0.34
FID-537	reference	

Model adj r^2 = 0.32
[a]Full Model: SY—Sample Year, SP—Specific Process, MP—Major Process, FID—Factory ID.
[b]The coefficients for 7 of 39 specific processes in the model.
[c]The coefficients for two of 56 factories in the model.

Estimation of exposure for study subjects

Of the 4581 jobs in cotton factories, 1803 were in production operations where the primary predictive model (SY, MP, SP and FID) could define estimates. A series of secondary models was, therefore, created to allow complete matching with the subjects' work histories. These models and the number of jobs that each was able to provide estimates for are given in Table 4. In total, 88% of the jobs were assigned exposure levels from these models. Of the remaining 565 jobs, 520 were administrative jobs in non-production processes and assigned zero exposure. The remaining 45 jobs were assigned exposure levels

Table 3. Predicted VE-equivalent dust concentrations versus measurements; by Specific Process

Specific Process	n	Measurements		Estimates		Evaluation		
		Mean ($mg\ m^{-3}$)	SD ($mg\ m^{-3}$)	Mean ($mg\ m^{-3}$)	SD ($mg\ m^{-3}$)	R. Bias[a] (%)	R. Prec.[b] (%)	R. Acc.[c] (%)
Overall	43	0.71	0.50	0.70	0.22	−1.8	61.3	61.3
Opening	8	1.12	0.55	0.96	0.14	−14.1.1	45.5	47.6
Carding	8	1.06	0.54	0.74	0.11	−30.0	44.5	53.9
Drawing	8	0.83	0.58	0.71	0.11	−14.0	59.4	61.1
Combing	7	0.39	0.15	0.56	0.07	41.5	42.3	59.3
Roving	8	0.34	0.11	0.75	0.11	117.6	28.5	121.0
Spinning	4	0.27	0.20	0.23	0.02	−16.9	66.7	68.8

[a]R. Bias: Relative bias, $(bias/mean_m)*100$.
[b]R. Prec.: Relative precision, $(precision/mean_m)*100$.
[c]R. Acc.: Relative accuracy, $(accuracy/mean_m)*100$.

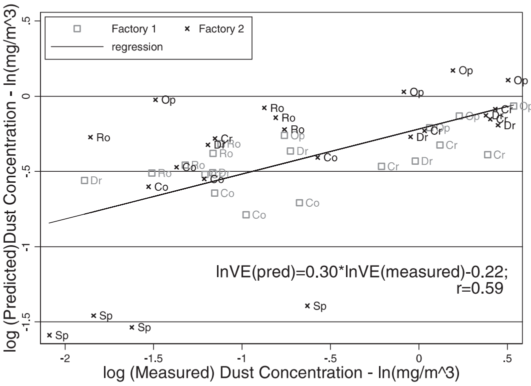

Fig. 2. Predicted Exposure Estimates versus Cotton Dust Measurements in two Shanghai textile factories (Op, Opening; Cr, Carding; Dr, Drawing; Co, Combing; Ro, Roving; Sp, Spinning).

according to the mean levels in their major process, as defined by the job dictionary.

Dust exposures in jobs held prior to 1975, the first year that measurement data were available, were estimated based on modeled exposures in 1975. Although there was an inverse relation between the sampling year and the historical dust concentration, it was not reasonable to extrapolate this association back in time to the beginning of the first job in 1932. If a linear extrapolation were assumed, the mean dust concentration of 5 $mg\ m^{-3}$ in 1975 would increase to 46.0 $mg\ m^{-3}$ in 1932. If the term of the job overlapped the boundary of the range of the model then the assessment was performed in two parts accordingly. Of the 4581 cotton jobs held by the study subjects, 1508 ended before 1975 and 448 ended before 1955.

Estimation of endotoxin

The specific endotoxin levels per milligram of cotton dust derived from literature are summarized in Table 5, and depicted in Fig. 3. Endotoxin concentrations in air for 1625 jobs within these seven processes were calculated based on the product of the

Table 4. The predictive models used to estimate cotton dust exposure

Models[a]	r^2	Jobs[b,c,d]	Relative accuracy (%)
Full			
SY, SP, MP, FID	0.32	1803 (39.4%)	61.3
Secondary			
SY, SP, FID	0.32	36 (0.8%)	61.3
SY, SP, MP	0.19	1678 (36.6%)	61.4
SY, SP	0.19	235 (5.1)	61.4
SY, MP	0.13	258 (5.6)	61.9
SY, FID	0.26	6 (0.1)	70.8

[a]SY—Sample Year, SP—Specific Process, MP—Major Process, FID—Factory ID
[b]4581 Cotton and 2662 non-cotton jobs defined based on the type of textile operation and the fibers used; non-cotton jobs assigned zero exposure.
[c]520 jobs (11.4%) in administration and in non-production processes of cotton factories were assigned no exposure.
[d]45 jobs (1.0%) assigned exposures based on re-coded major process.

specific endotoxin levels and the estimated cotton dust concentration. For the remaining jobs, endotoxin concentrations were approximated by the average levels associated with their major processes. All specific processes involving the handling of finished cloth or thread were assigned specific endotoxin levels equal to those measured in weaving, 160 EU mg^{-1} of cotton dust. For wet processes such as Bleaching and Dyeing or for processes working with completely finished products such as in Printing, endotoxin exposures were estimated based on one-tenth the specific endotoxin levels present in weaving or 16 EU mg^{-1} cotton dust.

Cumulative exposure to cotton dust and endotoxin

We estimated cumulative exposures for 3812 textile workers included in the epidemiological study of lung cancer. Study subjects worked between 1932 and the end of follow-up, December 31, 1998. Cumulative cotton dust exposure was computed as

Table 5. Endotoxin levels by specific process as measured in five surveys

Sample Year	1983[a]		1987[b]		1993[c]		1999[d]		2006[e]		Average	
Specific Process	*n*	GM (EU mg^{-1})	*n*	GM (EU mg^{-1})	*n*	GM (EU mg^{-1})	*n*	GM (EU mg^{-1})	*n*	GM (EU mg^{-1})	*n*	GM (EU mg^{-1})
Opening	14	1280	30	2970	86	5230	16	8360	8	665	154	2560
Carding	13	1800	36	3190	99	9700	16	19400	8	2730	172	4940
Drawing	4	1350	32	5400	91	7400	12	8630	8	4100	147	4530
Combing	7	1970	32	5360	58	5710	8	2880	8	2540	113	3380
Roving	4	1760	31	1650	56	2730	12	1870	8	2800	111	2110
Spinning	8	51.1	16	263	20	591	8	375	8	332	60	251
Weaving									8	159	8	159
Total	50	913	177	2710	410	5460	72	4970	56	1160	765	2390

[a]Olenchock *et al*. (1983).
[b]Kennedy *et al*. (1987).
[c]Christiani *et al*. (1993).
[d]Christiani *et al*. (1999).
[e]Astrakianakis *et al*. (2006).

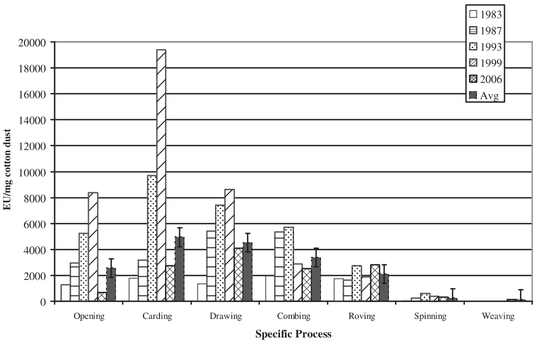

Fig. 3. Endotoxin per unit of cotton dust for seven specific processes collected during five sampling surveys (Olenchock *et al*., 1983; Kennedy *et al*., 1987; Christiani *et al*., 1993, 1999; Astrakianakis *et al*., 2006).

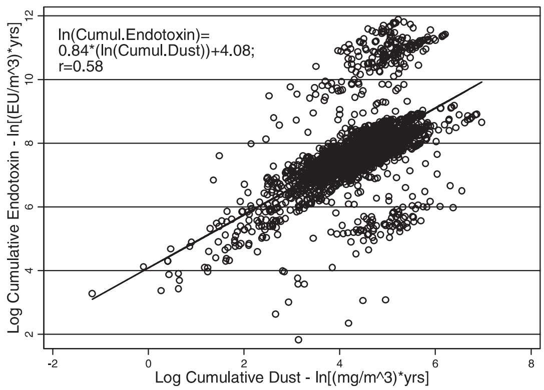

Fig. 4. Estimated cumulative exposure—cotton dust versus endotoxin, by subject. Comparison of exposures for 2539 subjects exposed to cotton dust and endotoxin.

the sum of the products of the estimated concentration and the duration for each of the jobs held by the 3812 workers in the study. Of these, 1221 subjects who had only worked in non-cotton jobs were assigned no exposure to cotton dust and endotoxin. Based on the aggregation of exposures over all jobs held by the 2592 cotton-exposed subjects, the mean cumulative exposure level was 113.8 mg m^{-3}*years. We could not estimate endotoxin exposures for 53 women who had worked as machinists, who were exposed to wool in early processing stages, or who held sanitation jobs with the potential for endotoxin exposure because no endotoxin data were available in these processes. Therefore, these subjects were removed from subsequent analyses. Among the remaining 2539 exposed women, the mean and median cumulative exposure to endotoxin was 6708 EU m^{-3}*years.

The relation between the estimated levels of cumulative cotton dust exposure and the approximated cumulative endotoxin exposure is presented in Fig. 4. The graph includes results for the 2539 subjects exposed to both cotton dust and endotoxin. The correlation among exposures for these subjects is 0.59.

The distribution of exposure levels evident in Fig. 4 reflects the approach used to assign endotoxin exposures. Among most subjects, the cumulative exposure to cotton dust and endotoxin follow a proportional association. A small group of subjects accumulated a much greater amount of endotoxin exposure for their level of cumulative dust exposure as compared with the majority of study subjects. Most of their work experience was in the early stages of fiber manufacturing where high-endotoxin content fiber was processed (e.g. carding or drawing). Conversely, a small group of subjects accumulated a lower amount of endotoxin exposure because most of their work experience was in parts of the factory where low or no endotoxin content fiber was processed (e.g. administration). Of course, the longer a subject worked the greater the dust and endotoxin exposure; however, the cumulative exposures are not associated with the number of jobs held.

DISCUSSION

Cumulative exposure to cotton dust was estimated for 7242 jobs held by 3812 study subjects based on historical measurement data collected from Shanghai textile factories. These measurements include data from 19% of the cotton factories in this study. The distribution of measurements is weighted towards older factories where exposures are often thought to be higher than newer operations owing to antiquated equipment and less rigorous control of exposure. However, in this dataset, higher measurements were associated (very weakly) with newer operations, perhaps owing to the larger scale of production in newer operations. Although requested, no information was available regarding production levels over time for these factories.

A source of uncertainty in exposure estimation is the accuracy of the historical concentrations measured by the CDS. The CDS is a device that collects each sample over a short duration of time (20–25 min) and at a very high flow rate, typically 20 l.p.m. The open-faced configuration of the CDS total dust sampler results in the capture of large particles that are not inhaled and are, therefore, not biologically relevant. Previous results show that CDS measures dust concentrations between two and ten times higher, depending on the process, than the VE (Astrakianakis *et al*., 2006). The imprecision in the measurements results in highly uncertain exposure estimates, and because of the over sampling, may not closely relate to the true dose.

The high mean concentration level of 2 mg m^{-3} found in the historical data is not surprising given the characteristics of the sampling device used to collect the samples. Even after converting this concentration to the inhalable dust fraction using equation (1) (0.45 mg m^{-3} as a VE-equivalent concentration), the GM of the dust levels in the historical database still exceed the current OSHA permissible exposure limit of 0.2 mg m^{-3} in yarn manufacturing [29CFR Cotton Dust, 1910.1043 (1978 (r1986))]. By comparison, Merchant *et al*. (1973) measured GM levels of 0.41 mg m^{-3} using a VE in a North Carolina textile factory.

Another source of uncertainty is the use of area samples rather than personal samples to model exposure. We previously demonstrated that personal exposure of inhalable cotton dust may be as much as 4.5 times the equivalent area measurement (Astrakianakis *et al*., 2006). In cotton industries outside China, Simpson *et al*. (1999) reported a median concentration of inhalable cotton dust based on personal samples of 1.04 mg m^{-3} in Lancashire cotton factories. Other UK studies comparing area and personal sampling measurements have found as much as one order of magnitude difference among sampler results (Ogden *et al*., 1993).

Cotton dust exposure model

The exposure model includes predictor variables for factory (FID), major process (MP), specific process (SP) and sample collection year (SY). The proportion of the variance in the sampling data resolved by the model is 32%. Of these four variables, FID is the most important, explaining 25% of the variance alone. Other predictors of exposure such as size of the workforce, age of the factory and season were not included in the final model. Although season was a statistically significant predictor of concentration it was not included for practical reasons. Estimated cotton dust exposures for the case–cohort study are cumulative concentrations based on annual rather than seasonal exposure levels. Not surprisingly the historical measurements collected in winter, when natural ventilation is at a minimum, were significantly higher than measurements collected in spring/fall, which in turn were higher than measurements collected in summer.

Our approach to the exposure reconstruction was similar to that of Stewart *et al*. (1995) and Seixas *et al*. (1991). A series of 'iterative' steps were used to assign exposures, directly or weighted according to what information was available, to each of the 7242 jobs. If the job was held during a year within the range of the model (1975–1999) and dust measurements for the factory, major process and specific process were available in the historical database, then an estimate using the full model could be assigned to that job. When less information was available, one of five secondary models was applied.

Exposure model validation

The accuracy of the model was evaluated using the dataset of measurements collected by Christiani *et al*. During four surveys over a 15 year period they collected cotton dust samples from two cotton factories also included in the current study. They collected respirable dust using a VE, requiring conversion of our CDS-based predicted estimates for comparison. The full model and all the secondary models performed with a similar degree of accuracy (61–72%). Overall, the error in the predictions was due mainly to a lack of precision in the estimates. By comparison, the magnitude of the bias of the predictions from all models was very small, <5% (data not shown).

Comparing the accuracy of the predictions by manufacturing process reveals the model overestimated dust concentrations in roving. Almost no historical data ($n = 2$) were available for the roving processes in these two factories. The average historical exposure measurement in this specific process across all factories is much greater than the independent measurements of exposure for either of these two factories, providing one possible explanation.

Although the validation results are satisfactory, it is important to emphasize that the measurement data were only available for 11 processes in two factories. Moreover, these processes were those for which the most historical data were available during the predictive model development and, therefore, this is a test of the estimates at their most stable.

An additional qualification regarding this evaluation stems from the error introduced with the conversion of the estimates. As described above the estimates were converted from CDS (total dust) to VE (respirable dust) in order to allow comparison with the data collected by Christiani and colleagues. The proportion of the variance resolved by the conversion model was 0.67. Therefore, the evaluation of the accuracy of the estimates includes error from both the predictive model and the conversion model.

Undoubtedly, these sources of uncertainty create some error in the estimates of cumulative exposures to cotton dust and endotoxin. Nonetheless, to our knowledge, these data represent the most complete quantitative estimates of exposure for the cotton industry generated for any epidemiological study.

SUMMARY AND CONCLUSION

Quantitative estimates of cumulative cotton dust and endotoxin exposure were calculated for 3812 women enumerated in a case–cohort study of lung cancer risk. The estimates were derived from a linear model based on a large set of historical measurements specific to these factories. These women held 7242 jobs in 503 factories from 1932 to 1998. The quantitative estimates of cotton dust exposure were compared with independently collected measurements from two of the study factories for the purpose of validation. The relative bias of the estimates was very low (<2%) and the relative accuracy was within 61% of the measurements and considered satisfactory based on suggested standards. The validation results suggest that inaccuracies in the estimates are due primarily to the imprecision of the underlying CDS data and the cumulative error of the predictive and conversion models. Despite the absence of historical endotoxin measurements, it was possible to assign retrospective exposures based on the predicted cotton dust estimates and approximations of endotoxin per unit of dust in specific processes.

Acknowledgements—This study was supported by grant R01CA80180 from the US National Cancer Institute, National Institutes of Health and R01OH2421 from the US CDC (NIOSH), Department of Health and Human Services. We thank our Shanghai colleagues Dr Dao-Li Gao, Ms Wang Wen Wan, the STIB industry hygienists and the BSE workers for their assistance throughout this study. We would also like to thank Dawn Fitzgibbons, Roberta Ray, Karen Wernli and Wenjin Li of the FHCRC for their tireless effort in supporting the Shanghai occupational study.

REFERENCES

Astrakianakis G, Seixas N, Camp J *et al*. (2006) Cotton Dust and Endotoxin Exposure Levels in Three Shanghai Textile Factories; a Comparison of Samplers. J Occup Environ Hyg; (Accepted).

Berry G, Molyneux M. (1981) A mortality study of workers in Lancashire cotton mills. Chest; 79: 11s–15s.

Blair A, Stewart P. (1992) Do quantitative exposure assessments improve risk estimates in occupational studies of cancer? Am J Ind Med; 21: 53–63.

Blair A, Rothman N, Zahm S. (1999) Occupational cancer epidemiology in the coming decades. Scand J Work Environ Health; 25: 491–7.

Bouyer J, Hemon D. (1993) Studying the performance of a job-exposure matrix. Int J Epidemiol; 22: S65–71.

Christiani D, Wegman D, Eisen E *et al*. (1993) Cotton dust and gram-negative bacterial endotoxin correlations in two cotton textile mills. Am J Ind Med; 23: 333–42.

Christiani D, Ye T-T, Zhang S *et al*. (1999) Cotton dust and endotoxin exposure and long-term decline in lung function: results of a longitudinal study. Am J Ind Med; 35: 321–33.

Hornung R. (1991) Statistical evaluation of exposure assessment strategies. App Occup Environ Hyg; 6: 516–20.

Intercooled Stata for Windows (2004) V8.2. College Station, TX: StataCorp LP.

Kennedy S, Christiani D, Eisen E *et al*. (1987) Cotton dust and endotoxin exposure-response relationships in cotton textile workers. Am Rev Respir Dis; 135: 194–200.

Kromhout H, Symanski E, Rappaport S. (1991) A comprehensive evaluation of within- and between-worker components of exposure to chemical agents. Ann Occup Hyg; 37: 253–70.

Kuzmickiene I, Didziapetris R, Stukonis M. (2004) Cancer incidence in the workers cohort of textile manufacturing factory in Alytus, Lithuania. J Occup Environ Med; 46: 147–53.

Levin L, Gao Y-T, Blot W *et al*. (1987) Decreased risk of lung cancer in the cotton textile industry of Shanghai. Cancer Res; 47: 5777–81.

Levin L, Zheng W, Blot W *et al*. (1988) Occupation and lung cancer in Shanghai: a case control study. Br J Ind Med; 45: 450–8.

Merchant J, Ortmeyer C. (1981) Mortality of employees of two cotton mills in North Carolina. Chest; **79** 6S–11S.

Merchant J, Lumsden J, Kilburn K *et al*. (1973) Dose response studies in cotton textile workers. J Occup Med; 15: 222–30.

Ogden TL, Bartlett I, Purnell C *et al*. (1993) Dust from cotton manufacture: changing from static to personal sampling. Ann Occup Hyg; 37: 271–85.

Olenchock S, Christiani D, Mull J *et al*. (1983) Endotoxins in baled cottons and airborne dusts in textile mills in the People's Republic of China. Appl Environ Microbiol; 46: 817–20.

Rappaport S. (1991a) Assessment of long-term exposures to toxic substances. Ann Occup Hyg; 35: 61–121.

Rappaport S. (1991b) Selection of the measures of exposure for epidemiology studies. Appl Occup Environ Hyg; 6: 448–56.

Rappaport S, Smith T. (1991) Exposure assessment strategies, exposure assessment for epidemiology and hazard control, industrial hygiene science series. Chelsea, MI: Lewis Publishers.

Sartorius. (July 1997). Sartorius basic plus, operating instruction [PDF]. Sartorius AG. Retrieved, Available at: www.sartorius.com

Seixas N, Sheppard L. (1996) Maximizing accuracy and precision using individual and grouped exposure assessments. Scand J Work Environ Health; 22: 94–101.

Seixas N, Moulton L, Robins T *et al*. (1991) Estimation of cumulative exposures for the national study of coal workers' pneumoconiosis. Appl Occup Environ Hyg; 6: 1032–41.

Simpson J, Niven R, Pickering C *et al*. (1999) Comparative personal exposures to organic dusts and endotoxin. Ann Occup Hyg; 43: 107–15.

Stewart P. (1999) Challenges in retrospective exposure assessment. Scand J Work Environ Health; 25: 505–10.

Stewart P, Triolo H, Zey J *et al*. (1995) Exposure assessment for a study of workers exposed to acrylonitrile. II. A computerized exposure assessment program. Appl Occup Environ Hyg; 10: 698–706.

Stewart P, Lees P, Correa A *et al*. (2003) Evaluation of three retrospective exposure assessment methods. Ann Occup Hyg; 47: 399–411.

Thomas D, Gao D, Ray R *et al*. (2002) Randomized trial of breast self-examination in Shanghai: final results. J Natl Cancer Inst; 94: 1445–57.

Wang Q-S, Boffetta P, Parkin D *et al*. (1995) Occupational risk factors for lung cancer in Tianjin, China. Am J Ind Med; 28: 353–62.

Wernli K, Astrakianakis G, Seixas N *et al*. (2005) Development of a job exposure matrix (JEM) for the Textile Industry in Shanghai, China. J Occup Environ Hyg (Accepted).

Wu-Williams A, Xu Z, Blot W *et al*. (1993) Occupation and lung cancer risk among women in Northern China. Am J Ind Med; 24: 67–79.