8. Data Analysis and Reporting of Aggregated Data

8. Data Analysis and Reporting of Aggregated Data ■

8.1 Introduction

Data analysis is an important component of surveillance. Data analysis is required to decipher the message that the surveillance data are trying to provide. As such, it will assist in identifying the most problematic pesticides, the risk factors associated with acute pesticide-related illness, and emerging data trends. This chapter will briefly discuss analysis of pesticide poisoning surveillance data. A more detailed review appears in standard epidemiological and surveillance textbooks [Teutsch and Churchill 2000; Rothman and Greenland 1998].

Data dissemination is another important function of PPSPs. The control and prevention of acute pesticide-related illness depends on the dissemination of surveillance data to ensure that educational, consultative, and regulatory interventions can be effectively targeted. In addition, surveillance partners (that is, HCPs, other government agencies, and labor and industry groups) welcome reports on surveillance findings and their impact. Keeping partners informed can promote visibility and support for PPSP. Additional details on the content and audience for PPSP surveillance reports are provided in this chapter.

8.2 Routine Descriptive Analysis

The pesticide poisoning data compiled by PPSPs can be analyzed in terms of person, place, and time. These analyses are useful for identifying variations that may be amenable to intervention.

8.2.1 Person-Based Analyses

8.2.1.1 Case Series

The case series focuses on persons and is the most basic level of data analysis and presentation. It serves as a useful tool for raising attention about emerging pesticide problems if they involve a particular pesticide or a particular type of pesticide usage (e.g., mosquito control or pesticide use in schools). A case series provides a narrative of the cases, including a description of the circumstances that led to exposures, and recommendations for preventing similar events. It also presents a distribution of case characteristics that can include disease classification, severity category, occupational versus nonoccupational, occupation, age, sex, and other characteristics that highlight the subgroups accounting for the greatest (and perhaps least) number or percent of cases. Reports of new and novel emerging problems that are reported in a timely manner should be considered for publication in the MMWR or another peer-reviewed publication. Examples of case series reports that were published in the MMWR appear in Appendix A.

8.2.1.2 Bivariate Analyses

Bivariate analyses involve examining the association between two variables. For example, an analysis of pesticide active ingredients and severity categories can assist in identifying the most problematic pesticides. Bivariate analyses can provide clues about risk factors that will allow the PPSP to target particular populations for outreach, education, or further study. Additional bivariate analyses can examine pesticide chemical class or functional class by risk factors such as demographic characteristics, occupation, industry, type of exposure (e.g., drift, spray, or contact with pesticide residue), or activity of the person at the time of exposure. The clustering of cases in particular worksites or across an industry can be used to determine whether a worksite or industrial group meets program criteria for targeted on-site investigations.

8.2.1.3 Rates

Ideally, risk factor data should be expressed as rates. The general form of a rate is:

$$\text{Rate} = \frac{\text{number of cases in specified time}}{\text{average or mid-interval population}} \times 10^{n}$$

where the numerator represents the number of cases meeting the specified criteria (e.g., males or farmworkers with acute pesticide-related illness), the denominator represents the size of the population during the specified time period (e.g., 1 year), and the size of n generally ranges from 2 to 6 depending on the size of the numerator (that is, n determines if the rate will be as low as *per 100 population* or as high as *per million population*).

The pesticide poisoning data collected by PPSPs is placed in the numerator. The denominator data must usually be sought elsewhere. When calculating county-wide or State-wide rates, an important source of denominator data for population estimates is the U.S. Census. When calculating rates among working populations, the denominator data can be obtained from the CPS, which is conducted by the Bureau of Labor Statistics (BLS) and is a useful source of employment population estimates.

Note that the denominator data involve limitations. Because the numerator and denominator data are often obtained from different sources, it is important to ensure that similar criteria are used for deriving the numerator and denominator. That is, the denominator should represent that portion of the population at risk for the exposure of interest. For example, when calculating rates of occupational pesticide poisoning, the denominator should consist of the number of employed workers, and not the total workforce or the total population, because the total workforce includes the unemployed who are not at risk of occupational pesticide poisoning. Similarly, when the denominator is the State population, the numerator should contain only residents of that State, because nonresidents who were poisoned in the State will not be included in the denominator. It is also important that the numerator and denominator are from the same time period (e.g., from the same calendar year).

Many factors influence the reliability of denominator (population) estimates, especially among agricultural workers. For example, migratory patterns and difficulties with tracking undocumented workers make it difficult to reliably count the total number of farmworkers. Another way of approaching the issue of denominators is to utilize the amount of pesticide used or sold in the geographic area in question. This is feasible only in States that routinely collect this type of data (e.g., California and New York). Even in these States, pesticide use information is very limited, as data collection is often confined to restricted-use products or products applied by licensed applicators.

Despite the limitations, the calculation of rates is very important. In contrast to simple counts, a rate allows a more informed comparison across groups (e.g., occupations, industries, age categories, sex, race, counties, States, etc.). Simple counts may erroneously suggest that a problem exists within a demographic group that accounts for a high proportion of the total population. By calculating rates, it may be found that the number poisoned per 100,000 persons does not vary across demographic groups. Examples of the use of rates appear in some SENSOR-pesticides publications [Calvert et al. 2003, 2004].

8.2.1.4 Time-Based Analyses

Data can also be analyzed to determine whether trends occur over time. Data comparing the number of cases reported in different years can indicate whether there were sudden shifts in the pattern of occurrence versus a stable number of reports. Although such data can be presented in tables, use of graphs can greatly facilitate the identification of trends and aberrations. Graphs can also provide visual clues on the impact of changes in case definition, program outreach efforts, introduction of a new pesticide product, or regulations restricting the use of a product. If the size of the population changes over time, it is more appropriate to examine time trends by graphing rates and not counts.

When constructing graphs, it is preferable to keep them simple. A simple, straightforward graph is much easier to interpret than a complex, cluttered graph. The graph should be able to stand alone. This requires having the title, axes, legend, and data source clearly and concisely labeled. Any data exclusions should be noted.

8.2.1.5 Place-Based Analyses: Mapping of Data

Mapping of data to show the geographic distribution of cases can be a useful tool for presenting information and for examining the spatial relationship between cases and sources of exposure. Many States with a PPSP provide a map indicating the number of cases by county. States currently involved in surveillance can geocode data and thereby produce detailed mapping using geographic information system (GIS) software. Geocoding involves including some way of recording the location of exposure that will permit geographic identification. This is not always feasible since descriptions of locations are not always clear (e.g., when workers do not know the address where they were working, pinpointing the exposure location is very difficult). The current version of SPIDER does not capture location data in a format needed for detailed geocoding (that is, location accuracy in SPIDER does not extend beyond address or Zip Code). It is anticipated that a future release of the SPIDER program will include fields (e.g., longitude and latitude) to record more precise exposure locations.

GIS mapping enables illness and injury data to be overlaid with information about crop distribution, location of waterways, roads, schools, etc. If the State has a system that records pesticide application data, this could also be mapped. Mapping can be useful to examine issues of large-scale vector control programs by mapping pest habitat areas, treatment areas, distribution of the disease of concern in animals or humans, and complaints and illness associated with pesticide use. Mapping can also be useful for identifying agricultural fields that abut residential, commercial, or school property, as these areas may be at risk for off-target pesticide drift.

This approach can be useful for generating hypotheses about risk factors that influence exposure, which can be further explored through epidemiologic studies. Data mapping may provide additional clues about geographic clustering of particular uses of pesticides or the prevalence of problems on the urban-rural interface. This information may be useful in developing interventions or providing information to policy makers responsible for land use or pesticide use regulations. The State of Washington is currently piloting a GIS approach for examining their data on agricultural pesticide-related illness and injury [Baum 2001b].

Any mapping of data must take into account confidentiality. A number of techniques are available to manipulate data to prevent identification of locations. Intentional errors (random shifts within a range) can be introduced into maps that contain street-level detail, making it difficult to determine exact locations. Maps with such introduced error must be annotated to make it clear that points are not exact locations. Simply presenting maps that do not provide street-level detail may also protect confidentiality. However, data presented on a regional level can still identify persons if they include age or race information. Other techniques to protect confidentiality include suppression for small numbers in detailed area maps or use of large symbols that interfere with the ability to precisely locate points [Spinello 2001].

8.2.2 Data Dissemination: Defining What Reports Will Be Useful For a Specific Program

Reports can be comprehensive or short and concise. Each PPSP must evaluate its own needs for different types of reports. It is important to determine the intended audience along with the purpose of the report, as the presentation of data will need to be constructed accordingly. A comprehensive summary report can provide an overview of the surveillance program, including tables and graphs on the overall magnitude of the problem, interpretation of the findings, and a description of recent program accomplishments. These reports are often required by Federal funding agencies and are also produced for multiagency coordinating boards. An annual report to legislatures must sometimes address statutory requirements (e.g., the need to demonstrate how the PPSP is meeting requirements for timeliness of response). There may also be a State interest in presenting data describing agricultural versus nonagricultural applications, illnesses affecting school students, or other issues. PPSPs with multiagency boards typically present data showing referrals from, or cases investigated by, partner agencies.

Reports prepared for the public or for HCPs should be brief–the content limited to one or a few issues. A case vignette or case series presented with some additional descriptive statistics may be useful for communicating a risk or prevention message to the public or engaging HCPs' interest in case identification, treatment, and reporting.

Reviewing reports produced by other PPSPs may provide other helpful ideas. In addition, the SPIDER program provides some standard descriptive tables that surveillance programs may find useful to include in their reports. See Appendix C for sample tables found in SPIDER. Finally, a recent article that provides national findings from the SENSOR-Pesticides Program may provide helpful ideas for report content [Calvert et al. 2004].