National Collaborating Centre for Environmental Health

Centre de collaboration nationale en santé environnementale

Backgrounder for Evaluating the Quality of Epidemiological Studies on Environmental Health

Epidemiology is the study of causes, distribution, and control of disease in human populations. Epidemiological methods can provide the most compelling evidence for environmental risks to humans. When evaluating the quality of an epidemiological study in environmental health, some key questions to consider include:

(1) What type of study design is used and what are the limitations?
(2) Was the exposure assessment appropriate?
(3) What are the sources of bias and error in exposure and outcome assessment?
(4) Were associations statistically significant?
(5) Is there evidence of causality?
(6) Is the study original and relevant?

Examples from hypothetical environmental health studies involving exposure to pesticides are used to illustrate some of the epidemiological concepts.

(1a) Experimental Study Design	Example
Experimental study: The investigator manipulates the study factor and randomly assigns subjects to the exposure (or treatment) groups and to the non-exposed groups.	Over a season, farms were chosen at random to have their berry crop sprayed either with malathion or an inert control substance. Researchers were blinded to the exposure group when assessing the incidence and severity of skin rashes in farm workers.
Intervention trial: An intervention (typically oriented towards education and behavioural change) is applied to the study group and not to the unexposed controls. Random assignment is preferred.	Three out of six similar communities (all having similar rates of household pesticide use at the outset) were assigned to an intensive pesticide awareness campaign and three were not. The use of alternative non-pesticide pest-control methods were assessed after one year.

(1b) Observational Study Design	Example
Cohort study: A group of persons without the disease are selected to represent various degrees of exposure (e.g. from highly exposed, to unexposed) and are followed over a lengthy period of time to observe if they develop the outcome of interest. This can be done retrospectively (back in time) or prospectively. Multiple health outcomes can be measured.	Pregnant mothers were enrolled in a study on household pesticides. Their children were followed in households where it was determined if they were using synthetic or natural indoor insecticides or no pesticides at the time of pregnancy. The principal outcome was the incidence of asthma in their children.
Case-control: The study group includes people with the outcome or disease status of interest (cases) and people without (controls). Past exposures of cases are compared with those of controls to evaluate which exposures are associated with the outcome. This design is suitable for rare outcomes, such as cancer.	Emergency room records were used to identify children treated for severe asthma attacks. The comparison group was children treated for injuries during the same period at the same hospital emergency department. The families were then interviewed about household pesticide use, among other exposures.
Cross-sectional: Exposures and outcomes are determined at the same point in time, often by questionnaire. These studies are relatively quick and less expensive.	A cross-sectional survey of children attending an asthma camp included questions for parents about their children's asthma severity, as well as on household pesticide use, environmental tobacco smoke and other exposures.
(1c) Descriptive Study Design	**Example**
Case report (or case series) presents a detailed description of new or unique findings on a disease in one (or more) individuals. This generates hypotheses on disease causation.	The case of 5 children with fatal asthma was described and it was noted that all of their mothers were employed as berry pickers.
Ecological or correlation studies: Examine rates of disease by population-level exposures. Groups (rather than individuals within groups) are the unit of analysis. Ecological fallacy is the inability to apply group-level relationships to individuals.	Residents of urban areas have asthma rates that are increasing over time and sales data on pesticides have also increased over time. However, we cannot assume that asthmatics are more likely to have used pesticides.
Surveillance studies: Focussed, systematic, and routine collection of health data, allowing for monitoring of trends in morbidity and mortality rates as well as on related risk factors.	From a provincial health surveillance program, asthma incidence is routinely collected and analyzed. Asthma incidence was found to have increased each year since 1990. This coincides with the introduction and use of household pyrethroid pesticides, but further

	studies are needed to confirm the link.
Disease clusters: Occurrence of a greater than expected number of cases of a particular disease within a group of people, a geographic area, or a period of time.	A number of employees of a pesticide manufacturer and some members of their families were diagnosed with leukemia. The numbers of cases were higher than expected in the general population.
(2) Exposure Ascertainment	**Example**
Exposure definition: What was the primary exposure of interest? Was the exposure accurately measured? Was it measured on a continuous scale, or on an ordinal or binary scale? If categorical, on what basis were they distinguished?	The level of exposure to the pesticide diazinon was determined as low-medium-high based on the distribution of acetylcholinesterase blood level across the three categories.
Route of exposure: Was the route of exposure (inhalation, ingestion or skin contact) defined?	Homeowners applying herbicides often do not wear adequate protective equipment and can be exposed through skin contact, as well as inhalation.
Exposure duration: Was environmental exposure based on current exposure only? Were all sources of exposure and changes over time considered?	Cumulative exposure to the organophosphate took into account current restricted use as well as past permitted uses.
Exposure measurements: Was exposure determined for individuals or was it by groups (ecological)? How was personal exposure measured? If biological markers were used, were they appropriate for short- or long-term exposures?	Certain pesticides change the level of a specific enzyme in blood. The magnitude of the change indicates dose absorbed for individual exposures. The health of workers in farms that have had aerial spraying was compared to those in farms with backpack spraying.
Exposure validation: Were subjective or objective measurements used and were they validated (i.e. do they reflect what they were intended to measure)?	Exposure history by questionnaire was compared to a biological marker of cumulative exposure to validate the questionnaire
(3a) Random Source of Error	**Example**
Random error: Reflects fluctuation around a true value because of poor precision, sampling error, and variability of measurement. Non-differential misclassification (e.g., no difference in misclassification of the exposure between those with and without the outcome of interest) will weaken associations, such that the observed effect estimate will be closer to null	Exposure to synthetic household pesticides was determined by estimating cumulative exposure (number of years by frequency of use by type of pesticide) based on subject interviews. This imprecise estimate of exposure could result in misclassification of exposure for all individuals in the study and bias the results toward the null.

(no effect) than the actual relationship.	

(3b) Systematic Sources of Error (Bias)	Example
Confounding: The association between an exposure and outcome is affected by the relationship of a confounder variable with both the exposure and the outcome. A variable cannot be a confounder unless it is distributed differently among the compared populations and it must not be in the causal pathway. "Unknown confounders may have affected the result" is a criticism often applied to any study.	A significant association was found between pesticides exposure in children of farm labourers and their performance in intelligence tests. However, socioeconomic status, was not considered and is related both to the probability of being exposed and to the outcome of IQ deficits in children.
Selection bias: An error due to systematic differences in characteristics between those selected or not selected for a study. Examples include *self-selection bias* (of study participants), the *healthy worker effect* (those employed are healthier than the general population) *and differential loss to follow-up (*loss of study participants that is related to both the exposure and the outcome).	Asthma rates among pesticide applicators were lower than that of the general population. Asthmatic workers would tend to avoid such occupations and the comparison with the general population includes those with asthma (healthy worker effect).
Information bias: Results from a systematic difference in the way the exposure or outcome is measured between compared groups. • *Recall bias* is a differential level of accuracy or completeness of recall to memory of past experiences between study groups. For example, cases may be more likely than controls to recall and report prior exposures. • *Interviewer bias* is a systematic difference in soliciting, recording or interpreting information. • *Differential misclassification:* errors in exposure or outcome classification that are more or less likely to occur in the different comparison groups.	Interviewers, aware of the health status of the children participating in the case-control study, surveyed the parents about past exposure to household pesticides. A strong association between use of cosmetic herbicides and nervous disorders was found. Parents with ill children are more likely to recall suspected exposures, particularly if the problem has received media attention. Interviewers may probe the known cases more thoroughly for parents with ill children.
(4) Statistical significance	**Example**
P-value: The probability of rejecting the null hypothesis of no difference or relationship, and therefore accepting the result, when no difference or relationship actually exists. A significance level of 5% is typically set as the accepted probability (Type I error).	The relationship between a history of insecticide exposure in the homes of children with a diagnosis of leukemia was found to be statistically significant ($P<0.05$). There is less than a 5% probability that the there no relationship between the exposure and leukemia.

Confidence intervals (CI): An estimated range of values that is likely to include the true estimate, demonstrating uncertainty. A 95% CI means that there is 95% probability that the true estimate is within the stated CI. Narrow CIs imply greater precision and are affected by sample size and variability.	For children prenatally exposed to insecticides the odds ratio for leukemia was 1.8 (95% CI 1.4, 2.0). (This means that there is 95% chance that the true estimate is between 1.4 and 2.0.) However, early childhood exposure to insecticides was not associated with leukemia (OR 1.1, 95% CI 0.8, 1.3).
Power: The probability a true association will be found when one exists is highly affected by sample size. Type 2 or β (beta) error occurs if a result is non-significant (e.g. $p>0.05$) when there is actually a true difference between groups. Power is calculated as (1- β) and is typically set at 80%.	20 men and women were compared for digital tremor after using synthetic household pesticides and no differences were observed. However, it was later determined that the study was underpowered and to potentially show significant results a sample size of at least 90 subjects was needed per group.
(5) Selection of Hill's Guidelines for Assessing Causation	**Example**
Temporal Relationship: Exposure always precedes the outcome. This is the only absolutely essential criterion.	Some residents near to fields sprayed by paraquat, between 1974 and 1999, were later diagnosed with Parkinson's disease in the study period from 2005 to 2010.
Strength: This is defined by the size of the association as determined by the effect measurement. The stronger the association, such as a large odds ratio, greater than "2", the more likely it is to be causal, if statistically significant.	For the association between the herbicide paraquat and Parkinson's disease, the odds ratio was 2.0 (95% CI1.6, 2.9), indicating that those living close to fields sprayed with paraquat had twice the risk of having Parkinson's disease.
Consistency: The association is consistent when results are replicated in studies in different settings using different methods.	The increased risk for Parkinson's disease, related to use of paraquat and other pesticides, was demonstrated in 3 case-control studies and a cohort study undertaken in the prairie regions of Canada.
Consideration of Alternate Explanations: Multiple explanations or hypotheses should be considered before making conclusions about the causal relationship between any two items under investigation.	The association between Parkinson's disease and paraquat may be explained by socioeconomic factors; residences close to crop areas are generally of lower socioeconomic status.
Plausibility: The association agrees with currently accepted understanding of pathological processes. However, new research findings must be considered.	Paraquat was found to interfere with dopaminergic systems, known to be compromised in Parkinson's disease.

Coherence: The association should be compatible with existing theory and knowledge. However, research that disagrees with established theory and knowledge are not automatically false.	The association of orchard work and pesticides with Parkinson's disease has been observed in many epidemiological studies, but it has not been possible to identify the specific pesticide(s) involved.
(6) Originality	**Example**
Does the new research add to the literature? e.g. increase in size or length of follow-up, difference in studied population such as in age, sex or ethnicity, more rigorous methodology Is the study generalizable to other populations or settings?	A prospective cohort study in Europe on neuro-degenerative disease in women linked occupational pesticide exposure to lower cognitive performance, adding to previous case-control study findings of occupationally exposed men.

Acknowledgements

Helen Ward wrote the document. Emily Peterson and Angela Eykelbosh provided valuable input and review. A previous version was reviewed by Mona Shum and Tom Kosatsky.

References

- Aschengrau A, Seage III GR. Essentials of epidemiology in public health. Sudbury, MA: Jones & Bartlett Learning; 2003.
- Elwood M. Critical appraisal of epidemiological studies and clinical trials. 3rd ed. Oxford, UK: Oxford University Press; 2007.
- Friis RH, Sellers T. Epidemiology for public health practice. 4th ed. Sudbury, MA: Jones & Bartlett Learning; 2009.
- Greenhalgh T. Assessing the methodological quality of published papers. BMJ 1997;315:305-308.
- Rothman KJ. Epidemiology, an Introduction. Oxford, UK: Oxford University Press; 2002.

A more detailed assessment tool and dictionary for epidemiological studies is available from:

- National Collaborating Centre for Methods and Tools (2008). *Quality Assessment Tool for Quantitative Studies*. Hamilton, ON: McMaster University. http://www.ephpp.ca/PDF/Quality%20Assessment%20Tool_2010_2.pdf and http://www.ephpp.ca/PDF/QADictionary_dec2009.pdf

Production of this document was made possible through a financial contribution from the Public Health Agency of Canada.