Environmental Health Perspectives
Vol. 62, pp. 407-414, 1985

The Problem of Multiple Inference in Identifying Point-Source Environmental Hazards

by Duncan C. Thomas*

Point-source environmental hazards are often identified by examination of unusual clusters of disease cases. The very large number of potential clusters give rise to the statistical problem of "multiple inference," i.e., the more clusters examined, the greater the risk of "false-positive" associations emerging by chance alone. This paper first distinguishes the situation of clusters identified by anecdotal observation from those that emerge from systematic searches. The latter may or may not include a systematic enumeration of potential causal factors associated with each potential disease cluster. If exposure information is not systematically available, empirical Bayes procedures are suggested as a basis for ranking the observed clusters in order of priority for further investigation. If exposure information is systematically available, empirical Bayes procedures can be used to select associations to report or to rank them in order of priority for confirmation. In addition, procedures are described for testing the global null hypothesis of no exposure–disease associations and for estimating the number of true-positive associations. These approaches are advocated in preference to classical frequentist approaches of multiplying *p* values by the number of tests performed.

Environmentally related diseases have become a major concern to the public and headlines are made almost daily. The large number of claims requiring some kind of response from public health officers raises new challenges for which the epidemiologic community has yet to formulate any systematic or widely accepted solution. The natural conservatism of the scientific discipline, arising from the conventional requirement of a high degree of statistical significance and the low credibility scientists frequently attach to these claims, is often met with cynicism by the general public (*1*). The scientific approaches may not lead to appropriate responses to the public health problem. The situation is further compounded by the frequent failure to distinguish between the kinds of statements that can be made based on anecdotal evidence and those based on systematic study. Finally, there is the lack of consensus among statisticians about how to deal with the problem of multiple inference in large scale exploratory studies (*2*).

The statistical issues of multiple inference for point source environmental hazards are the subject of this paper. The interpretation of an isolated anecdotal observation is discussed first. Next, approaches to systematic examination of clustering behavior in the absence of exposure information are described. Finally, the problem of multiple inference in studies of many possible exposure-disease associations is addressed.

*Department of Preventive Medicine, University of Southern California School of Medicine, 2025 Zonal Avenue, Los Angeles, CA 90033.

Investigation of an Isolated Cluster

Reactive Epidemiology

On September 17, 1981, the Montreal *Gazette* carried a lead headline alleging that a particular street in the district of Chomedy, Quebec had suffered an "epidemic" of cancer, strongly suggesting an unspecified environmental cause. In the ensuing furor, Dr. Walter Spitzer of the McGill Cancer Center was commissioned to carry out an epidemiologic survey of the area to determine whether there was in fact an excess of cancer and, if sustantiated, to explore possible causal factors.

The methods used—careful definition of the population and period of time at risk, selection of suitable control populations, complete ascertainment of cases, confirmation of diagnoses, etc.—were all part of standard epidemiologic practice and have been described elsewhere (*3*). No unusual excesses were in fact observed. An important use of epidemiology is the proper documentation of whether or not a disease excess in fact exists. Properly done, such a study can relieve unnecessary anxiety or show clearly the need for study of causal factors or intervention.

Several statistical issues require consideration, however. First, how widely should one define the population

at risk and the period of time? Second, should one include the observations which led to the investigation? Finally, how are statements of statistical significance to be interpreted?

The first two questions are closely related. Commonly, as in the Chomedy scare, neither the geographic area nor the period of time are well defined in the initial reports. The epidemiologist must weigh several conflicting considerations. First, one must decide whether the purpose of the study is to determine the truth of the claimed excess or to determine whether the excess is a reflection of a more general phenomenon. If the former, then it would be pointless to exclude the initial observations, and expansion of the area or time period too widely would risk diluting any excess. On the other hand, if the purpose is to look for a general phenomenon, then the initial observations must be excluded in order to obtain an independent replication. In either case, if the suspected causal factor is localized in space, then the study population should be similarly localized. If it is localized in time, then the period of ascertainment should not predate the exposure (except perhaps for comparison purposes), and it is likely that prospective observations will be required for testing the hypothesis. If the cause is unknown, then both space and time must be defined widely enough to allow a reasonable range of comparisons. Finally, the investigator must consider sample size limitations: no excess, or lack of excess, is convinving if based on only a few cases.

The interpretation of claims for statistical significance is more subtle. It is unlikely that a study would have been carried out had the initial observations not given some grounds for concern. Hence, if these observations are included in the study data, the probability of a "statistically significant" excess, even if the null hypothesis is true, is certainly larger than the claimed significance level. This liberal bias arises out of a nonrandom selection of study population (i.e., the inclusion of a subpopulation in which it is known in advance that an excess exists) and does not arise if the initial observations are excluded. It may still be helpful to make significance statements, but their proper interpretation is as follows: "An excess of this magnitude or greater would be observed by chance alone in alpha percent of clusters selected at random; however, as the present cluster was not selected at random, the probability that such an excess would have come to attention by chance alone cannot be assessed." (It is worth noting that a similar liberal bias affects the reporting of findings in the scientific and lay press generally, thereby making it difficult to combine all the evidence in interpreting any particular association.)

Analytical Epidemiology

If a sufficiently large excess does appear to exist, the obvious next step is a search for possible causes, perhaps by correlational ("ecological") or case-control approaches. If enough factors are considered, it is probable that at least one "statistically significant" determinant will be found. Such factors deserve to be reported, but unless they are stated as hypotheses in advance, biologically plausible, and based on independent prior evidence, most scientists would agree that action should not be taken without independent replication.

Similar issues of multiple inference arise when the starting point is a particular exposure factor (rather than a particular disease cluster) and a variety of possible health outcomes are considered. These issues are addressed below in the general context of systematic explorations of associations between many diseases and many exposures.

Systematic Examination of Clustering Behavior

Most diseases will show more than random variation in incidence rates because there are real regional variation in risk factors or completeness of ascertainment. Systematic studies of clustering can help put anecdotal observations into context. Such variation can also be useful for generating hypotheses. For example, the "cancer maps" of the U.S. counties (*4*) have often been used to generate hypotheses, either through systematic correlation analyses (*5*) or through astute observations of isolated clusters associated with particular exposures.

In order to identify the "high risk" clusters for further investigation or to rank the clusters for correlational analysis, one must adopt some parameter of excess risk and some procedure for estimating it. On the one hand, it could be argued that the most important clusters to investigate are those most likely to represent causal associations, and it is sometimes stated (*6*) that the best single index of causality is the strength of the relative risk (RR). On the other hand, it could be argued that a more appropriate index for public health purposes is the number of excess cases (the "attributable number," AN). Whatever parameter is adopted, chance variation must still be taken into account. A simple ranking of clusters on their point estimates would take no account of the strength of the evidence (as summarized in a p value) whereas a ranking of p values would take no account of the magnitude of the excesses. An *ad hoc* compromise would be to rank some lower confidence limit on the chosen parameter, but the choice of confidence level would be arbitrary and different choices would produce different rankings. Empirical Bayes (EB) estimators (*7*) were developed to provide a unified approach to this problem. These refer to a system of statistical inference in which prior probabilities are estimated from the data rather than specified *a priori*, preferred estimates being those with the maximum posterior probability. In contrast, the more commonly used maximum likelihood (ML) estimates are based on a system of statistical inference in which only information in the data, rather than prior probabilities, are used, preferred estimates being those which maximize the probability (likelihood) of the observed data.

ML estimators have the paradoxical property that, though asymptotically unbiased for any particular association, they are biased when considered as an ensemble: the largest estimate, for example, is probably an overestimate of its population value and the smallest estimate an underestimate. Classical Bayes estimators therefore pull each estimate $\hat{r}_i$ back toward the center of some "prior" distribution $f(\rho_i)$ by an amount that depends on their variances s_i^2, in order to obtain a "posterior" estimate of the true population value ρ_i. Empirical Bayes procedures differ only in that the prior distribution is not postulated arbitrarily but is fitted to the data by specifying a parametric form $f(\rho|\theta)$ and estimating its parameters θ. For example, suppose one assumed that the observed numbers of cases D_i followed a Poisson distribution with parameter $= E_i\rho_i$where E_i are the expected numbers on a standardized incidence ratio (SIR) or proportional incidence ratio (PIR) basis using rates for the combined population, and ρ_i is the true relative risk to be estimated. The then natural ("conjugate") prior for ρ_i is the gamma distribution (Appendix 1), with shape parameter k and scale parameter η to be estimated. These procedures offer two distinct advantages for hypothesis generation. First, they provide an improved ranking of high-risk clusters, allowing for differences in their random variability and in the magnitude of their risks. Second, they provide an estimate of the true variability of the population rates after removing the chance variation in the sample.

Some of the variability in population rates may be attributable to known confounding factors. This variation can be removed in several ways. The simplest is to stratify on these factors and carry out a separate EB analysis in each stratum. A better approach would be to start with estimates which were standardized for these factors and combine all strata in a single EB analysis. Finally, covariate information z could be added to an EB analysis of nonstandardized estimates by allowing the parameters to depend on z. For example, again taking the prior distribution for ρ_i to be gamma, one might allow the scale parameter to depend on covariates, say $\eta = \exp(a + bz)$, and estimate the regression coefficients a and b together with the shape parameter k (assumed not to depend on z). This approach can sometimes be applied when standardized estimates cannot be obtained, e.g., when only one aggregate data on confounders are available, and is explored below as a way of incorporating information on exposure.

Details of these approaches are described elsewhere (*8*) with applications in a slightly different context, namely, identifying cancer clusters associated with particular occupations or occupational exposures. In that project, exposure information was available, so the study became a systematic exploration of associations rather than simply of disease clusters; these applications are therefore described in the next section. Another important difference was the use of individual rather than aggregate data, thereby avoiding the "ecological fallacy" (*9*), that associations across aggregate units may not reflect differences in risk between individuals at different exposure levels.

It is worth repeating that distributions of clusters can be helpful for putting particular clusters (e.g., those exposed to a hazardous waste disposal site) into context, but that such comparisons have straightforward interpretations as statistical significance claims only if the clusters being tested were selected *a priori* (i.e., by exposure), not by the knowledge that they showed excesses of disease.

Systematic Exploration of Many Possible Exposure–Disease Associations

A question of greater scientific interest than the descriptive problems discussed above is whether exposure to hazardous waste disposal sites has any adverse effect on health and, if so, which sites or which chemicals are associated with which adverse outcomes. Suppose that a series of population units have been selected, each classified in various ways by exposure (e.g., whether or not it is near any disposal site, whether near a particular site, whether the site contains a particular chemical, whether it has contaminated the water supply, etc.) and observed and expected numbers of cases of various diseases ascertained. There would appear to be three basic scientific questions to be addressed: (1) to test the global null hypothesis that there are no true exposure-disease relations; (2) if rejected, to estimate the number of true-positive associations; and (3) to select from all possible associations, those that are most likely to be true associations (or at least those most in need of further study). In addition, one might add a fourth objective of a more descriptive nature: to provide for each exposed cluster an assessment of the ranking of its disease risks relative to similar nonexposed clusters, recognizing that because of the large number of possibilities, some exposed clusters will appear to be at high risk by chance alone.

Descriptive Approaches

Before describing approaches to dealing with the scientific questions, it would be helpful to review the descriptive approaches that were developed to address the fourth objective in a study by Dr. T. Mack of cancer incidence in relation to proximity to waste disposal sites in Los Angeles County (*10*). For this purpose, census tracts were selected as the unit of observation (1290 in number) and classified (a) as "exposed" or not, depending on whether they contained a disposal site (about 50 in number), and (b) into one of 11 racial/socioeconomic (SES) strata. The observed incident cancers from 1972 to 1981 in each census tract were counted in each sex and in about 90 anatomical sites. The observed numbers were compared with expected numbers on age-adjusted population (SIR) and proportional incidence (PIR) bases,

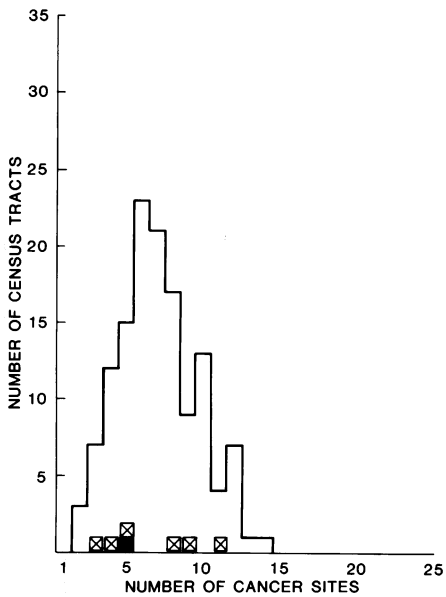

FIGURE 1. Frequency distribution of 133 sample census tracts by number of cancer sites with observed high risk. Criterion of high risk: SIR or PIR >2, cancer sites with high prior etiologic plausibility only. Solid box denotes census tract with direct hazardous risk exposure, open boxes denote surrounding census tracts.

using rates for all Los Angeles County. Anatomical sites were further classified by the biological plausibility of any associations with chemical exposures.

The first step of the analysis has been to develop various definitions of "excess incidence" based on the size of the SIR and PIR, their lower 95% confidence limits, the number of excess cases (or "attributable number," AN), and the concordance of the sexes. Frequency distributions for a random sample of each of the race/SES strata were then obtained for the number of anatomical sites in each plausibility group showing excesses by these criteria (see, for example, Fig. 1). For those anatomical sites showing excesses, the SIR was also plotted against the AN on a scatter diagram. These frequency distributions of numbers of excess sites, and scatter diagrams for particular anatomical sites could then be used for comparison with the corresponding values for specific exposed census tracts (the indicated points in Fig. 1).

For many census tracts, their placement in relation to other comparable nonexposed tracts was sufficient to show that their rates were not unusual. Some, however, do appear to show unusual excesses and are a cause of concern, requiring further investigation. However, because of the large number of potential associations (200 exposed tracts × 90 anatomical sites × 2 sexes), and because of the complexity of the criteria defining "excesses," it is difficult to evaluate the statistical significance of any particular association or to assess whether the frequency of disposal site associated excesses is any different from what would be expected by chance. For this reason, approaches to the first three objectives enumerated above are still needed.

We are, however, attempting to summarize the pattern as follows. For any particular waste disposal site, several census tracts may be considered to be exposed; if there is more than one, a subjective grading of the degree of potential exposure is assigned to each. Similar subjective weights are assigned to each anatomical site, based on a classification of biological plausibility similar to that used by Buffler (*11*), and additional weights are assigned for concordance between the sexes and concordance across exposed census tracts. The percentile rankings of RRs for those associations showing excess are then combined, incorporating the various weights, as described in Appendix 2, resulting in a single summary score for the disposal site. The absolute value of the score depends in a complex way on the number of exposed tracts, the number of anatomical sites, and the various weights, and cannot be interpreted in isolation. Because of its complexity, the sampling distribution of the score can also not be computed analytically, but it can be simulated by computer. Basically, a race/SES-stratified sample of sets of the same number of census tracts is drawn from the pool of comparison tracts, randomly assigned the same exposure values as the actual exposed tracts, and the summary score recomputed. By repeated sampling, a distribution of scores is obtained for comparison against the observed score. The end result is a percentile rank for the overall pattern of site-specific excess incidence for each disposal site (*11*).

Analytical Approaches

Regression Models. To test the global null hypothesis of no exposure-disease associations, an ecological correlation analysis (*9*) can be done. Specifically, a Poisson-error model (*12*) of the form

$$D_{ij} = E_{ij} \exp \{a_j + b_j z_i\}$$

can be fitted by using linear models packages such as GLIM (*13*). However, its validity rests on the assumption that the only sources of variation in the observed relative risks $r_{ij} = D_{ij}/E_{ij}$ are the systematic effects of measured covariates z_i and Poisson errors. As noted earlier, there is almost certainly additional variation due to unmeasured risk factors or ascertainment errors, and the failure to include this source of variation will lead to an underestimate of the error variance and hence an overestimate of the degree of significance. This can be overcome by fitting the marginal likelihood from the compound Poisson model

$$D_{ij} \sim \text{Poisson } (E_{ij}\rho_{ij})$$

and

$$\rho_{ij} \sim \text{Gamma } [k, \eta = \exp\{a_j + b_j z_i\}$$

which is given as Eq. (A-4) in Appendix 1. Here, if z_{ij}

are binary, $\hat{b}_{ij}$ is simply $\ln(D_{ij}/E_{ij})$ as for the simpler Poisson model above, but its standard error now correctly reflects the additional sources of variation. Once a_j, b_j, and k are estimated, the EB estimates of ρ_{ij} are the residual natural logarithm of residual risk (ln RR) for all census tracts, after removing the effects of the measured exposure variables, which could be used for further hypothesis generation.

For a truly global test, we could constrain $b_j = b$ and confine z to the single variable, "exposed" or not. Nevertheless, to the extent that any associations are specific to particular anatomical sites or to particular disposal sites, this global test would have very low statistical power. On the other hand, to the extent that each anatomical site is allowed a different set of coefficients and many exposure variables are included, the test ceases to be a global one and the problem of multiple inference remains.

Estimating the Number of True Positive Associations. It is therefore appropriate to estimate the number of "true-positive" associations, to aid in the interpretation and to guide the selection of associations to report. Various ways of doing this have been reviewed elsewhere (*8*). Two of the most promising techniques are p value plotting (*14*) and comparison against a randomization distribution of p values. Both approaches are applied to sets of p values, which can be derived from simple comparisons of observed and expected counts or from the more sophisticated regression models described in the previous section.

The first approach assumes that the true distribution of p values is a mixture of a theoretical uniform distribution for "true negative" associations and some other distribution for the "true positive" associations, whose form is unspecified but is assumed to be concentrated near the "significant" end. By plotting the cumulative distribution of observed p values and fitting a straight line through most of the distribution, the point where that line intersects the vertical axis provides an estimate of the number of true negative associations. In the example in Figure 2, taken from the occupational cancer example (*8*), the estimated number of true positive associations is about 19 to 24 (out of 218 based on at least five expected exposed cases). With small numbers of cases, however, the assumption of uniformity of p values under the null hypothesis can break down. Such nonuniformity was clearly evident when we considered all 684 associations, which included many with fewer than one expected exposed case.

On the other hand, exclusion of those associations based on small numbers may cause one to miss some large relative risks for rare exposures or rare diseases. This problem can be overcome by comparing the observed distribution of p values not against a theoretical uniform distribution but against an empiric distribution obtained by randomizing the exposures against the diseases (Fig. 3). For example, if each disposal site represented a separate binary exposure, the randomization distribution would be simply the observed distribution of p values for all census tract × anatomical site associations. Techniques have been described elsewhere (*8*) for comparing this distribution to that for the exposed tracts to obtain an integrated measure of separation between the two distributions and hence an estimate of the number of true positive associations. In the occupational example, the resulting estimate is about 25 of the complete set of 684 associations.

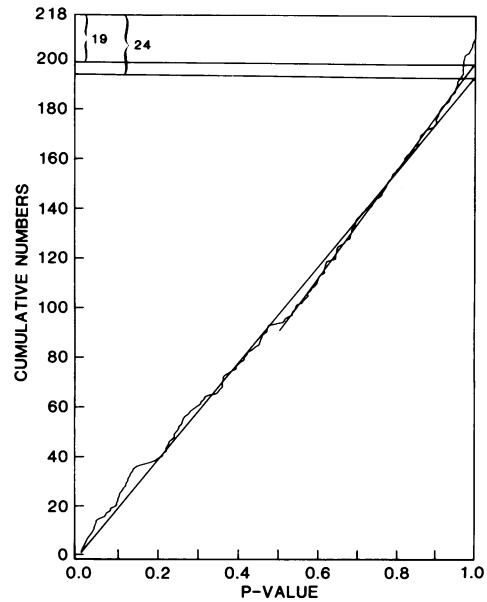

FIGURE 2. Cumulative distribution of p values for 218 occupational associations based on at least five expected exposed cases. Fitted lines are used to estimate the number of true positive associations.

Selecting Associations To Report. The procedures described above provide an estimate of the number of true positive associations but cannot identify which associations are the true ones. Nevertheless, it is clear that those with larger relative risks or smaller p values are more likely to represent the true positive associations, so it is sensible to single them out for reporting or for further investigation. Various ways of selecting a subset of associations for reporting have been reviewed elsewhere (*8*), including multiplying p values by the number of tests performed, splitting the sample for searching and testing, and Bayesian methods. None of these approaches seems suitable. Adjustment of p values produces very low power for detecting true associations; splitting the sample offers no advantage over simply adopting a more conservative significance level and has poorer power than a single analysis at that more conservative level; and Bayesian methods require a consensus about explicit prior distributions.

An approach meriting further exploration entails the use of cost-benefit criteria, not to weigh the advantages and disadvantages of reporting particular associations, but to assess the performance of alternative decision rules. It is unlikely that any consensus could be developed to weight the potential benefits of particular exposures or the potential costs of particular diseases.

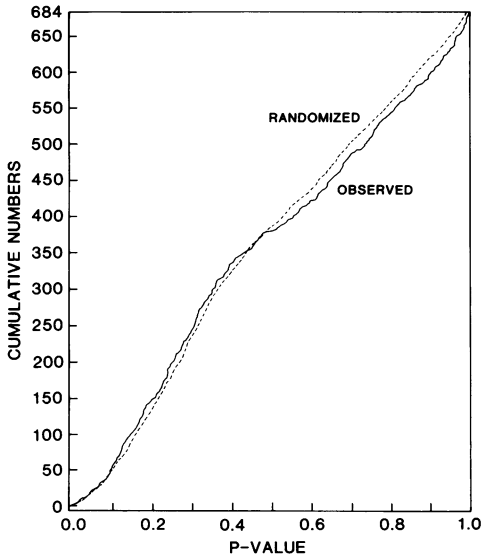

FIGURE 3. Cumulative distribution of p values for all 684 occupational associations: (—) observed associations; (--) randomized associations.

However, it is reasonable to suppose that, on average, the benefits of reporting a true positive association would be proportional to the true AN, and the cost of reporting a false positive association would be proportional to the apparent AN. This leads to a theory for evaluating alternative decision rules that is different from the classical Neyman-Pearson or Bayesian schools, and is a promising avenue of further research.

In the meantime, our preference is not to use binary decision rules to decide which associations to report, but to report all associations together with a ranking based on EB estimates of the parameter which is considered to best meet the objectives of the study, e.g., the RR or the AN. Suppose we let the parameter of interest be r_{ij} = lnRR for the association between waste disposal site i and anatomical site j, and assume r_{ij} is normally distributed with true unknown mean ρ_{ij} and known variances s_{ij}^2. The estimate $\hat{r}_{ij}$ might be obtained simply as $\ln(D_{ij}/E_{ij})$ for exposed census tracts, or as $\hat{b}_{ij}$ in the more sophisticated regression models described above. As a prior distribution, it is reasonable to assume that there is a finite probability that the true $\rho_{ij} = 0$; this leads us to postulate a three-parameter prior distribution

$$f(\rho_{ij}) = \alpha\delta(0) + (1 - \alpha)\, N(\mu,\sigma^2)$$

where $\delta(x)$ is the Dirac δ-function and α, μ, and σ^2 are parameters to be estimated. The resulting EB estimate of $\hat{b}_{ij}$ could then be expressed as RRs or ANs for ranking purposes. An example of the rankings obtained in these various ways for a selected subset of associations from the occupational cancer study is provided in Table 1. Because the data are preliminary, the exposures are identified only by code numbers.

Comment

The problem of deciding which disposal sites (or other exposures) pose a public health threat when many possible associations are under investigation is difficult. Part of the problem lies in the difference in interpetation between the disease cluster that was identified by anecdotal evidence and the one that emerged as a result of systematic investigation. The latter are generally more informative. Various approaches were outlined above for identifying clusters that might be investigated to search for possible causes and for exploring hypotheses regarding various exposures that have already been measured. Applications of these approaches to the Los Angeles County waste disposal site cancer data are described in a separate report (*10*).

The basic study design we are using is very simple and could easily be implemented by almost any cancer registry. The statistical methods are less straightforward, but computer programs can be made available. The hard part is the interpretation. Empirical Bayes techniques (even with the incorporation of prior knowledge) are no substitute for scientific judgment and no cure for faulty data. Before reporting any associations from such an analysis, one must seriously consider the biologic credibility of the association, the potential for bias in the design, the concordance with other data, and so on. Furthermore, the statistical methods advocated here, while arguably more powerful than classical frequentist techniques when many hypotheses are to be considered simultaneously, are still limited in their power by the available sample sizes. "Nonsignificant associations" do not necessarily imply no association; power calculations should always be done before dismissing a credible hypothesis. Finally, the methods still require further development. In particular, it would be useful for the kinds of environmental associations considered here to be able to incorporate into the EB analysis information on contiguity, as is proposed, for example, in the randomization procedure described in Appendix 2.

The reception of the rather convoluted statistical procedures outlined here by the general public remains to be seen. A member of an exposed cluster may be convinced that there is a causal link and not be impressed by the fact that many other hypotheses considered failed to show an effect. Indeed, the relevance of these other hypotheses to the interpretation of the particular association is not even agreed by scientists (*2*). Therefore, the first priority should be to develop a consensus of scientists on how to deal with the problem of multiple inference, and then find ways of explaining the approach convincingly to the public. The first step is to recognize that the adoption of extremely conservative decision rules (such as multiplying p values by the number of tests performed) is not an appropriate response to either the statistical or the public health problem. Hopefully

Table 1. Maximum likelihood (ML) and empirical Bayes (EB) estimates of relative risk and attributable number for selected occupational associations.

		Relative risk estimates		Attributable number estimates	
Cancer site	Exposure factor	ML	EB	ML	EB
Lung	E1	1.55	1.43	24.2	19.3
Bladder	E2	1.89	1.52	18.6	11.3
Stomach	E1	1.85	1.48	14.4	8.7
Lung	E3	3.88	1.86	24.9	8.0
Kidney	E4	1.58	1.39	10.4	7.4
Colon	E5	1.88	1.45	13.9	7.3
Prostate	E6	2.22	1.54	15.0	6.9
Stomach	E7	1.53	1.37	8.1	5.8
Rectum	E4	1.90	1.42	8.6	4.5
Colon	E8	1.97	1.50	6.6	3.5
Prostate	E9	4.71	1.96	11.7	3.1
Bladder	E10	2.01	1.42	5.6	2.4
Pancreas	E2	3.12	1.51	6.7	1.9
Kidney	E11	2.85	1.47	5.1	1.4
Kidney	E12	2.81	1.45	3.8	1.0
Rectum	E13	4.76	1.56	5.2	0.9
Lung	E14	3.61	1.41	3.1	0.5

some of the approaches outlined above will prove to be more suitable solutions.

Appendix 1

Implementation of Empirical Bayes Procedures

Let D_{ij} be the observed number of cases of disease j in census tract i, E_{ij} the expected number, and z_i a vector of exposure and confounder information on census tract i. Let ρ_{ij} represent the true (unknown) relative risk, incorporating both measured and unmeasured covariates. Then the likelihood of observing d_{ij} is the Poisson probability

$$P(D_{ij}|E_{ij}\rho_{ij}) = (E_{ij}\rho_{ij})D_{ij} \exp\{-E_{ij}\rho_{ij}\}/D_{ij}! \quad \text{(A-1)}$$

The conjugate prior for the Poisson distribution is the gamma distribution, with parameters k for shape and η for scale:

$$P(\rho|k,\eta) = \frac{(\rho/\eta)^{k-1} \exp\{-\rho/\eta)}{\eta\Gamma(k)} \quad \text{(A-2)}$$

This produces a family of skewed distributions, with mean $k\eta$ and variance $k\eta^2$. To add covariates, one can let η depend on z while keeping k constant, for example:

$$\eta_{ij} = \exp\{a_j + b_j z_i\} \quad \text{(A-3)}$$

To estimate a, b, k, the marginal likelihood $L = \Pi_{ij} L_{ij}$ can be maximized, where

$$L_{ij} = \int P(D_{ij}|E_{ij}\rho_{ij})\, P(\rho|z_{ij},a_{ij},b_{ij},k)d\rho$$
$$= \left(\frac{\Gamma(D_{ij} + k)}{\Gamma(k)\Gamma(D + 1)}\right)\left(\frac{(\eta_{ij}E_{ij})D_{ij}}{(1 + \eta_{ij}E_{ij})D_{ij+k}}\right) \quad \text{(A-4)}$$

Once these estimates have been obtained, the posterior distribution of ρ_{ij} is ρ

$$P(\rho_{ij}|a_{ij},b_{ij},k,D_{ij},E_{ij},z_{ij})$$
$$= P(D_{ij}|E_{ij},\rho_{ij})\, P(\rho_{ij}|z_{ij},a_j,b_{ij},k)/L_{ij}$$
$$= \text{Gamma}\left[\frac{D_{ij}+k,\ \eta_{ij}}{(1+\eta_{ij}E_{ij})}\right] \quad \text{(A-5)}$$

The posterior expectation of ρ_{ij} is therefore

$$\hat{\eta}_{ij} = \frac{D_{ij} + k}{E_{ij} + (1/\eta_{ij})} \quad \text{(A-6)}$$

which is the relative risk parameter, adjusted for z, that is suggested for ranking.

Where z includes exposure variables of interest, the MLEs of b can in turn be considered to be a family of random variables and subjected to EB analysis. For this purpose, we propose assuming $\hat{b}_{ij}$ are normally distributed with known variances s_{ij}^2 and mean β_{ij} to be estimated. (If the z_i are binary, the $\hat{b}_{ij}$ are simply $\ln(D_{ij}/E_{ij})$ for the subset of census tracts with $z_i = 1$). For a prior distribution, we postulate that there is a non-zero probability that $\beta_{ij} = 0$, so we take

$$P(\beta_{ij}) = \alpha\delta(0) + (1 - \alpha)\, N(\mu,\sigma^2).$$

EB estimation of β_{ij} is described by Thomas et al. (Appendix 2) (*8*).

Appendix 2

Proposed System for Scoring Excess Rates Associated with Particular Waste Disposal Sites

Let s represent a particular disposal site and let $i = 1, \ldots, I_s$ indicate the census tracts exposed to it. For

each exposed census tract, let z_i be a grading of its degree of exposure, and let l_i indicate its race/SES stratum.

For each anatomical site $j = 1, \ldots, J$, let w_j be a subjective weight to be assigned to the prior credibility of associations with enivronmental exposures. For each sex $k = 1,2$, let S_{ik} denote the set of all anatomical sites that show "excess" incidence in census tract i by some criterion. Our primary criterion is:

$$\text{Either PIR} > 1.5 \quad \text{or} \quad \text{SIR} > 1.5$$

and

$$\text{Corresponding Poisson probability} < 0.025$$

Further, let S_{i*} represent the subset of anatomical sites for which both sexes show excess incidence, and let C_{jk} represent the set of exposed census tracts which show excess incidence at anatomical site j and sex k.

For those anatomical site × sex combinations showing excess incidence, let p_{ijk} be the percentile rank of RR_{ijk} among all census tracts in the same stratum, scaled so that small p values indicate large RRs.
Our proposed summary score is then

$$\begin{aligned}\hat{S}_s = &\sum_i \sum_k \sum_{j \in S_{ik}} z_i w_j \ln p_{ijk} \\ &+ \alpha \sum_i \sum_{j \in S_{i*}} z_i w_j \ln (p_{ij_1} p_{ij_2}) \\ &+ \beta \sum_j \sum_k \sum \sum_{(i_1, i_2) \in C_{jk}} z_{i_1} z_{i_2} w_j \ln (p_{i_1jk} p_{i_2jk})\end{aligned}$$

where α and β are coefficients indicating the weight to be assigned for concordance between the sexes and concordance across exposed tracts.

The distribution of S_s can be simulated by drawing repeated stratified samples r of size I_s from the population of all census tracts, maintaining the same distribution across race/SES strata. Each sampled census tract would then be randomly assigned an "exposure score" z_i from the scores for the actual exposed census tracts in the same stratum. For each sample r, the statistic S_{ir} would be evaluated and the observed S_s would be compared against the distribution of S_{sr}.

This work was supported in part by Public Health Service grants CA 17054 and CA 14089, National Cancer Institute, and contract number 82-79920, State of California Department of Health Services.

I wish to thank Dr. Thomas Mack for many helpful suggestions, Mr. Richard Pinder for technical assistance, and Mrs. Beth Woodin for preparation of the manuscript.

REFERENCES

1. Neutra, R. Epidemiology for a distrustful community. Environ. Health Perspect. 62: 393–397 (1985).
2. Cole, P. The evolving case-control study. J. Chron. Dis. 32: 15–17 (1979).
3. Spitzer, W. D., Shenker, S. C., and Hill, G. B. Cancer in a Montreal suburb: the investigation of a nonepidemic. Can. Med. Assoc. J. 127: 971–974 (1982).
4. Mason, T. J., McKay, F. W., Hoover, R., Blot, W. J., and Fraumeni, J. F., Jr. Atlas of Cancer Mortality for US Counties, 10-1950-1969. US DHEW, Washington, DC, 1975.
5. Blot, W. J., and Fraumeni, J. F. Geographic patterns of lung cancer: industrial correlations. Am. J. Epidemiol. 103: 539–550 (1976).
6. Breslow, N. E., and Day, N. E. Statistical Methods in Cancer Research: I. The Analysis of Case-Control Studies. (IARC Scientific Publications, No. 32) International Agency for Research on Cancer, Lyon, 1980, pp. 58–69.
7. Morris, C. Parametric empirical Bayes inference: theory and applications (with discussion). J. Am. Statist. Assoc. 78: 47–65 (1983).
8. Thomas, D. C., Siemictycki, J., Dewar, R., Robins, J., Goldberg, M., and Armstrong, B. G. The problem of multiple inference in studies designed to generate hypotheses. Am. J. Epidemiol., in press.
9. Morgenstern, H. Uses of ecological analysis in epidemiologic research. Am. J. Publ. Health 72: 1335–1344 (1982).
10. Mack, T. M., and Thomas, D. C. Methodology for evaluating cancer risk in small communities: evaluation of cancer risk in the residential neighborhood near BKK landfill. Report to the State of California Department of Health Services, University of Southern California, Department of Preventive Medicine, Los Angeles, June 3, 1985.
11. Buffler, P. A. Potential health effects that might be detected at waste disposal sites: analysis based on population studies. Environ. Health Perspect. 62: 423–456 (1985).
12. Frome, E. L. The analysis of rates using Poisson regression models. Biometrics 39: 665–674 (1983).
13. Nelder, J. A. General linear interactive modelling. Numerical Algorithms Group, Royal Statistical Society, Oxford, 1976.
14. Schweder, T., and Spjotvoll, E. Plots of p-values to evaluate many tests simultaneously. Biometrika 69: 493–502 (1982).