SECTION II

EPIDEMIOLOGICAL DATA

EPIDEMIOLOGICAL APPROACH TO EVALUATION OF TOXIC EFFECTS

by
T. Partanen

This paper is a short survey on epidemiological strategies and techniques in the evaluation of toxic effects. Emphasis is put on outcome and exposure indicators, as well as evaluation of effects, with the understanding that epidemiological research should be added to knowledge from experimental and other scientific activities in the field (1).

Outcome Indicators

Epidemiology often deals with binary types of variables, e.g. cancer/no cancer, exposure to agent X/no such exposure. In this case, the outcome indicators refer to what Miettinen (unpublished document, 1979) calls "measures of occurrence" (of disease, sign, symptom, etc.): that is, outcomes in the health-disease system.

The basic epidemiological outcome indicators are prevalence and incidence rates. Prevalence rate is the proportion of subjects with an outcome condition (disease, etc.) out of the target group or population at issue. For example, in a group of 343 rayon viscose workers exposed to carbon disulfide (CS_2), 16.8% reported angina pectoris (AP) in a standardized interview done in 1967 (1). This percentage is the prevalence rate of AP in this group at that time. Whether or not this finding indicates something of the effect of CS_2 is obscured until answers have been given to such questions as: What would the prevalance have been without exposure? Have some of the workers stopped working in the plant because of disease? What are the determinants other than CS_2 exposure that may affect the rate? Answers to these questions were attempted in the study by Hernberg et al. (2).

Incidence rate is either directly related to person-time space (incidence density or force of morbidity) or is cumulative (Miettinen, unpublished document, 1979). The incidence rates are based on new ("incident") cases (of

disease, etc.) occurring in a population during follow-up. The follow-up may be retrospective, prospective or both. The incidence density is the number of new cases per population and time (i.e. person-years). The follow-up usually concerns groups (cohorts) without the condition at the beginning of the follow-up. For example, the incidence of mortality from malignant neoplasms in Finnish locomotive engineers (drivers) between 45 and 54 years of age was estimated to be 1.76/(1000 y) in 1967 (3).

Cumulative incidence rate is the proportion of a population which has contracted the disease (etc.) during a selected time period. For example, the cumulative coronary mortality in a group of CS_2 exposed workers in a 5.5-year period was 4.7%, whereas the same figure for a comparable, nonexposed group was 0.9% (4).

Certain relations exist between prevalence and incidence rates, dependent on certain assumptions concerning, for example, average duration of the disease. A number of outcome indicators are inferred from the basic rates. Thus, rate ratio (relative risk[a] or risk ratio) or RR is the ratio of two incidence (or prevalence) rates. For example, Hernberg et al. (4) estimated that the RR for coronary heart disease (CHD) death for the CS_2 exposed and nonexposed was 5.6; that is, the mortality of the exposed was 5.6 times that in the nonexposed. The RR may be estimated in different ways; in addition, it can usually be estimated in a case control design even when the risk components (disease incidence rates) are unknown.

Risk difference (attributable risk) or RD is the difference between two risks or rates. For example, the cumulative RD estimate in the example above was 3.8/(1000 x 5.5 y); that is, in 5.5 years, 3.8 additional CHD deaths per 100 persons occurred over the reference, or nonexposure, rate.

---

[a] Risk refers here strictly to the probability of contracting a disease for a person with fixed characteristics (e.g. age) during a fixed time period. Risk may also loosely refer to prevalence or incidence rates. Hence, epidemiologists speak of "risk ratios".

A frequently encountered indicator is the odds ratio (OR). It is the cross-product ratio ad/bc of the frequencies (numbers of persons) in the double-binary set-up below.

	Diseased	Non-diseased
Exposed	a	b
Nonexposed	c	d

The OR is usually an estimate of the RR regardless of whether the design is a cohort (follow-up) or case control type. For example, from a case control study of childhood cancer according to maternal occupation, the OR for cancer/no cancer versus maternal occupational branch of the food industry/other was estimated to be 4.01, an estimate of RR (5).

The risk difference (attributable risk) can be expressed relative to the reference risk (incidence), to the incidence in the nonexposed or to the incidence in the population in general (exposed and nonexposed combined). The first of these alternatives is the attributable risk per cent or etiological fraction (for the exposed):

$$(1) \quad ARP = \frac{R_e - R_r}{R_e} \times 100 = (100).(RR-1)/RR$$

where R_e = risk (incidence) in the exposed;
R_r = risk (incidence) in the nonexposed.

The difference $R_t - R_r$ (incidence in the total population minus that in the nonexposed) can be expressed as the proportion of the (weighed) population incidence R_t (exposed and nonexposed combined). This proportion is the population attributable risk PARP = (100) $(R_t - R_r)/R_t$. For example, the recent controversial statement that 20%-40% of all cancers are attributable to occupational factors (6) is in essence an estimate of PARP. Cole & Merletti (7) point to some difficulties in the interpretation of PARP in general and, specifically, in the case of Bridbord et al. (6). PARP can also be estimated in a case control study (8).

Finally, the standard(ized) morbidity (or mortality) ratio (SMR) is the actual number of deaths (or "incident cases") in the exposed divided by the "expected" number of deaths, the latter being calculated as the product of the death rate in the reference population and the number of the exposed (i.e. hypothetical number of deaths in the exposed if the referent rate were operating). Usually, the SMR is calculated over strata of a factor that has a different distribution in the exposed and the nonexposed, or reference, population. The set-up is as follows.

Factor level (stratum) (e.g. age group)	Number of deaths in the exposed (D_e)	Number of exposed (N_e)	Mortality rate in the reference population (R_r)
1			
2			
.			
.			
.			
k			

Here, SMR = (observed)/(expected) = $\dfrac{\sum_i C_e \Big/ \sum_i N_e}{\sum_1 R_r N_e}$

It is usually multiplied by 100. For example, in a recent study (9), the following results (Table 1) emerged from a 19-year period of retrospective follow-up of 1413 men exposed to fibrous glass. The authors state that "digestive and urinary cancers are elevated, although differences between observed and expected are not statistically significant".

Table 1. Results from 19-year period of retrospective follow-up of 1413 men exposed to fibrous glass (9)

Cause of death (7th rev.)	Observed	Expected[a]	SMR
All cancer (140-205)	40	45.1	88.7
Digestive (150-159)	18	13.6	132.9
Respiratory (160-164)	9	13.7	65.8
Genital (177-179)	1	2.9	34.7
Urinary (180-181)	5	2.4	207.7
Leukemia, aleukemia (204)	2	2.0	100.0
Lymphomas (200-203, 205)	3	2.9	101.8
All other cancer	2	7.5	26.6

[a] Based on United States experience.

The RR, RD, ARP, PARP and the SMR are the most frequently used outcome indicators in epidemiology[a]. They are often adjusted for factors that are correlated with either

[a] The proportional morbidity (mortality) ratio (PMR) should perhaps be added. It is usually the ratio of a disease-specific incidence and another (or "all causes") incidence.

exposure or outcome so as to ensure or at least increase comparability between series (e.g. of the exposed and the nonexposed) in these factors (e.g. age). When the basic outcome variate is continuous - as, for example, with a change in systolic blood pressure in mmHg - other indicators are usually used.

Selection and self-selection are often operating in industrial populations in the meaning that workers entering industry are healthier on the average than those not entering; a turnover of nonhealthy workers to lighter jobs or out of working life also occurs. Thus, mortality rates of industrial groups are usually between 60% and 90% of those in the population at large (10) - the so-called "healthy worker effect", which is the most marked at the beginning of exposure (if coincident with the beginning of the employment)[a].

Latency periods, or induction times, pose some problems in obtaining delayed effect indicators, as, for example, in the case of cancer. To quote Saracci (1):

> "By the time of the response, say a cancer excess, becomes known, the stimulus to be controlled, say a chemical carcinogen, may already have been removed and substituted by other substances whose effect will only show up in a further twenty or thirty years."

Cancer cases (partly) due to exposure may be expected earlier if one accepts that exposure sometimes acts as a "promoting agent" so that "fruits of prevention efforts do not have to wait through the 20-year latency period" (11). All in all, a tendency exists to search for early precursors or other indicators of early effects. Thus, subclinical indicators are often considered. Also, birth defects are under study because the latency periods may be relatively short (e.g. if exposure has taken place during pregnancy) and because some relation appears between teratogenesis (and mutagenesis) and carcinogenesis.

---

[a] Hernberg, S. Sixth international advanced course on epidemiologic methods, 1979: lecture materials. Helsinki, Institute of Occupational Health (mimeographed documents).

Relatively high specificity as to exposure is a "good" quality for an outcome indicator because the control for other covariates ("confounding factors") is simpler.

Exposure Indicators

A serious problem in "industrial epidemiology" is the difficulty in obtaining measures of exposure. This difficulty is a reason for the high prevalence of "binary" epidemiology: measure of exposure is often on the nominal, crude yes/no level. Exposures to "new" substances also pose problems, mainly due to the induction periods. Ideally, quantitated dose or exposure indicators would be needed for estimating the exposure-response relationships and for finding "safe" doses when such exist. Important dimensions are "final total dose" (e.g. in concentration and time), "final highest exposure intensity", maybe also "final duration of exposure", age at onset of exposure and transient exposure levels (10).

Hernberg (see footnote on page 94) has listed dimensions of measures of exposure and measurement methods. His list is duplicated in Table 2.

Assessment and Evaluation of Effects

Epidemiological effect evaluation may be thought as taking place at three levels: data screening (for associations, alarms, and hypothesis generation), assessment of existence of causality and quantification of exposure-response relationships. Data screening may utilize registers of the diseased and registers of exposures, including linkage between the two. The former registers include registers of cause-of-death and death certificates, hospitalization and hospital discharge registers, primary care records, disability registers, sickness benefit and other compensation registers, cancer registers, registers of other diseases, mass screening registers, industrial health records, injury registers and registers of special surveys. The exposure registers generated within the monitoring of exposures are relatively rare. Here, the main sources of information are measurements made in industry as well as a few registers or workers exposed, for example, to carcinogenic agents. Most often, however, registers of the diseased,

Table 2. Dimensions of measures of exposure and measurement methods[a]

Dimensions of exposure:
- Biological monitoring
- Air samples from breathing zone
- "Background" samples (static predetermined sites)
- Classification by work area, type of work, occupation
- Exposed/nonexposed
Measures of exposure:
- Quality
- Intensity
- Duration
- Fluctuations
- Calendar time
- Concurrent other exposure
Methods of measurement:
- Duration
- Measurement: retrospective, current, prospective
- Comparison of present situation by rating known changes and interviews with hygienists, foremen, workers
- Reconstruction

[a] For reference see footnote on p. 94.

or outcome registers, are used for case-finding; environmental data is added ad hoc, usually retrospectively.

An example of data screening is the Finnish study on childhood cancer and parental occupation (3). It is a case control study, with childhood cancer cases reported to the Finnish Cancer Register between 1959 and 1975 and controls drawn from maternity welfare centre files. Parental occupations were manually recorded at the maternal welfare centres for both cases and controls: cases were identified by name and person ID number. The maternal occupations which occurred more frequently among the cases than the controls included farmers' wives (only

during 1969–1975), pharmacists, saleswomen, bakers and unspecified factory workers. The apparent paternal risk occupations were farmers, motor vehicle drivers, machine repair men and painters. Also at risk were men who reported an academic degree instead of an occupation. In some of these occupations, exposures to harmful chemicals are possible, but chance correlations cannot be excluded.

Within one study, demonstration of causality is, strictly speaking, very difficult in the nonexperimental set-up which is usual in epidemiology. Statements of probability can be done, but they are not sufficient. Statistical inference is usually done with relatively powerful statistical tests of null hypotheses, such as RR = 1 or RD = 0 or by calculating confidence limits for outcome parameters of interest (Miettinen, unpublished document, 1979). Susser (12) points to "bias toward scepticism in conventional procedures of inference": "Much statistical strategy aims to avoid false positives, inferences that give credence to causality where none exists". He calls statisticians and epidemiologists "properly professional sceptics". Indeed, avoidance of false negatives has been a largely neglected matter; however, it is in the interests of those directly involved - the workers. In addition, statistical criteria for "significance" might be relaxed in medium-sized studies, especially when dealing with a severe outcome, as, for example, mortality.

Strength of statistical association should be considered separately from statistical testing and calculation of confidence intervals because sample sizes affect the latter. Thus, a practically insignificant association may become significant when using statistical inference techniques merely because sample sizes are very large. The matter of cleaning up the association from effects of other confounding factors is an important issue in nonexperimental research, especially when the response under consideration is relatively nonspecific with reference to its determinants.

Consistency of a finding through independent studies (also experimental) is becoming a standard criterion, as, for example, in the evaluation of carcinogenicity of chemical agents. Consistency of association through different strata of the target population also increases the credibility of the inference. Inconsistencies, however,

do not necessarily negate the inference because the effect of an agent may be modified by other factors. In this context, the statement is often made that demonstration of an exposure-response relationship, if not due to confounding factors, contributes to the credibility of causality.

The biological reasonableness of the association (criterion of coherence (12)) is usually considered a quality that increases one's belief in causation. It is, however, not a universally applicable criterion since, strictly taken, it has an empirical flavour of slowing down development of theoretical inventions that are necessary for the progress of science. Keeping the nature of this criterion in mind, it is probably in most instances, however, a sound measure in the evaluation of causality. Causality being demonstrated as obvious, the interest is usually turned to the form of exposure-response relations, interactions (synergisms/antagonisms) and effect-modifying external or host factors.

A simple example of an attempt towards empirical demonstration of an exposure-response relationship will be illustrated with the relationship between industrial styrene exposure and its effects on CNS functions. The current exposure was measured with urinary mandelic acid concentrations in workers by means of five determinations per worker during five weeks. Prevalence of EEG abnormalities, as specified in the study protocol, was used as one of the CNS indicators (13). Results are seen in Table 3. The "validity" index is the sum of the sensitivity and specificity in the fourfold tables generated for EEG abnormality and mandelic acid concentration in urine, using different cut-off points for the latter. The authors state that the validity index and the $\chi^2_{[2 \times 2]}$ had their maximal values at mandelic acid concentrations below 699 mg/l and that, at 800 mg/l, the prevalence of abnormal EEGs probably reaches a plateau.

Examples of results from some psychological tests measuring CNS functions, as related to exposure, are given in Table 4.

Table 3. Distribution of subjects into categories according to their electroencephalographic (EEG) findings and urinary mandelic acid concentrations

The statistical test results and validity calculations are derived from fourfold tables with cut-off points at the lower limit of the mandelic acid concentration category (13).

Mandelic acid concentration (mg/l)	EEG finding			χ^2	p	Prediction of EEG by mandelic acid concentration		
	Normal	Abnormal	Total			Sensitivity	Specificity	Validity
- 699	34	4	38	.	.	.	.	.
700- 799	7	3	10	4.60	<0.05	0.83	0.47	1.30
800-1199	10	7	17	3.66	(<0.06)	0.70	0.56	1.26
1200-1599	8	3	11	0.30	NS	0.39	0.70	1.09
1600-1999	4	1	5	0.04	NS	0.26	0.81	1.07
2000-	10	5	15	0.36	NS	0.22	0.86	1.08
Total	73	23	96					

Table 4. Distribution of subjects into categories according to their psychological test results and urinary mandelic acid concentrations
The statistical test results and validity calculations are derived from fourfold tables with cut-off points at the lower limit of the mandelic acid concentration category (13).

Mandelic acid concentration (mg/l)	Performance on psychological test			χ^2	p	Prediction of psychological performance by mandelic acid concentration		
	Better than or equal to cut-off value	Poorer than cut-off value	Total			Sensitivity	Specificity	Validity
Symmetry Drawing								
- 399	20	2	22	.	.	.	.	.
400- 799	19	5	24	3.19	NS[a]	0.92	0.28	1.20
800-1199	15	3	18	4.34	<0.05	0.72	0.55	1.27
1200-1599	9	3	12	9.26	<0.01	0.60	0.76	1.36
1600-1999	3	2	5	12.98	<0.001	0.48	0.89	1.37
2000-	5	10	15	12.84	<0.001	0.40	0.93	1.33
Total	71	25	96					
Bourdon-Wiersma Vigilance Test								
-1199	58	8	66	.	.	.	.	.
1200-1599	10	2	12	1.76	NS[a]	0.50	0.71	1.21
1600-1999	5	-	5	2.30	NS[a]	0.38	0.83	1.21
2000-	9	6	15	5.36	<0.05	0.38	0.89	1.27
Total	82	16	98					
Mira Test								
-1199	54	12	66	.	.	.	.	.
1200-1599	9	3	12	4.59	<0.05	0.52	0.74	1.26
1600-1999	3	2	5	6.39	<0.05	0.40	0.86	1.26
2000-	7	8	15	5.59	<0.05	0.32	0.60	1.22
Total	73	25	98					

[a] Nonsignificant

Theoretical work has been done in the modelling of exposure-response relationships and the effects of multiple exposures. Linearity of the response to low-level exposure to carcinogens is under discussion but is hard to demonstrate with epidemiological data. Multivariate models for categorical and continuous data have been developed, also with provision for interactions (14-16). Multivariate models of linear and logistic type have been fitted to epidemiological data. The latter takes the form:

$$(2) \qquad \ln(p_D/(1-p_D)) = \sum_{i=1}^{k} b_i X_i$$

where p_D = probability of developing disease in a period;

$\{b_1\}$ = linear set of coefficients in function $\Sigma b_i X_i$, estimated, for example, as coefficients of the discriminant function;

$\{X_1\}$ = set of exposure indicators, confounding factors and/or effect modifiers (Miettinen, 1979).

From (2), it follows that:

$$p_D = \frac{1}{1 + e^{-\Sigma b_i X_i}}$$

which gives the probability of D as a weighted multiple logistic function of the k "predictors". If the Xs are binary, the estimated e^{b_i}:s give the estimates of the RRs for the X_i:s, and testing for $b_i = 0$ equals testing of $RR_i = 1$.

REFERENCES

1. Saracci, R. Epidemiological strategies and environmental factors. International journal of epidemiology, 7(2): 101-111 (1978).

2. Hernberg, S. et al. Coronary heart disease among workers exposed to carbon disulfide. British journal of industrial medicine, 27: 313-325 (1970).

3. Hannunkari, I. et al. Work conditions and health of locomotive engineers. II. Questionnaire study, mortality and disability. Scandinavian journal of work, environment and health, 4(Suppl. 3): 15-28 (1976).

4. Hernberg, S. et al. Excess mortality from coronary heart disease in viscose rayon workers exposed to carbon disulfide. Scandinavian journal of work, environment and health, 10: 93-99 (1973).

5. Hemminki, K. et al. Childhood cancer and the parental occupation in Finland. Journal of epidemiology and community health (in press).

6. Bridbord, K. et al. Estimates of the fraction of cancer in the United States related to occupational factors. Prepared by the National Cancer Institute, the National Institute of Environmental Health Services and the National Institute of Occupational Safety and Health, 1978.

7. Cole, P. & Merletti, F. Chemical agents and occupational cancer. Journal of environmental pathology and toxicology, 3(4): 399-417 (1980).

8. Cole, P. & MacMahon, B. Attributable risk in case-control studies. British journal of preventive and social medicine, 25: 242-244 (1971).

9. Enterline, P.E. & Marsh, G.M. Environment and mortality of workers from a fibrous glass plant. In: Lemen, R. & Dement, J.M., ed. Dusts and disease (occupational and environmental exposures to selected fibrous and particulate dusts). Park Forest South, Illinois, Pathotox Publ., 1979, pp. 221-231.

10. Pasternack, B.S. & Shore, R.E. Statistical methods for assessing risk following exposure to environmental carcinogens. In: Whittemore, A., ed. Environmental health. Quantitative methods. Philadelphia, SIAM, 1977.

11. Demopoulos, H.P. Summary of conference. Major discussion points and societal interrelationships. Journal of environmental pathology and toxicology, 3(4): 473-481 (1980).

12. Susser, M. Judgement and causal inference: criteria in relationships. American journal of epidemiology, 99: 385-388 (1977).

13. Härkönen, H. et al. Exposure-response relationship between styrene exposure and central nervous functions. Scandinavian journal of work, environment and health, 4(1): 53-59 (1978).

14. Nurminen, M. & Mutanen, P. Statistical modelling of epidemiologic data. Scandinavian journal of work, environment and health, 6: 146-152 (1980).

15. Rothman, K.J. Synergy and antagonism in cause-effect relationships. American journal of epidemiology, 99: 385-388 (1974).

16. Kupper, L.L. & Hogan, M.D. Interaction in epidemiologic studies. American journal of epidemiology, 108: 447-453 (1978).