How valid are estimates of occupational illness?

While the annual BLS survey measures few chronic, long-latent, or fatal illnesses, estimates derived from other studies can prove statistically flawed and inaccurate

HARVEY J. HILASKI AND CHAO LING WANG

Incidence rates of occupational disease, published each year by the Bureau of Labor Statistics, understate the total impact of the work environment on workers' health.[1] This is so because the statistics virtually exclude chronic types of illnesses, as well as illnesses having a long latent period whose relationship to the job often surfaces only after retirement or death.

Alternative methods of measurement confirm that an undercount exists, but differ concerning its magnitude. This article examines some of the alternative methods of estimating occupational diseases and suggests that a consensus on the adequacy and reliability of the estimates is not likely.

One of the first studies to highlight the scope of occupational disease in this country was a pilot study sponsored by the National Institute for Occupational Safety and Health (NIOSH). Confined to cross-sectional samples of workers in designated small industries in Oregon and Washington, the study was designed to determine the usefulness of a set of medical procedures for diagnosing occupational disease, and to ascertain how much new data on occupational illnesses would be generated by this method. The results of the study, published in 1975, underscored the issue of a large undercount in current occupational illness statistics, primarily those of BLS.[2]

Harvey J. Hilaski is an economist in the Office of Occupational Health and Safety Statistics, Bureau of Labor Statistics. Chao Ling Wang is a biostatistician in the same office. Mary Kay Rieg of the *Review* staff provided special editorial assistance.

When analysts compared counts of cases of "probable occupational disease" from the pilot study with those from employer logs maintained under regulations of the Occupational Safety and Health Administration (OSHA), and with employer workers' compensation records, they found that nearly 90 percent of the cases (approximately 400) uncovered in the pilot study were not listed in employers' files.[3] In effect, the findings indicated that the procedures normally followed by employers in recording and reporting occupational illnesses in fulfillment of legal requirements result in a gross underestimate of occupational disease in the United States.

It should not be surprising that most of the "probable occupational disease" cases found in the pilot study were not included in employer records. Consider, for example, that of the 346 cases discovered in the study, hearing loss was most prominent, amounting to about 28 percent. This condition usually has a gradual onset, with the result that the worker may be unaware of any defect in hearing unless he or she undergoes a hearing test such as that administered in the study. Moreover, hearing loss is often part of the aging process; without baseline data and subsequent periodic testing of the work environment and resulting effects on the worker, the occupational relationship can be seriously challenged.[4]

The NIOSH study results may be questioned in several aspects, including possible bias inherent in the special procedures used to assess the health status of surveyed workers. However, the design is a feasible method for dealing with the detection of occupational diseases which are not readily apparent. Unlike the BLS annual survey, which measures the incidence of occupational disease, or the number of new cases recognized by employers as occupational in keeping with regulatory criteria, the NIOSH study measured the prevalence of occupational illness, or the level of occupational illnesses existing at a given time. For some purposes, prevalence is the more meaningful measure, in that it reflects the universe of persons with disease whose medical and economic needs may warrant special attention. However, high prevalence does not necessarily indicate high risk; instead, it may reflect an increase in survival, perhaps due to improved medical care. Conversely, low prevalence may reflect either a rapidly fatal process or cure of disease. The incidence measure is better adapted to keep a "running tab" on the health effects of workplace hazards, except for that component of occupational disease which is chronic in nature or of long latency.

Indirect estimates

Other studies and reports on occupational health problems have also pointed to an undercount in current illness estimates, but they are largely based on indirect evidence. The lack of reliable measures of occupational illnesses nationally has necessitated use and manipulation of surrogate data from epidemiologic or other studies to produce rather specific estimates of occupational diseases. On the basis of such data, a 1979 Labor Department report to Congress claimed that a conservative estimate of the prevalence of byssinosis among nearly 560,000 workers exposed to cotton dust at current levels was 83,610.[5] In 1980, another congressional report by the Department (hereinafter referred to as the Interim Report) provided several estimates of respiratory disease prevalence or deaths from worker exposure to asbestos, silica, beryllium, cadmium, chromium, arsenic, nickel, coal tar products, and diisocyanates.[6] Among other things, the Interim Report stated that, of 1 million workers currently exposed to silica, an estimated 59,000 will develop "some level" of silicosis; it also predicted 43,230 lung impairments as a result of exposure to diisocyanates, chemicals used to produce plastic products.

Much like the NIOSH study results, these estimates suggest a substantial undercount in the regularly published national statistics, which for 1980 showed 2,200 cases of all "dust diseases of the lungs."[7] (The BLS survey does not ask employers to specify recognized occupational illnesses such as asbestosis, byssinosis, and silicosis and thus provides no direct estimates of the incidence of these diseases.) It seems worthwhile, therefore, to examine more closely the major methodologies and techniques commonly used in deriving indirect estimates of occupational disease.[8]

Prevalence of disease. Prevalence rates were used to derive the estimate for silicosis cited above. These rates express the proportion of the population having a disease at a given time, regardless of time of onset. But because epidemiological studies, the major source of prevalence estimates of occupational disease, are confined to a particular population with a specific exposure, extrapolation of the findings to other, larger populations requires caution. Table 1, excerpted from the Interim Report, will be used to illustrate some of the limitations in the use of prevalence rates for this purpose.

The exposed population figures, obtained from a variety of sources, including the Mine Safety and Health Administration and NIOSH, are not representative exclusively of the exposed work force. The more important figures in this table are the prevalence rates, which were obtained from various special studies. Estimates of the numbers of disease cases were calculated by multiplying the exposed populations by the prevalence rates.

The crucial problem in applying prevalence rates derived from special studies to an entire production work force is the lack of assurance that the composition and levels of exposure of the larger population are the same as those of the worker group studied. The group of workers selected for epidemiological study is usually not a statistical sample of all production workers in the industry; therefore, the prevalence rates from such a study cannot be generalized to the larger population.

Use of the specialized prevalence rates in deriving the silicosis estimates probably led to biased results. For example, the 10-percent prevalence rate for the granite industry, which was adapted from an article published in the *Archives of Environmental Health* in 1974,[9] was not a prevalence measure of silicosis or other respiratory disease, as would be expected. The source data related only to current dust exposures in the Vermont granite

Table 1. Estimated number of cases of silicosis among currently exposed workers

Industry	Population exposed	Prevalence per 100 persons	Estimated number of disease cases
Total	1,057,000		59,102
Mining:			
Metal	24,000	3.4	816
Coal	126,000	.5	630
Nonmetal	7,000	.8	56
Quarry:			
Granite			
Sand or gravel	4,000	10.0	400
	40,000	10.0	4,000
Stone, clay, and glass products	511,000	2.0 – 20.0	31,500
Foundries:			
Iron and steel	192,000	4.0 – 9.0	16,700
Nonferrous	75,000		
Abrasive blasting	78,000	6.5	5,000

SOURCE: U.S. Department of Labor, *An Interim Report to Congress on Occupational Diseases*, June 1980, table A-2, pp. 130–31.

sheds. But for estimating purposes, it was assumed that sand and gravel workers have the same "prevalence rate" as the granite workers, under further assumption of a similar degree and duration of exposure. In the case of the stone, clay, and glass products industry, a specific prevalence rate was selected for estimating purposes, although special studies showed a range of rates. While the chosen rate—6.16 percent—appears conservative, there is no reason to assume that it is a better estimate than any other within the wide 2- to 20-percent range of individual study results. If prevalence rates had been calculated from sample data, the precision of the estimates could be measured by standard errors. But for the data in table 1, the standard error of estimated prevalence is not available, and the precision of the estimates is therefore not known.

Dose-response. Dose-response studies generally attempt to establish a statistical relationship between dose (exposure) and response (onset of disease or death). For simplicity, this will be illustrated by example. Eight of 22 North Carolina textile manufacturing plants were selected for a 1970–71 study of cotton textile workers.[10] A dose-response curve was fitted to the resulting data on the prevalence of byssinosis among a group of 1,259 workers in the cotton preparation and yarn area. (See table 2.) From the fitted curve, the byssinosis prevalence rates and their 95-percent confidence limits for workers in the cotton preparation and yarn area were predicted at various cotton dust levels of exposure:

Dust level (mg/m^3)	0.1	0.2	0.5
Predicted prevalence per 100 workers	6.5	12.7	25.8
95-percent confidence interval	5.0–8.5	10.8–14.9	22.5–29.3

The referenced study suggested that a reasonably safe level of lint-free cotton dust is 0.1 mg/m^3 in the cotton preparation and yarn area, because nearly 94 percent of the workers exposed at this level had no symptoms of byssinosis.

As indicated in table 3, the authors of the Interim Report used the results of the North Carolina study to calculate the total number of expected byssinosis cases by multiplying 1977 BLS data for production workers in six yarn manufacturing industries by the prevalence rate of 25.8 per 100 workers—the prevalence predicted by the textile worker study for dust level exposure of 0.5 mg/m^3. But it is unrealistic to assume that all workers in yarn industries generally are exposed to such high levels of cotton dust. This assumption might not even be true of cotton preparation and yarn workers nationwide, the types of workers among whom the prevalence study was done. Even if it were, however, the results of this special study should not be construed as necessarily representative of production workers in all yarn manufacturing industries, because the occupational composition of other industry segments would probably not be similar to that of the cotton preparation and yarn area.

Table 2. Data on byssinosis prevalence by median dust level

Item	Median dust level (mg/m^3 rounded)									
	.05	.07	.10	.15	.23	.34	.51	.77	1.2	1.7
Sample size (n)	145	71	193	279	78	208	147	65	31	42
Cases of byssinosis (r)	5	2	14	27	10	39	37	30	14	17
Prevalence ($r/n \times 100$)	3.4	2.8	7.3	9.7	12.8	18.8	25.2	46.2	45.2	40.5

Closer scrutiny of the dose-response relationship in the cotton preparation and yarn area makes the selection of the 25.8 prevalence rate even more puzzling, as only about 285, or 23 percent, of the 1,259 workers fell in the categories exposed to median dust levels 0.5 mg/m^3 or above. According to the data in table 2, the average overall byssinosis prevalence for the entire study was 15.5 per 100 workers, or 195 persons.

For the study of illnesses such as byssinosis, information on duration or years of exposure to the hazard is also crucial. This factor should be taken into account because a worker with 10 or 20 years of work exposure would seem to be more susceptible to byssinosis than a worker with a few years of exposure. Moreover, a worker's exposure level may change over the years, due to changes (not always for the worse) in working conditions, including ventilation, industrial hygiene practices, and so forth. In short, a worker normally experiences different amounts of exposure, of varied duration, over the course of his or her employment. If the dose-response relationship curve is not adjusted for the extent of exposure, its accuracy is diminished. Unfortunately, comprehensive data on the intensity, duration, and fluctation of exposure are rarely available, particularly in retrospective studies of the type used in making the above estimates.

In general, dose-response relationship curves are nonlinear, monotonic (increasing or decreasing), and have lower and upper asymptotes (usually, but not always, 0 and 100 percent). The models tend to operate well in a restricted range of exposure or dosage levels, but not over the entire range; that is, they may be useful in determining the "safe" level of exposure, but they are not suitable for developing national or industry-wide statistics on occupational illnesses.

Standardized mortality ratio. The standardized mortality ratio has been widely used as a summary index of mortality in occupational epidemiologic studies. The ratio is a method commonly used to accomplish indirect age adjustment by applying age-specific death rates of a standard population to a study population to yield a

Table 3. Expected numbers of byssinosis cases, selected yarn manufacturing industries

Yarn manufacturing	Standard Industrial Classification Code	Exposed workers	Assumed prevalence (in percent)	Expected cases
Broad woven fabric mills, cotton	2211	86,600	25.8	22,340
Circular knit fabric mills	2257	1,000	25.8	300
Yarn spinning mills	2281	34,400	25.8	8,880
Texturizing, throwing, twisting, and winding mills	2282	13,900	25.8	3,590
Thread mills	2284	3,600	25.8	930
Tire cord and fabric	2296	1,600	25.8	410

SOURCE: U.S. Department of Labor, *An Interim Report to Congress on Occupational Diseases*, June 1980, table A-1, p. 126.

number of "expected" deaths. It is defined as:

$$\text{SMR} = \frac{\text{total observed deaths in a study population}}{\text{total expected deaths in the population}}$$

A ratio greater than 1 means that more deaths have been observed in the study group than would be expected on the basis of rates for the standard population; conversely, fewer deaths than expected are indicated by a ratio less than 1. A test, such as chi-square, is generally used for determining the level of significance of the results.

If the focus is on mortality from hazards in the workplace, an ideal standard population would include all workers in the Nation. But because mortality data are not available in this detail, the total U.S. population, or the male population, is generally selected as the standard. Consequently, misunderstanding sometimes arises in applying standardized mortality ratios to estimate the total number of deaths caused by certain diseases in industry.

In the Interim Report, the ratio was used to derive estimates of work-related lung cancers.[11] Data from three sources were used in the computations:

(1) The number of workers exposed to beryllium compounds and oxides in end-user processes, estimated at 50,000, was obtained from the NIOSH 1972 National Occupational Hazard Survey.
(2) The mortality ratio of 1.6 for lung cancer among beryllium workers was taken from epidemiologic studies.
(3) The 1976 U.S. age-adjusted incidence for lung cancer of 116 per 100,000 males over age 20 was used.

The number of expected deaths from lung cancer among workers in end-user industry processes was then calculated as: 50,000 × 1.6 × (116/100,000), or about 93 lung cancer deaths. Is this a valid estimate of all lung cancer deaths in this industrial population which were due to exposure to beryllium compounds?

First, it is important to understand that a standardized mortality ratio of 1.6 does not mean that the mortality rate of the study population is 1.6 times that of the standard population and can thus be used as a multiplying factor to obtain the number of deaths in a broader population. Even when exposed to the same health hazards and having the same age-specific death rates for all age groups, different study populations will yield different values of the standardized mortality ratio if their age distributions differ. Obviously, then, the ratios should not be used in estimating the number of deaths due to disease in populations which differ in composition from a study population.

Second, a high ratio does not imply that all deaths from lung cancers, for example, are caused by occupational exposure; it tells us only that the study population has an unusually high mortality risk. We do not know what percentage of deaths actually resulted from exposure, in this case to beryllium. Even if the standardized mortality ratio were interpreted to mean that the number of deaths of the study population is 1.6 times that of the standard population, 58 (that is, 93/1.6) deaths per year from lung cancer would have occurred in the study population irrespective of any exposure to beryllium. That leaves 35 deaths per year (or about 38 percent of total estimated deaths) which might be attributed to, or aggravated by, exposure to beryllium compounds. At most, the estimate of 93 deaths depicts the total cancer toll among the occupational group, not the excess cancer resulting from beryllium exposure.

Relative risk. Because incidence or other direct measures of occupational disease are generally lacking, epidemiologic study of occupational morbidity often relies on a measure of excess risk of a disease among workers in specified working environments to determine the association between certain occupational factors and the incidence of disease. One such measure is relative risk.

As we will see, relative risk also is subject to misuse in making estimates of occupational disease in industry, perhaps because of confusion about its definition. This can be illustrated again by an example from the Interim Report. From the National Occupational Hazard Survey, 98,090 workers were estimated to be exposed to chromates in chromate pigment production and, as noted, the 1976 U.S. male age-adjusted incidence rate of lung cancer for those over age 20 was 116 per 100,000. Based on these two figures and a chosen relative risk of 5, the estimated number of lung cancers per year among the 98,090 workers was 570 cases: (exposed population) × (incidence rate of male population) × (relative risk) = (98,090) × (116/100,000) × 5 = 570 cases.

The report stated: "Based on three studies reporting relative risks from 2.3 to 38, a relative risk of 5 will be used for workers exposed to chromate compounds. . . ."[12]

Seemingly, this is a conservative choice, but in reality there is no way to tell whether the relative risk of 5 is appropriate or not, because we have no information on its precision. More basic issues are whether relative risk should even be used to estimate the incidence of lung cancer cases, *and* what is involved if one does so.

Relative risk is a measure of the strength of the association of the disease with a certain factor, such as exposure to a specific chemical, and thus is an important statistical tool in retrospective epidemiological studies. It is defined as the ratio of the incidence rate of those exposed to a factor to that of those not exposed. Conversely, relative risk can be used to compare groups of subjects diagnosed as having a disease to determine if the groups differ in the proportion of persons who had been exposed to the specific factor or factors. However, because retrospective study entails looking at the historical frequency of the suspected cause in a diseased group and a control group, the incidence rates of the diseased among the exposed and unexposed cannot be estimated directly but only approximated by relative risk, an odds ratio (or risk ratio).

Consider the following tabulation, in which the total T_i of the ith group of workers in a study population T $(T = \sum_{i=1}^{K} T_i)$, where K equals the total number of groups, is divided as:

	With disease	*Without disease*	*Total*
Exposed to factor	A_i	B_i	$A_i + B_i$
Not exposed to factor .	C_i	D_i	$C_i + D_i$
Total	$A_i + C_i$	$B_i + D_i$	T_i

Data classified in the table may be obtained from prospective, cross-sectional, or retrospective studies. According to the tabulation above, the proportion of workers exposed to a factor and having the disease is $A_i/(A_i + B_i)$, while the corresponding proportion of unexposed workers with the disease is $C_i/(C_i + D_i)$. Thus, the relative risk of disease for exposed workers is: $A_i(C_i + D_i)/C_i(A_i + B_i)$. But because the incidence of a specific disease in a population tends to be low, the calculation $(A_iD_i)/(B_iC_i)$—that is, $(A_i/B_i)/(C_i/D_i)$—provides a close approximation of the relative risk (but not of the incidence of disease) for the ith group of individuals. For an overall relative risk, a commonly used formula is:[13]

$$R = \sum_{i=1}^{K} (A_iD_i/T_i) \;/\; \sum_{i=1}^{K} (B_iC_i/T_i)$$

Because age is an important factor affecting incidence of disease, it should be accounted for in measurements of overall relative risk. In such an age-adjusted risk ratio, T_i is the total of the ith age group. The chi-square test is commonly used to determine whether the relative risk is significantly different from 1. Using a risk ratio from a particular study and an incidence rate for the general male population to estimate the extent of occupational disease in a larger population exposed to a specific factor assumes similarity among the age and sex distributions of all groups. However, in the Interim Report's estimate of disease from exposure to chromate compounds, the composition of the general male population and the exposed population may not be similar to that of any special study population. Therefore, use of the relative risk of 5, selected from a range, may produce biased results.

In brief, relative risk is a measure only of the strength of the association between the disease and the exposure factor. If significantly different from 1, it indicates only that the disease is strongly associated with the exposure factor, not that the factor necessarily causes the disease. Any firmer conclusion would require further study. Relative risk is surely a critical measure for assessing the etiologic role of a factor in disease, but it is not suitable for estimating the incidence of disease.

Cancer related to occupational factors

What fraction of cancer incidence in this country is attributed to occupational exposure to carcinogens in the workplace? An unpublished 1978 report prepared jointly by several research institutes indicated that about 20 percent of all cancers are occupationally related and stated: "If recent evidence is considered and if the full consequences of occupational exposures in the present and recent past are taken into account, estimates of at least 20 percent . . . may even be conservative."[14] The report concluded that earlier estimates that only 1 to 5 percent of all cancers in the United States were attributable to occupational factors had not been scientifically documented and that Dr. Philip Cole's 1977 estimates of less than 15 percent for men, and less than 5 percent for women, contained a large element of uncertainty.[15] The results from the joint report have been cited in numerous publications, and questions have been raised concerning their validity.

The 20-percent overall estimate resulted from a two-step merger of the results of several separate studies. The first step developed estimates of the fraction of cancers due to asbestos exposure, while the second compared the risks from asbestos exposure with those from five other high-exposure substances, with the final result based on that comparison. Details of the estimation procedure follow.

According to the report, about 8 to 10 million workers have been exposed to asbestos since the beginning of World War II, and approximately 4 million have had heavy exposure. On the basis of a longitudinal study of a cohort of insulation and shipyard workers, the report indicated that, of deaths of heavily exposed workers, 20 to 25 percent were from lung cancer, 7 to 10 percent from pleural or peritoneal mesothelioma, and 8 to 9 percent from gastrointestinal cancer. Summed up, this suggests that 35 to 44 percent of cohort deaths were attributable to cancer diseases. Accordingly, the joint study group concluded that, over the next 30 years, at least 1.6 million (about 40 percent of the 4 million) heavily exposed workers would die from the asbestos-related cancers listed above. Based on an assumption suggested by data from a second source, the excess risk to the remaining less heavily exposed workers (4 to 7 million) was estimated to be one-fourth of that for the heavily exposed (a 10-percent risk, obtained as ¼ × 40 percent), yielding a cancer estimate for this group of 0.4 to 0.7 million. This raises the total to between 2.0 and 2.3 million cancer deaths over the next three decades, with expected averages of 58,000 to 75,000 cancer deaths per year associated with asbestos alone. The joint study indicated that this excess number of cancer deaths would account for roughly 13 to 18 percent of all expected cancer deaths.

In the second step, the study presented data on carcinogenic risks to workers found to be exposed to five substances in addition to asbestos during a 1972–74 National Occupational Hazard Survey.[16] Table 4, adapted from the study, shows selected results. The risk ratios were either standard mortality ratios or risk ratios selected from a range of values obtained from other epidemiological studies. The report indicated that the figures were not precise estimates, but reasonable ones for comparison purposes, because they were all derived by the same method.

Other conclusions pertaining to the second step of the joint study:

- Excess cases for the other five substances combined are about 33,000 cancers per year, versus 13,900 for asbestos alone (table 4). The data show that these five agents together pose hazards similar to or greater than those posed by asbestos.
- The projected numbers of excess cancers are only for the 1972–74 groups of N-size workers. But because of workplace turnover, the actual number of workers exposed over time will be several times larger than N.
- Consequently, the excess number of cases from asbestos (13,900) among the 1972–74 group underestimates the annual expected number of cancer deaths related to asbestos—58,000 to 75,000, as derived in step 1—by a factor of 4 to 5.
- Because the data for the five other substances listed in table 4 were derived in the same way as those for asbestos, the estimates may likewise underestimate the number of cancers attributable to these substances.

Table 4. Risks of disease associated with exposure to selected substances

Chemical substance	Estimated number of workers exposed, 1972-74 (N) (in millions)	Diseases of risk	Risk ratio (R) (assumed)	Age-adjusted incidence of disease (I) (per 100,000)	Projected number of excess cases (R-1)NI
Asbestos	1.6	Lung cancer	6.6	116	10,400
		Mesothelioma	(one third	of lung cancer)	3,500
Arsenic	1.5	Respiratory cancer	4.7	131	7,300
Benzene	2.0	Leukemia	5	17.9	1,400
Chromium	1.5	Respiratory cancer	5	131	7,900
Nickel	1.4	Respiratory cancer	5	131	7,300
Petroleum products	3.9	Lung cancer	3	116	9,100

SOURCE: National Cancer Institute, National Institute of Environmental Health Services, National Institute for Occupational Safety and Health, *Estimates of the Fraction of Cancer in the United States Related to Occupational Factors*, September 1978, (unpublished), table 2, p. 33–38.

- According to results from the first study step, asbestos alone will account for between 13 and 18 percent of all cancer deaths over the next 30 years. The data for the five other substances suggest at least 10 to 20 percent additional cancer deaths. Hence the study conclusion that occupationally related cancers may make up 20 percent or more of cancer deaths in forthcoming decades.

Closer examination of the joint study findings indicates that they may not be fully supported by data from the various studies used in their development, a conclusion corroborated in a report prepared for the Office of Technology Assessment, U.S. Congress.[17]

First, a study of causes of nearly 2,300 deaths among a cohort of 17,800 asbestos insulation workers contributed the finding that 35 to 44 percent of workers heavily exposed to asbestos died of cancer. The joint study group selected 40 percent as an approximation of the cohort's fatal cancer risk, and applied this percentage to a national population of 4 million workers considered to be heavily exposed to asbestos. This extrapolation was based on the unstated and probably unjustified assumption that the cohort of asbestos insulation workers was representative of the worker population of mixed industries. The mixed industry population might well differ from the cohort group not only in levels and length of asbestos exposure, but also in terms of population factors, such as age, sex, race, and percentage of smokers, which play significant roles in risk assessment. From the statistical point of view, application of the results from a study population to other populations of different composition and exposure experience usually produces biased estimates. The assumption that the risk of cancers for the less heavily exposed group of 4 to 7 million workers is one-fourth that of the heavily exposed group is similarly questionable. Consequently, the resulting total estimate of 58,000 to 75,000 asbestos-related cancer deaths per year is highly suspect.

Second, the analytic results by "causes" of death in the cohort group of asbestos insulation workers showed only the percentage distribution by disease, irrespective of cause; they did not indicate what fraction of cancer was induced or aggravated by asbestos alone or by any other specific exposure. Although it may be highly abnormal to find 20 to 25 percent of lung cancers and 8 to 9 percent of gastrointestinal cancers as causes of death among a group of workers, the actual percentage of cohort-group deaths specifically associated with exposure to asbestos remains uncertain.

Third, the method of estimating excess cancer cases or deaths for each exposure substance shown in table 4 may result in either overestimates or underestimates. Risk ratios do not in themselves provide the magnitude of risk, because they are greatly affected by the composition of the study population. In that the risk ratio is a ratio of observed to expected disease cases or deaths, any small increase in observed cases will greatly increase the risk ratio if the expected number of cases is small. This is especially true when the study population is small. Assume, for example, that study results indicate that 1,400 deaths were observed when only 1,000 were expected, yielding a risk ratio of 1.4. The chi-square value for the level of significance will be: $X^2 =$ (observed-expected)^2/expected$=160$. To achieve the same level of significance for an expected number of 10, the observed number would have to be 50, and the resulting risk ratio would be 5. Thus, a small study population has a better chance to yield a large risk ratio than a large study population, if both experience the same hazards and have the same population composition. It is not surprising, therefore, to find that risk ratios vary widely among epidemiologic studies of workers exposed to the same chemical substance. It is more important to know whether the value of the risk ratio is significantly different from 1, which indicates that the risks for the exposed and nonexposed groups are not identical, than to determine the absolute magnitude of the ratio itself.

Fourth, the joint report considered 13,900 excess cases per year due to asbestos exposure to be underestimated by a factor of 4 to 5 under the assumption of high turnover of the work force. Because the report had already estimated 58,000 to 75,000 asbestos-related deaths per year in an exposed population estimated at 8 to 11 million, further inflation by this factor results in an incredible number of occupational cancers. Based on the age-adjusted incidence shown in table 4, the number of expected lung cancer deaths, excluding mesothelioma, among a 1.6 million male population over age 20 would be 11.6 million $\times$ (116/100,000) or 1,856, in the absence of any exposure to asbestos. It follows, then, that excess lung cancer deaths due to asbestos exposure alone (10,400) amount to 85 percent of all lung cancer deaths in this 1.6 million exposed population—that is, 10,400/(1,856 + 10,400). Add to this calculation the mesothelioma cases, and the figure becomes 88 percent, a rather astounding share considering all other cancer causes.

Estimates of the excess number of predominantly respiratory type cancer cases due to the five other substances for the given groups of exposed workers also are indicated in table 4. For these groups, the expected numbers of cancer cases in the absence of exposure would be: 1,970 (arsenic), 360 (benzene), 1,970 (chromium), 1,830 (nickel), and 4,520 (petroleum products), by direct application of the corresponding incidence rate. This means that the proportion of total cancer cases associated with each of these five chemicals would amount to 79 percent, 80 percent, 90 percent, 80 percent, and 67 percent, respectively. These proportions appear unreasonably high.

Finally, the joint study group considered the excess number of cancers attributed to the five other substances to be underestimated also, based on the two-step findings for asbestos and applying the same logic. However, this inference is not justifiable, because estimates for the other substances were obtained independently, and the magnitudes of the estimates were greatly affected by the value of the risk ratio chosen for each.

Disability-impairment data

Three types of disability data are available to the occupational health analyst: Social Security Administration data from the various surveys of disabled adults; the Social Security Administration's disability applicant files; and the National Center for Health Statistics' Health Interview Survey data. These data sets also suggest that there are greater numbers of health problems with some occupational connection than published data from employer-based surveys would indicate, but they do not necessarily establish a causal connection between work and disease. Therefore, it may be beneficial to discuss briefly the conceptual framework of these data bases.

The disability study. This study was an analysis of disability data obtained from an interview survey of 18,000 persons age 20 to 64, who had been selected from the 1970 5-percent census sample.[18] The survey was conducted by the Bureau of the Census for the Social Security Administration. Of the 18,000-member sample, 11,700 had been selected from among those persons who had a disability prior to October 1969 as indicated on the 1970 census questionnaire. A mail screening in 1971 resulted in selection of another 1,200 recent onset cases and 5,100 nondisabled persons. Disability in the study was defined "severe," if it precluded work altogether, or "partial," if it limited the kind or amount of work performed.

Based on survey results, it was estimated that about 15 million persons age 20 to 64 nationwide were disabled from all causes. To determine the job-relatedness of a disability, the survey respondent was asked: "Was your (main condition or illness) caused by your job?" Of the resulting estimated 2.4 million job-caused disabilities, 1.7 million were attributed by the author of the study to occupational disease. (For purposes of the study, occupational diseases were defined as all cases of disability which were not caused by an accident on the job.) The study further stated: "Because of limited understanding of what diseases are occupational in nature, it is likely that the actual number of occupational disease cases is much higher" (page IV).

The study indicated that, of the 1.7 million persons disabled by occupational disease, 1.1 million or two-thirds were partially disabled. Among the 0.6 million severely disabled, musculoskeletal and cardiovascular conditions accounted for almost 60 percent of all conditions reported. About 25 percent had mental and digestive ailments and 9 percent, respiratory conditions. Cancer caused by occupation was estimated among the severely disabled at little more than 1 percent, slightly over 6,000 cases.

Can we say that these are good estimates? As the author of the study points out, the survey provided a snapshot of the population at a given time.[19] Thus, the resulting estimates measured the prevalence of occupational disease, counting as they did all existing disabilities without regard to time of onset or diagnosis. The occupational relationship was subjectively perceived by respondents and may or may not have been corroborated by objective medical evidence. Except perhaps for the musculoskeletal impairments, the other conditions were of a type for which objective evidence of occupational causality might have been difficult to obtain.

Finally, while national estimates based on survey data of self-perceived work-related disabilities suggest that an undercount in employer-based occupational illness estimates does exist and in some identifiable parameter, it appears that those diseases which have long latent periods and have yet to be diagnosed are missed in this approach as well. Many long-latent diseases or aggravating disease symptoms of possible occupational origin are recognized beyond the cutoff age of most direct surveys. When disease appears after a worker has lived at least a normal life span, other factors, related to the aging process, enter which may lessen the urgency to determine precise causes.

Disability applicant files. The Social Security disability program provides benefits to disabled adults with work experience in employments covered by the Social Security Act, and to adults disabled since childhood who are dependents of disabled or retired work beneficiaries or of deceased insured workers. To qualify, claimants for Social Security Disability Insurance must prove that they are both disabled and unable to engage in any substantially gainful work due to their medical condition. Two sample data files from disability applicant records are maintained by the Social Security Administration (SSA), the Continuous Disability History Sample (CDHS) and the Longitudinal Sample of Disability Insurance Applicants (LSDIA). CDHS contains about 25 percent of allowed claims and 10 percent of denied claims, while LSDIA is a 5-percent longitudinal sample of disability applicants.

Disability applicant records contain demographic information, such as sex, race, date of birth, education, occupation, and industry of employment, as well as important medical information, such as diagnosis (primary, secondary, and tertiary), listings of impairments, principal body system involved, and severity and duration of impairment. It is important to note, however, that in the recording and coding of disability cases, work-related illnesses are not distinguished from nonwork-related ones. That is, for adjudicative purposes, an occupational relation to the disease or disability does not have to be established. (Although occupations are associated with worker claimants, no causal relationship is required or intended.) According to a 1974 report by the U.S. Department of Health, Education, and Welfare, the leading causes of disability by diagnosis were listed as chronic ischemic heart, schizophrenia, osteoarthritis, emphysema, displacement of intervertebral disc, diabetes mellitus, rheumatoid arthritis, acute cerebrovascular disease, malignant neoplasm of trachea and lung, neuroses, pulmonary tuberculosis, and mental disorders.[20] This listing also contains types of diseases for which, obviously, objective evidence of occupational causation would be hard to come by.

To qualify for disability insurance, a claimant must have a health condition sufficiently incapacitating to be unable to engage in any substantial, gainful work. Thus, a worker may have an occupational disease, but be disallowed disability benefits because pursuit of gainful employment is still possible. Therefore, disability estimates based on SSA records are not precise and comprehensive indicators of occupational impact. In addition, the occupational history of an applicant is limited to his or her longest full-time occupation in the 10-year period preceding the alleged date of onset. Because a job-caused disability, especially one of a chronic nature, may have been due to an earlier exposure, and perhaps, to a different job, there is a potential bias in the use of these statistics for epidemiological study. While the SSA disability files are an important source of data for development of morbidity ratios which identify disease and occupational relationships worthy of further study, they are not suitable for deriving estimates of occupational disease.[21]

Health Interview Survey data. The Health Interview Survey of the National Center for Health Statistics is a nationwide survey of approximately 40,000 households, conducted on a continuous basis. It is designed to gather information on personal and demographic characteristics, illnesses, injuries, impairments, chronic conditions, and other health topics. Respondents are asked whether they worked in the 2 weeks prior to the interview week, and in what occupation and industry. Each year's sample includes about 120,000 persons, of whom roughly 50 percent are employed.

As data are processed and tabulated, the center publishes analytical reports on various topics. While very few reports have been published on the work force population as the primary study target, data files are available for research purposes. Like the Social Security Administration disability applicant data the center's data can serve important epidemiological research objectives, but should not be used to derive precise estimates of occupational disease incidence.[22]

DESPITE THEIR SHORTCOMINGS, the results of the studies and applied methodologies discussed above do, in combination, point to a larger impact of the workplace on the health of workers than is borne out in regularly published statistics, although the magnitude of the understatement remains uncertain. Continued efforts towards improved or new methods are needed to produce national estimates of greater credibility for the chronic and long latent disease component of job origin. Such efforts might include improved techniques for diagnosing occupational diseases; more sophisticated and efficient means of monitoring workers' health; education and training of doctors and workers regarding health hazards on the job; conduct of epidemiological studies representative of national experience; and establishing methodology for determining the contribution of job exposure to the origin and course of disease. □

FOOTNOTES

[1] Harvey J. Hilaski, "Understanding statistics on occupational illnesses," *Monthly Labor Review*, March 1981, pp. 25–29.

[2] David P. Discher, Goldy D. Kleinman, and F. James Foster, *Pilot Study for Development of an Occupational Disease Surveillance Method* (Washington, U.S. Department of Health and Human Services, National Institute for Occupational Safety and Health, 1975).

[3] In the survey, the examining physician determined a condition to be one of five types: probable occupational disease; doubtful occupational disease; suggestive history; cannot be evaluated; probably nonoccupational. Probable occupational disease was considered to be present when "manifestations of disease are consistent with those known to result from excessive exposure to a given injurious agent; this injurious agent is present in the patient's working environment and significant contact in course of usual duties is likely." Discher and other, *Pilot Study*, p. 25.

[4] Although OSHA had a noise standard as early as 1971, environmental and audiometric testing was not formalized at the time of the pilot survey. Under the Hearing Conservation Amendment, which became effective Aug. 22, 1981, the permissible exposure level (PEL) remains at the 90 decibel (dB) level as an 8-hour time weighted average but an 85 decibel time weighted average was established as an action level which triggers the initiation of hearing conservation programs, including exposure monitoring, audiometric testing of employees, training, and some recordkeeping.

[5] U.S. Department of Labor, *Cotton Dust: Review of Alternative Technical Standards and Control Technologies*, Report to the Congress, May 1979.

[6] U.S. Department of Labor, *An Interim Report to Congress on Occupational Diseases*, June 1980. Also see "Labor Month in Review," *Monthly Labor Review*, August 1980, p. 2.

[7] See *Occupational Injuries and Illnesses in the United States by Industry, 1980*, Bulletin 2130 (Bureau of Labor Statistics, 1982), p. 32, table 7.

[8] This study was undertaken in the spring of 1980 as part of the Bureau's continuing evaluation of occupational safety and health statistics.

[9] Gilles P. Theriault, William A. Burgess, Lou J. DiBerardinis, and John M. Peters, "Dust Exposure in the Vermont Granite Sheds," *Archives of Environmental Health*, Vol. 28, 1974, pp. 12–17.

[10] James A. Merchant, John C. Lumsden, and others, "Dose Response Studies in Cotton Textile Workers," *Journal of Occupational Medicine*, Vol. 15, No. 3, 1973, pp. 222–30.

[11] *An Interim Report*, pp. 21–23.

[12] *Ibid.*, p. 26.

[13] See Nathan Mantel and William Haenszel, "Statistical Aspects of the Analysis of Data from Retrospective Studies of Disease," *Journal of the National Cancer Institute*, Vol. 22, No. 4, 1979, pp. 719–48, for a full discussion of the various relative risk formulas.

[14] National Cancer Institute, National Institute of Environmental Health Services, National Institute for Occupational Safety and Health, *Estimates of the Fraction of Cancer in the United States Related to Occupational Factors*, September 1978 (unpublished), p. 1.

[15] *Ibid.*, pp. 5–8.

[16] Exposure information for all substances, except arsenic, was derived from the National Occupational Hazard Survey conducted by NIOSH during 1972–74. Exposure information for arsenic was contained in criteria for a 1975 recommended standard developed by NIOSH.

[17] Richard Doll and Richard Peto, "The Causes of Cancer: Quantitative Estimates of Avoidable Risks of Cancer in the United States Today," *Journal of the National Cancer Institute*, June 1981, pp. 1238–45.

[18] Glen M. Shor, *Occupational Disease and Compensation: An Analysis of the 1972 SSA Survey of Disabled Adults*, prepared under contract for the Office of the Assistant Secretary for Policy, Evaluation and Research, U.S. Department of Labor, October 1979.

[19] Shor, *Occupational Disease*, p. 2.

[20] *Social Security Disability Applicant Statistics 1970* (Washington, U.S. Department of Health, Education, and Welfare, 1974).

[21] For a recent study of Social Security disability data, see *NIOSH Research Report, Occupational Characteristics of Disabled Workers* (Washington, U.S. Department of Health and Human Services, National Institute for Occupational Safety and Health, 1980).

[22] For analysis of Health Interview Survey data, see *NIOSH Research Report, Industrial Characteristics of Persons Reporting Morbidity During the Health Interview Surveys Conducted in 1969–1974* (Washington, U.S. Department of Health and Human Services, National Institute for Occupational Safety and Health, 1980).