American Journal of Industrial Medicine 12:419–430 (1987)

Assessing the Feasibility of Retrospective Cohort Studies

Kyle Steenland, PhD, Leslie Stayner, MS and Alice Greife, PhD

While most epidemiologic cohort studies are preceded by some sort of feasibility study, details of such prior investigations are rarely reported. Yet it is during such feasibility studies that critical decisions are made, such as site selection and definition of exposure. Here we present the details on one such feasibility study, conducted to determine the possibility of a cohort mortality study of workers exposed to ethylene oxide. Issues discussed include methods for estimating sample size and power, for estimating levels of exposure, and for assessing the adequacy of personnel records.

Key words: epidemiology, occupation, cohort studies

INTRODUCTION

Large retrospective occupational cohort studies are frequently conducted to determine whether a particular occupational agent is associated with a given disease. The usual advantage of such studies, in contrast to population-based case-control studies, is that they have good statistical power for studying occupational exposures that may be relatively rare in the general population, and that they are based on well-defined occupational histories. Such studies may be uniquely capable of testing a certain hypothesis.

The disadvantages are that occupational cohort studies require large populations, are quite expensive, and take a long time to do. Before undertaking such a study, it is well worthwhile for investigators to determine if their planned study is feasible.

The goal of a "feasibility" study is to determine if an exposed population is available that has: 1) adequate records to determine work history; 2) adequate data on exposures; 3) no serious confounding exposures; and 4) an adequate sample size to actually test the hypothesis in question.

In practice, such feasibility studies are usually conducted by all investigators before beginning a major study. They may range from informal phone calls to formal protocols and site visits. In addition to the critical decision about whether the overall study is feasible, feasibility studies frequently involve key decisions about how the major study itself will be conducted. While the basic legwork involved and the

National Institute for Occupational Safety and Health, Cincinnati, Ohio
Address reprint requests to Dr. Kyle Steenland, Mailstop R-15, National Institute for Occupational Safety and Health, 4676 Columbia Pkwy., Cincinnati, Ohio 45226.
Accepted for publication May 15, 1987.

conclusions reached are often the subject of informal discussions among epidemiologists, the details of such investigations are rarely published.

We here discuss some of the issues involved in conducting such a feasibility study. We use as an example an ongoing cohort mortality study of workers exposed to ethylene oxide (ETO), being conducted by investigators at the National Institute for Occupational Safety and Health (NIOSH), in collaboration with the National Cancer Institute (NCI).

METHODS OF ASSESSING FEASIBILITY

Hypothesis to be Tested and Choice of Industry to be Studied

Following the lead of earlier reports [Hogstedt et al, 1979a,b], the principal hypothesis to be tested in our study was that leukemia (or all lymphopoietic disease) is associated with exposure to ETO. We also wanted to be able to analyze the data for a dose-response, so that quantified estimates of past ETO exposure and some range of exposures, were necessary.

Because leukemia is a rare disease, one of the first concerns in designing this study was to identify a cohort of workers (known to have been exposed to significant levels of ETO), which was large enough to test the hypothesis. Our goal was to be able to detect a relative risk of approximately 2 for leukemia, with a power of about 80%. A second concern was that workers included in the study not be exposed to other potential leukemogens that might confound the study findings. A third concern was that the cohort had to be sufficiently concentrated in a finite number of locations to make the study logistically possible.

An initial review of the literature [Koketsu, 1978; NIOSH, 1981] indicated that manufacturing sites involved too few people, had mixed exposures (chemical plants), and had low exposures. (Because ETO is explosive, 100% ETO is produced outdoors, in closed systems.) Hospitals [Garry et al, 1979] had the largest population of exposed individuals but using these institutions would have required assembling a cohort from hundreds of hospitals, in each of which only a handful of people were exposed.

The medical supply industry appeared to have a large number of workers exposed. ETO is used indoors (often diluted with freon or carbon dioxide), and existing literature indicated that exposures were relatively high prior to the installation of engineering controls beginning in the late 1970s [HIMA, 1982; NIOSH, 1981; EPA, 1978; OSHA, 1983]. The literature also indicated that there were few known confounding exposures in the industry, with the exception of companies that produced pharmaceuticals. Furthermore, the number of companies in the medical supply industry, although large, did not seem prohibitive. It was not clear at this early stage how many sites would have to be included, but it was clear that the medical supply industry would involve far fewer sites than hospitals.

At this point, a preliminary proposal for a retrospective cohort mortality study in the medical supply industry was drafted. This proposal called for a feasibility study, to be conducted into two phases: 1) a mail and telephone survey; and 2) a series of walkthrough surveys at selected plants.

Mail and Telephone Survey

Using existing literature [EPA, 1978; OSHA, 1983; HIMA, 1982], we identified approximately 200 sites where ETO had potentially been used to sterilize medical

supplies. An initial telephone survey was conducted, and 75 companies were identified that had used ETO for sterilization purposes. These companies were sent a letter requesting information on dates of ETO use, volume of use, and the approximate number of workers thought to be exposed. The letter defined workers presumed to be exposed as sterilizer operators, material handlers, quality assurance personnel involved with sterile products, and any other personnel thought to be incidentally exposed. While we could not be sure that the company data would be accurate, we believed such data could be used to develop broad estimates of the extent of exposure at the plants in question. These broad estimates could be refined during later walkthrough surveys.

Pursuant to the Occupational Safety and Health Act, NIOSH is the government agency officially mandated to conduct safety and health research, and, if necessary, NIOSH has the legal authority to subpoena company records (Public Law 91-596, 91st Congress, Dec. 29, 1970). Other investigators who do not have this authority may have lower response rates to their inquiries and may be more concerned with the representativeness of their sample.

The response rate for our phone and mail survey was virtually 100%. Based on the information received from the companies, preliminary estimates were made of the number of person-years expected to be contributed by each plant. In general, these estimates were developed by taking the estimated number of workers exposed to ETO at the time of first use, and then estimating some rate of turnover of this population (eg., 5–10% a year). Using the expected follow-up period (year of first exposure until end of study) for this dynamic population, total person-years were then estimated. The follow-up period varied depending on the year of first ETO exposure at the plant, but on the average it was approximately 1968–1982.

Based on the results of the mail survey, it appeared that the estimated number of person-years (approximately 116,000) might be enough to result in a study of sufficient power (80% power to detect a relative risk of 2.7 for leukemia). We then solicited, and received, the cooperation of the trade association of the medical supply industry (Health Industry Manufacturers' Association).

We then developed a protocol for the study and began a process of review of the protocol by a peer review committee as well as by interested parties (companies and unions), following standard NIOSH procedures. The protocol established criteria for defining exposure and for inclusion of specific plants in the study (see below). Following this protocol review, we then selected 39 companies with the largest potential contribution of person-years for further feasibility assessment via walkthrough surveys.

Walkthrough Surveys

The primary objective of the walkthrough surveys was to establish the feasibility of including each individual plant site in the study. Based on the mail survey, some reasonable criteria were chosen for including a plant in the study. These criteria were as follows: 1) the plant must contribute at least 400 person-years to the study; 2) the plant must have adequate records for identifying past and present workers potentially exposed to ETO; and 3) the plant must not have any serious confounding exposures to a known or suspected leukemogen (eg., radiation, benzene).

The criteria for a lower limit for person-years were rather arbitrarily chosen, based on our mail survey, as a point that seemed to separate the larger plants from

the smaller ones. Furthermore, to ensure sufficient potential latency, to restrict the cohort to those with relatively higher exposures, and to eliminate exposures of short duration, it also was decided that to be eligible for the study an individual would have to have been exposed for at least 3 months prior to 1978. We anticipated a follow-up through 1984, and the 1978 cutoff thereby ensured a minimum of 7 years potential latency for cohort members. Given that the principal outcome of interest was leukemia, which has a shorter latency period than solid tumors, this seemed a reasonable cutoff point. This cutoff point also ensured that cohort members would have been exposed to relatively higher levels of ETO, since engineering controls were generally installed in the industry starting in 1978 [OSHA, 1983].

We first conducted an initial pilot study at five plants in order to test our assumption that the medical supply industry was a good choice for the study. In general, we found that these plants would potentially contribute more person-years than estimated from our initial crude assessment based on the mail survey. Exposure records frequently documented exposure not only to those directly exposed (eg., sterilizer operators), but also to large numbers of workers incidentally exposed to the gas. Furthermore, these five plants appeared to qualify for study because they had adequate personnel records, and because there were no significant potential confounding exposures to any suspect leukemogens.

The walkthrough survey for all plants was conducted by an epidemiologist and an industrial hygienist. At each site, we began with an opening meeting during which we discussed the history of ETO usage, industrial hygiene data, availability of personnel and other records, and confounding exposures. We then went on a plant tour focusing on areas where potential ETO exposures may have occurred. Key personnel on such a tour included the company industrial hygienist responsible for monitoring ETO levels, and long-time company or union personnel who were familiar with past production practices. Then the NIOSH industrial hygienist reviewed any ETO sampling data possessed by the company while the epidemiologist reviewed company personnel records. A preliminary decision was then made, in consultation with company and union personnel, about which job categories were likely to have involved ETO exposure.

Practical aspects of the procedures used in the walkthrough surveys are discussed below.

Adequacy of personnel records. Personnel records form the basis for the identification of study populations in most occupational retrospective cohort mortality studies. Adequate personnel records must include name, social security number, date of birth, and a work history sufficiently detailed to identify where an individual had worked over time. Information on sex and race is also desirable but was not essential for our study, since our investigators have the ability to obtain this information from the Social Security Administration. Job and department codes used in the company personnel files had to be sufficiently specific to identify employment in areas of the plant that had potential for exposure to ETO.

We conducted an inventory of records for both current and terminated employees. This inventory included identification of the types of records used, the period of time covered by the record system, and a total count of the number of records. If personnel records or other record systems could not be located for former employees, the plant was immediately excluded from consideration for our study.

The work history portion of the record was the most critical. Based on a plant tour and a review of hygiene records (see "Industrial hygiene considerations" below), an initial decision was made about which departments and jobs were likely to have involved exposure to ETO. It was critical that the work history contain sufficiently detailed information about job and/or department in order to determine at any given point in time whether an individual was exposed. In this study, job category was often not as important as department, because a whole area including many job titles would be exposed to the same level of gaseous ETO. An exception to this general rule was the sterilizer operators, who were assigned to a variety of departments, depending on the specific plant.

Minimally, then, the work history had to include departmental changes over time. If department names had changed over time, it was necessary to know exactly how they had changed and when. It was preferable if numbers were used to identify department on the work history, facilitating later coding. When only alphabetical names of departments were found, they had to be sufficiently consistent over time so that we could reasonably assign numerical codes to a finite number of alphabetical variations.

We also assessed whether personnel records appeared to be complete. Personnel records for terminated personnel are often arranged by year of terminaton. We checked to see if terminated files were available for all past years when ETO had been used. We also randomly checked files for alphabetical consistency. If we found that any large number of personnel records were missing, we excluded the plant, on the presumption that the missing records might introduce a bias in our findings. Two plants were excluded for missing records at the time of the walkthrough survey.

We also looked for any alternative lists of who had worked at a plant (e.g., payroll records, union records), which could later be used in a systematic computerized check of the completeness of personnel records. Our procedure generally is to microfilm all personnel rcords at the plant. We then create an index file of all these individuals, and only later select the exposed from this index file, based on their work in certain departments. Thus the index file represents all those who have worked at the plant, and it is the index file that we will use to match to alternative lists of employees.

Although most plants kept payroll records for the 7 years required by IRS, what we needed were alternative lists that went back in time to the point at which ETO had first been used. Only two plants had kept payroll records consistently back this far in time. One or two other plants with unions had kept seniority lists, which may prove useful as alternative lists.

If the company does not have such alternative lists for cohort verification, it is sometimes possible to use IRS quarterly earnings reports which are on file at the Social Security Administration. The permission of the company in question is required to obtain these records. Other investigators have used quarterly earnings reports to verify their cohort [March and Enterline, 1979]. There are practical problems encountered in using quarterly earnings reports [Nelson and Van Peenen, 1985]. One problem we encountered here is that some corporations have filed quarterly earnings reports that have combined individuals at several different plants, making cohort verification impossible.

When we discovered, during the walkthrough survey, that a plant had a significant number of missing records (e.g., no personnel records for terminated employees

with less than 10 years at the plant), we excluded that plant. However, it should be mentioned there are no clear guidelines as to what is "significant." Furthermore, while it may be intuitively desirable for personnel records to be complete, there may be little actual bias introduced in an assessment of cause-specific mortality by including plants with missing records.

Sample size and person-years estimation. Our method of estimating the number of workers exposed, and the number of person-years that they would contribute to the study, involved drawing a 5% random sample from personnel files at each plant. We then determined the percentage of individuals in this sample who were exposed (based on our initial decision about which jobs/departments involved exposure). This percentage was then extrapolated to the whole plant population. From the sample, for the exposed individuals, we calculated an average year of birth and average first year of exposure. Using these averages, and the estimated number of people exposed, we then calculated the estimated numbers of person-years to be contributed by the plant in question. For example, let us assume that our 5% sample showed that 30% of the workers were exposed, and let us also assume that the average year of first exposure for these individuals was 1968. If the total number of workers at the plant over time was 2,000, then 600 would be expected to have been exposed. Assuming 15 years of follow-up on the average (1968–1982), we would expect this plant to contribute 9,000 (15 × 600) person-years to the study.

Using the average age of this cohort at the midpoint of follow-up (1975), we could then use a U.S. leukemia death rate to estimate the number of expected leukemia deaths in each plant cohort. Based on the five plants in our pilot survey, we estimated that our follow-up period would be (on the average) 1968–1982, and that the average age of an exposed worker followed at the midpoint of the follow-up period (1975) would be about 40–44. Leukemia death rates do not differ greatly by sex or race. Nor have these rates changed much over the last 15–20 years, the period of the study. Hence we used an overall leukemia death rate (all sexes, races combined) in 1975 (3.3 per 100,000 deaths), to estimate expected leukemia deaths.

Industrial hygiene considerations. The industrial hygiene component of the walkthrough survey was directed toward assessing: 1) which jobs were likely to have been exposed to ETO; and 2) the quality of existing industrial hygiene data that could be used to quantitate employees' exposures to ETO.

Determining which job categories were likely to have been exposed was essential for estimating the total number of people exposed over time. Based on the plant tour and a review of existing industrial hygiene data, a preliminary decision was made about which jobs and departments had involved ETO exposure. The best documentation of exposure was the existence of actual industrial hygiene data. In those areas where industrial hygiene data had not been collected, we relied on the professional judgement of our industrial hygienist to determine if the possibility existed.

Quantitative exposure data were necessary if any type of dose-response analysis were to be conducted later. For those years when such quantitative data did not exist, or for areas of a plant likely to have been exposed but where no sampling data had been collected, we planned to create a mathematical model based on existing industrial hygiene data, which would enable us to estimate the unknown exposures. Using such a model, we planned to create an exposure matrix that could be used to estimate the exposure for each job in the study, at any given point in time.

Most companies surveyed did not begin to sample for ETO before 1978, at which time the first data indicating the potential carcinogenicity of ETO were published. Many of these early samples were collected with direct reading instruments and represented only area exposures, which are not generally representative of an individual's exposure (as measured with personal breathing-zone samples). However, the area samples were useful to estimate the potential exposure for people in a given job, as well as to document the extent of migration of ETO throughout the plant. By 1979, most companies had begun to use some type of personal sampling device. These personal samples were used to develop the exposure model, whenever possible.

It was also about this point in time (the late 1970s) that most companies began to install a variety of engineering controls in an attempt to reduce exposure to ETO. Some plants had conducted industrial hygiene sampling before and after such installation. We consider these data very useful for the construction of the exposure model, because they enable us to estimate the effect of engineering controls. Knowing the effect of such controls enables us to estimate earlier exposures at plants for which no sampling data exists prior to the installation of engineering controls.

Although we usually relied on existing company industrial hygiene data, we occasionally did our own sampling, if our hygienist believed we needed either to supplement or to validate company data.

RESULTS

Person-Years Estimate and Power Calculations

Of the 39 plants selected for consideration from our mail survey, 35 were visited, and 14 plants were selected for inclusion in our study.

Approximately half of the original 39 companies proved to be ineligible either because of missing records, inadequate number of workers exposed, or records inadequate for identification of the exposed. In addition, we decided to exclude those plants that lacked any industrial hygiene records; this led to the elimination of four plants.

Table I shows a comparison of the number of person-years for the 14 plants based on the estimates from the mail survey and the walkthrough surveys. Based on our walkthrough surveys, we have estimated that the study will include a total of approximately 19,000 exposed workers contributing approximately 261,000 person-years of observation to the study. In most cases, our original estimate of person-years from the mail survey was an underestimate. This was because, in the mail, survey most companies only identified workers directly involved in operations using ETO as being potentially exposed, without recognizing the extent of gas migration. In many instances during our walkthrough surveys, we discovered that other areas of the plant, and in some cases the entire plant, were potentially exposed to ETO as a result of inadvertent gas migration.

Knowing the approximate number of person-years at the 14 selected plants, based on the walkthrough surveys, enabled us to estimate the statistical power of the future study to detect a specific relative risk (Table II), as described previously in the Methods section.

We estimated that under the null hypothesis of no association between ETO exposure and leukemia, we would observe 8.6 cases of leukemia in our study. Using a formula developed by Beaumont and Breslow [1981], we further estimated that our

TABLE I. Person-Years Based on the Mail Survey Versus the Walkthrough Survey at the 14 Plants Chosen for the Study*

	Mail survey		Walkthrough survey	
Plant	Estimated no. exposed	Estimated no. person-yrs	Estimated no. exposed	Estimated no. person-yrs
01	60	470	1,530	18,000
02	170	690	2,750	19,300
03	680	8,800	510	7,200
04	2,590	45,700	1,530	21,420
05	400	5,240	3,691[a]	51,300[a]
06	150	1,550	539[a]	6,000[a]
07	30	460	260	2,600
08	290	2,740	940	13,100
09	90	1,100	470	4,940
10	70	820	1,540	20,000
11	30	280	790	10,000
12	90	960	1,650	18,900
13	70	550	2,080	24,500
14	30	250	970	26,100
Total	4,750	69,610	19,229	261,360

*For both mail and walkthrough surveys, individuals had to have worked in an exposed area prior to January 1, 1978 to be considered exposed. The mail survey made no requirement for duration of exposure. To be considered exposed in the walkthrough surveys, an individual was required to have worked in an exposed department for at least 3 months.

[a]At these two plants, we were unable to estimate the number exposed or the person-years during the walkthrough surveys because subsequent industrial hygiene sampling indicated that more departments were exposed than had been known at the time. The data presented here for these two plants were derived after microfilming and computerizing the personnel records from these plants.

TABLE II. Estimates of Minimal Detectable Relative Risk (SMR) for 14 Plants Included in Study, Based on Person-Years Estimated During On-Site Surveys*

Cancer type	No. cancer deaths expected[a]	Minimum relative risk detectable with 80% power[b]
Leukemia	8.6	2.0
All hematopoetic	19.9	1.7

*Mortality rates were derived from 1975 U.S. Vital Statistics for males and females combined, age group 40–44 (all race groups combined).

[a]At two plants (5 and 6), we were unable to estimate person-years during the walkthrough. Person-years were derived after filming and computerizing the personnel records from these plants.

[b]Assuming Alpha = .05 (one-tail).

study would have 80% power with an alpha level of .05 (one-tail) to detect a relative risk of 2.0 for leukemia, and 1.7 for all hematopoetic cancers.

We have now filmed the personnel records at all of the 14 plants, and detailed work histories have been coded for six of them. We have used standard life table analyses [Waxweiler et al, 1983] to calculate actual person-years at risk at these coded plants, as well as to calculate expected leukemia deaths.

The number of person-years estimated from 5% samples taken during our on-site surveys corresponds quite well to the actual person-years available for the entire

cohort at each plant based on the coded work histories (see Table III). This indicates that our procedure of sampling 5% of the personnel files gave an adequate estimate of person-years.

Although our original estimates of person-years at each plant were fairly accurate, our method of calculating expected deaths underestimated the actual expected deaths. The reason for this discrepancy is that we derived our expected leukemia deaths using an "average person-year" at each plant (see above). This procedure does not give sufficient weight to the contribution of older person-years to the expected number of deaths.

Leukemia rates go up exponentially with age (Fig. 1). The distribution of person-years, on the other hand, is approximately bell-shaped, but slightly skewed with a longer tail on the right (higher ages). Figure 1 also shows the distribution of person-years at plants 2 and 4. At plant 2, the cohort is relatively young (ETO was first used in 1968). Most plants in the cohort have this kind of distribution, and it was from this typical distribution that we estimated that the "average" person-year would involve a leukemia rate of 3.3 per 100,000. At plant 4, the cohort is older (ETO first used in 1957), and the curve is slightly shifted to the right compared with plant 2.

Table 4 shows the estimated expected leukemia deaths based on using a single death rate (3.3 per 100,000), versus the actual expected leukemia deaths using age, race, sex, and calendar-time specific leukemia death rates. The discrepancy between estimated and actual expected leukemia deaths is greatest at plants 3 and 4, sites where a large number of person-years were accumulated by older individuals.

In order to avoid the above problem, investigators merely need to collect all the necessary data when they sample the personnel records, so that the sample of the exposed may be divided by age, sex, race, and calendar-time; the appropriate leukemia rates may then be used to estimate expected deaths. This will obviously involve more work, but it would avoid discrepancies, as seen in Table IV.

Exposure Assessment

While work in developing a model to estimate past exposures is still in its initial stage, the walkthrough surveys did provide us with data on over 1,000 personal

TABLE III. Estimated (From Random Sample) and Actual (From Complete Coded Data) Person-Years at 14 Plants Included in the Study*

Plant	Estimated person-years[a]	Actual person-years[b]
1	18,000	16,991
2	19,300	19,781
3	7,200	8,831
4	21,420	28,174
10	20,000	18,720
11	10,000	11,759

**Estimated person-years are derived from a 5% random sample of personnel files conducted during walkthrough surveys, while actual person-years reflect the entire population after personnel files were coded and computerized. All person-years are for individuals exposed for at least 3 months prior to January 1, 1978.

[a]Based on 5% sample of personnel files.

[b]Based on entire cohort at plant.

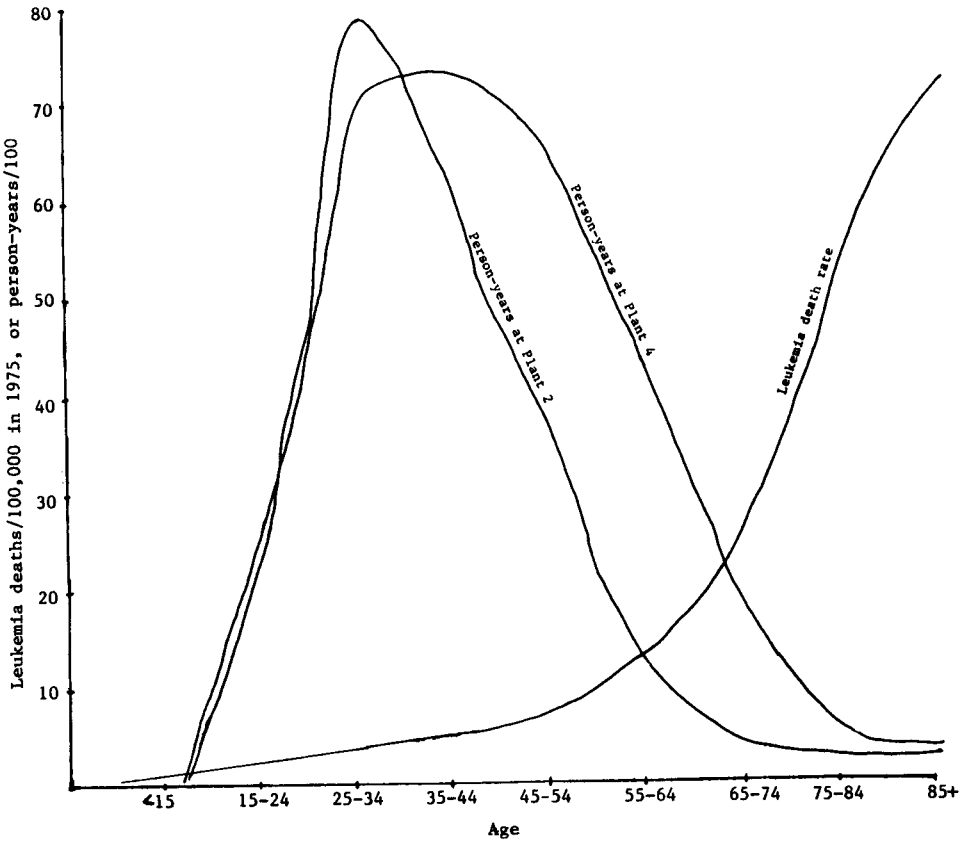

Fig. 1. Leukemia death rates and person-years at risk at two plants.

TABLE IV. Estimated Leukemia Deaths at Six Plants, Based on an "Average" Rate and on Age, Race, Sex, and Calendar-Time Specific Rates, Applied to Person-Years as Determined by Life-Table Analysis

Plant	Expected leukemia deaths using "average" rate[a] of 3.3/100,000 person-yrs	Expected leukemia deaths based on age, race, sex, calender-time, specific rates/100,000 person-yrs
1	.56	.46
2	.65	.58
3	.29	.51
4	.93	1.50
10	.62	.68
11	.39	.35
Total	3.44	4.08

[a]"Average rate" was determined for an average person-year in our cohort. An average person in our cohort was first exposed in 1968, at age 36. An average person-year would occur at the mid-point of the follow-up period (1975), at age 43. Hence the leukemia rate chosen was for a person aged 40–44 in 1975.

samples collected from 1977–1985, as well as numerous area samples. We are planning to rely on personal samples only in developing the regression model to predict exposure. Since many men in the same job were sampled more than once in the same year, we are calculating averages. The 1,000 samples will represent approximately 300 outcomes to be used in a weighted multiple regression. Predictor variables that we are coding for each sample include year or sampling, the presence or absence of engineering controls at the sterilizer, the presence or absence of aeration at the site of off-gassing product, type of product sterilized, job category, and the amount of ETO used over the specific calendar year at a given plant. Valuable data for developing the model are being contributed by plants from which we obtained sampling data during the walkthrough survey but that will not be used in the study because of problems with sample size or personnel records. A few plants with good sampling data will be excluded in developing the model, so that levels of exposure predicted by the model can be tested against actual data at these plants. While it is too early to conclude that a useful model with good predictive power can be created, it appears that sufficient data exist to attempt to build such a model.

We have mentioned earlier that our first estimates of who was exposed included primarily those directly exposed (e.g., sterilizer operators), while excluding many who were later determined to have been indirectly exposed (e.g., production workers). It should be pointed out that while, in general, those directly exposed would appear to have higher levels of exposure (e.g., 10–50 ppm over an 8-hour period), those who were indirectly exposed are not simply contributing a large number of person-years to the study at very low levels of exposure. Production workers in areas where ETO was in the air were sometime exposed to low background levels of about 1 ppm, but in other cases entire production areas were exposed to higher levels, such as 5–15 ppm.

DISCUSSION

It is clear from the above that while it may be possible to do a feasibility assessment solely by correspondence or phone, a more accurate assessment can better be accomplished with site visits. In our case there was no single "feasibility study." Instead, we conducted a phone survey of 200 plants, a mail survey of 75 plants, and a walkthrough survey of 35 plants. Finally, 14 sites were chosen for the study. Hence our feasibility study was really a series of steps in which we continued to learn more about the cohort, a kind of iterative procedure.

The walkthrough surveys provided, prior to the microfilming of the personnel records at any plant, a definition of who was exposed at each plant and a description of industial hygiene data and process changes over times. Obviously, such data are critical to the study. Thus the process of determining which plants would be included provided the essential data for later construction of the exposed cohort and coding of the work histories. Furthermore, cost was reduced by determining before microfilming that certain plants were to be excluded.

Our experience, as described above, may not be "typical," because of the nature of the agent being studied and the hypothesis in question. Nevertheless, the basic questions that must be answered in any study prior to beginning data collection are the same. First, the investigators must determine, based on a review of the literature, what is the hypothesis of interest. Second, an exposed cohort must be

identified. This cohort must be exposed to the agent in question at levels that are of interest to the investigator. The cohort also must be without serious confounding exposures. Work and exposure histories of the cohort must be adequately documented. Finally, the exposed group must be large enough so that there is enough statistical power to provide meaningful information about the hypothesis.

There is some question as to whether the site or sites chosen for a study are "representative" of the entire universe of the exposed population. However, the key point is not that the chosen sample is representative (as in a random sample), but rather that the level of exposure of the sample is of interest. Clearly, there is no reason to study a plant where the employees are all exposed to 1 ppm of a substance if other workers (in adequate numbers) are exposed to higher levels, and if it is still unknown if the higher levels cause disease.

The danger of not adequately addressing the above-mentioned questions before beginning a study is that a great deal of time and money may be spent in a study that ultimately may not be feasible, or the study may be completed but be of little value due to inadequate sample size or confounding exposures.

It would be helpful if investigators not only conducted thorough feasibility studies, but also described them in some detail in the methods sections of their papers. We have suggested some specific techniques for assessing the adequacy of personnel and exposure records and for sample size estimation. These techniques are grounded in common sense, and other investigators may use different methods. The general principles involved, however, will remain the same.

We are not advocating that our procedures here should be applied to all cohort studies. Other investigators studying other substances will have different problems. Our goal here has merely been to indicate the methods we used to determine the feasibility of our study, in the hopes that some of the lessons we learned can be applied elsewhere.

REFERENCES

Beaumont J, Breslow N (1981): Power considerations in epidemiologic studies of vinyl choloride workers. Am J Epidemiol 114:725.

EPA: Preliminary benefits analysis of ETO as a medical product sterilant. (1978) Mitre Technical Report 7774, EPA contract 68-01-3965.

Garry V, Hozier J, Jacobs D, Wade R, Gray D (1979): Ethylene oxide: Evidence of human chromosomal effects. Environ Mutagen 1: 375.

Health Industry Manufacturers Association (HIMA) (1982): Ethylene oxide worker safety seminar, Chicago, IL.

Hogstedt C, Rohlen O, Berndtsson B, Axelson O, Ehrenberg L (1979a): A cohort study of mortality and cancer incidence in ethylene oxide production workers. Br J Ind Med 39: 276.

Hogstedt C, Malmqvist N, Wadman B: Leukemia in workers exposed to ethylene oxide. (1979b) JAMA 241: 1132.

Koketsu M: Ethylene oxide industry-wide industrial hygiene study. (1978) Submitted to the National Institute for Occupational Safety and Health under Contract No 210-76-0162, SRI International, Menlo Park, CA.

Marsh G, Enterline P (1979): A method of verifying the completeness of cohorts used in occupational mortality studies. J Occup Med 21: 665.

Nelson N, Ven Peneen P: Use of Social Security Administration quarterly lists for occupational cohort verification. (1985) Letter to Editor, Am J Epidemiol 122: 5.

NIOSH: Current intelligence bulletin 35-ethylene oxide. (1981) NIOSH Publication 81-130, NIOSH, 4676 Columbia Pkway, Cincinnati OH 45226.

OSHA: Ecnomic and environmental impact study of ethylene oxide, (April 18, 1983). Prepared by JRB Associates, Contract No J-9-F-2-0069.

Waxweiler R, Beaumont J, Henry J, Brown D, Robinson C, Ness G., Wagoner J, Hyg S, Lemen R (1983): A modified life-table analysis for cohort studies. J Occup Med 25: 115.