Clues to the Etiology of Disease Clusters: A Practical Method

Daniel Smith and Sumi Hoshiko
California Department of Public Health, Environmental Health Investigations Branch
Richmond, CA 94804
Daniel.Smith@cdph.ca.gov

Annual Meeting of the Society for Epidemiologic Research, Anaheim, CA, June 2009

Investigating a community disease outbreak or cluster, such as a cancer excess, where environmental factors are suspected can be time-consuming if not focused. We describe simple approaches to thinking about risk factors and their prevalences that can guide the investigator looking for plausible explanations for an observed excess.

Typically the first step in a cluster investigation is a statistical assessment of the presence and magnitude of the excess. This involves calculating the observed number of cases in the index community and the expected number, based on rates in a reference area. The ratio of observed to expected cases, the standardized incidence ratio (SIR), reflects a combination of the relative risk (RR) of a risk factor and its prevalence. Therefore, given the SIR and possible RRs for suspected causes, one can calculate the proportion of the index community that would have to have been exposed.

The usual next step in an investigation examines the series of cases. We can screen for candidate risk factors by calculating the proportion of cases required to have a risk factor if that factor were responsible for the community SIR.

These methods can be used to help judge whether certain combinations of risk factors and exposure frequencies are plausible, and to rule in or out potential risk factors in a case series.

First steps:
1.) Compile O = the *Observed* number of cases in an index community, workplace, etc.
2.) Calculate E = the *Expected* number of cases, using rates from a reference area
3.) Examine O/E = Ratio of observed to expected = Standardized Incidence Ratio (SIR)

The community is concerned about possible environmental factors. What are plausible candidates for the exposure? What characteristics must the exposure have?

1.) Prevalence of exposure in the index community

From the SIR, we can make inferences about:

RR = the relative risk of a putative exposure
P_0 = the proportion of the reference population exposed
P_1 = the proportion of the index community exposed

The incidence in the index community (O) is proportional to the RR of exposure and the proportion of the population exposed (P_1) and unexposed ($1 – P_1$):
$O \propto P_1(RR) + (1 - P_1)(1)$

Likewise, the incidence in the reference community (E) is proportional to the RR and the proportion of the reference population exposed (P_0) and unexposed ($1 – P_0$):
$E \propto P_0(RR) + (1 - P_0)(1)$

The SIR (O/E) is the ratio of these two incidences: $[P_1(RR) + (1 - P_1)]/[P_0(RR) + (1 - P_0)]$

It follows that $P_1 = (SIR(P_0(RR - 1) + 1) - 1)/(RR - 1)$

If the exposure is unique, and no one in the reference population is exposed (that is, $P_0 = 0$), then $P_1 = (SIR – 1) / (RR – 1)$.

The least the RR of the exposure could be is the SIR, and this occurs when everyone in the community is exposed ($P_1 = 1$) to a unique exposure ($P_0 = 0$).

Example: 10 cases are observed, and 5 are expected, so SIR = 2. If a suspected risk factor has a RR of 2, it would have to be nonexistent in the reference area ($P_0 = 0$) and exposing all of the index community ($P_1 = 1$) to produce an SIR of 2.

Another suspected risk factor X is thought from the literature to have an RR = 3 and is normally present in about 10% of the population ($P_0 = 0.10$). If X were responsible for the doubling of the community rate, it would have to be exposing 70% of the index community. Is this plausible?

Various combinations of P_0, P_1, and RR can be explored to rule in or out plausible causes in the community.

2.) Prevalence of exposure among cases

A further step in the investigation is to conduct a case series review. If we examine cases in the index community, how will recognize the causal factor if we see it? If a single risk factor were responsible for the excess, in what proportion of cases would we find that exposure? We can use the SIR and estimates of P_1 above to calculate the proportion of cases that should be exposed.

Proportion of exposed cases = Exposed cases/total cases $P_c = P_1RR /(P_1RR + (1-P_1))$

Continuing the example above with SIR = 2: If a suspect risk factor X has RR = 3, and $P_0 = 0.1$, we found previously that P_1 would have to be 70%. If we now were to examine the 10 observed cases, how many would have to share an exposure to risk factor X?

$P_c = 0.7(3) / (0.7(3) + 0.3) = 0.88$.

So a risk factor present in only half the cases would not be sufficient to account for the cluster.

We can focus our investigation on risk factors that show the required prevalence among cases

Example: We recently examined a cluster of pancreatic cancer cases in an area around a Northern California city (population about 50,000)[1].

Observed: Between 2004-2006, 33 cases of pancreatic cancer occurred

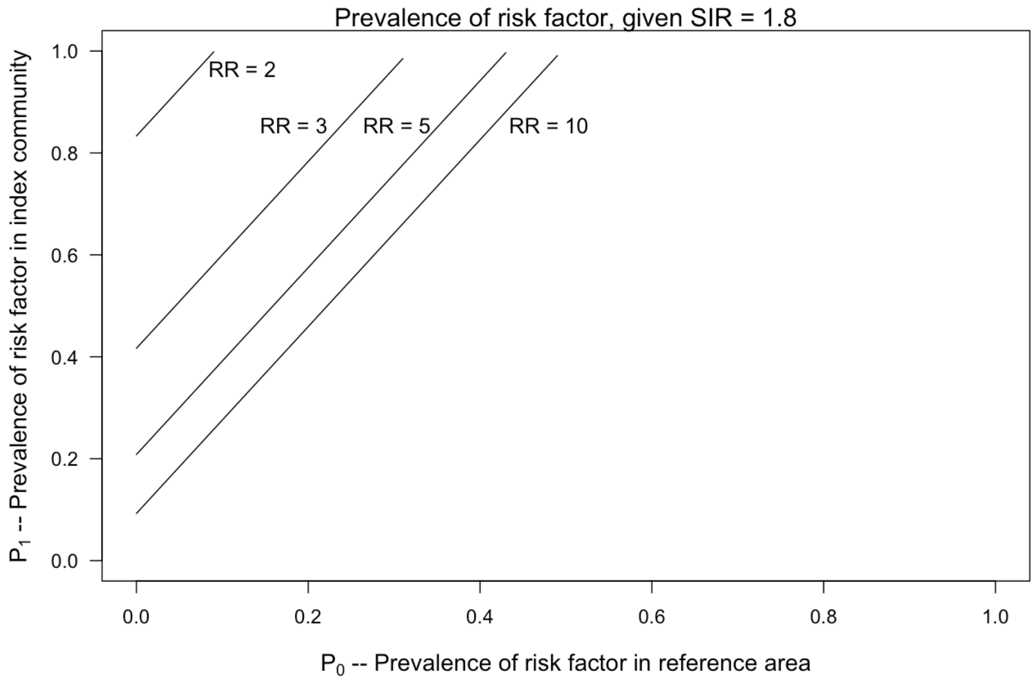
Expected: During this period, 18 cases (6 per year) would be expected

SIR = 33/18 = 1.8 ($p = 0.002$)

Q1. Prevalence of exposure in index community

The minimum RR = SIR = 1.8, so we must look for risk factors for pancreatic cancer of that strength or more.

For various combinations of P_0 and RR, we can make inferences about what P_1 must be.



For example, heavy smoking has an RR for pancreatic cancer[2] of around 5. If the prevalence of smoking in the reference area is 20%, the prevalence in the index community must be 58% to account for the observed SIR. Is that plausible?

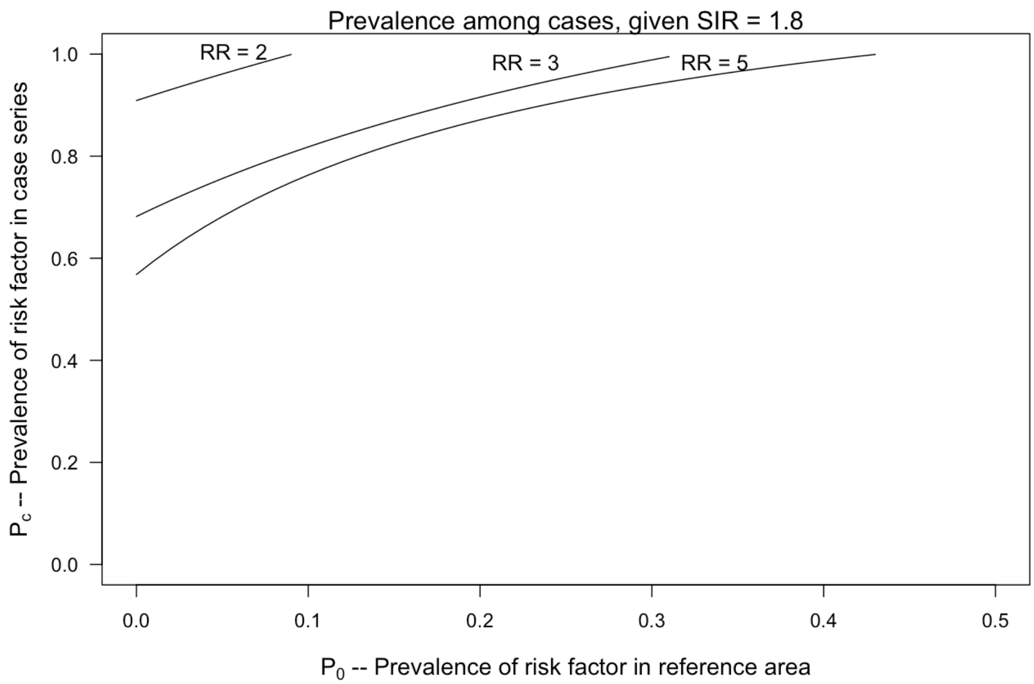
A meta-analysis of occupational exposures[3] has shown that some classes of chemicals have RRs around 1.5 to 2. Nearly all of the index community would have to be exposed to these chemicals ($P_1 > 80\%$) to account for the observed SIR.

We conclude: *If an environmental risk factor is responsible for the cancer excess, it must be among the higher RRs for pancreatic cancer, and be common among members of the index community.*

Q2. Prevalence of exposure among cases

If we examine the series of 33 cases, what should we be looking for?

If a single risk factor with an RR in the range of those reported in the literature (2-5) is responsible, it must be shared by the majority of cases (60% or more). A candidate risk factor present in less than half the cases would not be sufficient to account for the cluster.



The minimum P_c occurs when P_0 is zero and the RR is enormous. Then $P_c = (SIR - 1)/SIR$. So if SIR = 1.8, the culprit risk factor must present in at least 44% of the cases, and then only if its RR is practically infinite.

Even without a control series, some risk factors may be ruled out or in.

Results: Interviews were successfully completed with 25 out of 33 cases or next of kin. The most prevalent risk factors were family history of cancer (80%), history of smoking (60%), and history of Type II diabetes (40%). In these respects, as well as others, the cluster case series resembled the cases found in published case-control studies.

Assuming the interviewed cases were representative of all cases, we established the criteria that an environmental risk factor would have to be present in at least half of the cases for very unusual exposures, or two-thirds of the cases for more common ones, in order to have plausibly accounted for the observed excess and be a candidate for further study. No factors met these conditions.

Conclusions: Similar graphs can be drawn for any given SIR. The SIR of a cluster implies certain combinations of RR, P_0, P_1, and P_c. Examining these graphs and relationships can be used to judge the plausibility of various hypotheses, and focus an investigation down to the most likely risk factors.

References:
1. Smith D, Hoshiko S, Naik K, et al. Pancreatic Cancer Follow-up Investigation Report, Oroville Area, Butte County. California Department of Public Health, Environmental Health Investigations Branch, 2009. http://www.ehib.org/projects/OrovillePancreaticCancerInvestigation.pdf
2. Anderson KE, Mack TM, Silverman DT. Cancer of the pancreas. In: *Cancer Epidemiology and Prevention,* 3rd Ed. Schottenfeld D, Fraumeni JF, eds. New York: Oxford University Press, 2006.
3. Ojajarvi IA, Partenen TJ, Ahlbom A, et al. Occupational exposures and pancreatic cancer: a meta-analysis. *Occup Environ Med* 2000; 57:318.