4. Data Collection and Management

4. Data Collection and Management ■

4.1 Introduction

This chapter addresses some specifics of documenting PPSP procedures, including data collection and data management. Developing case investigation procedures, forms, and a data management system are important aspects of surveillance system design. (Case investigation procedures are covered in Chapters 5, 6, and 7.) If the appropriate information is not routinely collected, keyed, analyzed, interpreted, and disseminated, the goals of surveillance will not be realized. To some extent, the level of resources available to the program will dictate the amount of information that is routinely collected.

4.2 Data Standardization

At the outset, we cannot overemphasize the importance of using the standardized variables and the standardized case definition. Applying a standardized format for data collection makes aggregation of data across States feasible. The ability to aggregate data is valuable both at the State and Federal level. Potential users of the aggregated data include regulatory agencies, public health policy makers, researchers, programs conducting worker education, the public, and the medical community.

The large number of pesticide products on the market and difficulties in obtaining case reports makes the pooling of all available data particularly desirable. The ability to evaluate pesticide poisoning by product, crops, and geographic location greatly enhances the ability of States to evaluate whether limited case reports in their jurisdictions are reflecting larger problems linked to the specific uses of a pesticide product. Applications for this type of surveillance data at the national level cover a broad range of functions. Regulatory agencies can use the aggregated pesticide poisoning data to guide the development and amendment of regulations, target enforcement efforts, and evaluate the effectiveness of current control mechanisms. Pesticide users, commercial and agricultural pesticide applicators, and users of consumer pesticide products would all benefit from additional information that increases the understanding of risks associated with pesticide use.

4.3 Documentation of Procedures

The processes used to investigate cases, classify cases, enter and analyze data, provide feedback to reporters, and disseminate information are integral to a successful surveillance program. Documenting these procedures is often relegated to the bottom of the list of program management tasks. While programs can function without written documentation, it is certainly not advisable. Documentation provides guidance to all staff for a consistent approach to program objectives. Written policies and procedures make it easier to justify various decisions, including whether a particular pesticide exposure event will be investigated. As program procedures and policies change over time, written documentation facilitates the identification of those changes that might influence data analysis findings. Newly developed surveillance programs are often dependent on only one or two staff people

who are working on many fronts. It is common for procedures to be developed and passed along verbally. Regrettably, if these staff leave the program, the procedures are often lost and must be recreated by new staff. For some procedures, a simple, bulleted list will suffice for documentation; for others, more complete instructions are desirable. The program should also maintain an orientation checklist for new staff to ensure that critical issues are covered in training. Topics of particular concern that should be documented include confidentiality, and safeguards to ensure employee safety and health when performing investigations.

Procedures should be developed for the following activities:

1. Case report management (intake, investigation, closure, classification, and feedback to reporters)

2. Data entry, quality assurance, and control

3. Data analysis

4. Data dissemination

Written policies and procedures should be developed for protecting the confidentiality of case report information at all stages of intake, investigation, and analysis. Examples of PPSP procedure documentation may be available through requests to established PPSPs. (Links to State program offices are in Appendix G under State PPSP Contact Information.) At later stages, policies and procedures for archiving data should also be implemented. These issues and the use of data to target activities and develop intervention strategies will be discussed in Chapters 5 and 6.

4.4 DATA COLLECTION

The PPSP collects data on each poisoning case, and these data are organized using variables.

For all variables that are collected, States are encouraged to use standardized formats. Recommended formats are listed in *Standardized Variables for State Surveillance of Pesticide-related Illness and Injury*. (Copies of this document can be obtained from the NIOSH Web site, http://www.cdc.gov/niosh/topics/pesticides/, or by calling 1–800–365–4674.) The standardized variable format includes variable names, definitions, variable types, widths, and clarifying comments for variables that are considered desirable for all States to collect. Core variables that are critical for States to collect and transmit to CDC are indicated by asterisks. The variables are divided into general subject areas. Within any given subject area, variables are available that allow States to provide a brief narrative description about the data. Some additional discussion and clarification of the variables are provided below. The nature of the data collected for this condition usually dictates that States use a relational file structure and not a flat file structure.

4.4.1 STANDARD VARIABLES TO BE COLLECTED BY PPSPS

4.4.1.1 ADMINISTRATIVE AND DEMOGRAPHIC VARIABLES

The variables in this category include information about the source(s) of the report, relevant dates, event identifiers, county and State of exposure and residence, sex, age, Hispanic ethnicity, and race. These variables are used to describe the demographic characteristics of cases, track the geographic distribution of cases, and ensure that cases and events are linked without duplications. Not all of the variables needed at the State level are included in the standardized variable document (e.g., personal identifiers and addresses of the cases). However, these and other identifying and tracking variables are captured in the SPIDER database pro-

gram. Note that race is not captured in the standard format currently recommended by CDC, since the CDC recommended format makes collection and analysis of race information more complex for persons who are multiracial. The race variable found in the standardized variable document is structured according to the CDC standard in effect at the time the document was initially developed and is considered easier to use by participating States.

4.4.1.2 OCCUPATION AND INDUSTRY DATA

Coding of the occupation and industry of an affected individual can be accomplished by using one of several different standard coding systems. Codes for occupation can be based either on U.S. Bureau of Census (BOC) codes [NCHS 2003] or the 2000 Standard Occupational Classification (SOC) codes [OMB 2000]. Codes for industry can be based upon BOC codes [NCHS 2003] or North American Industry Classification System (NAICS) codes [OMB 1997]. Note that industry and occupation codes are periodically updated, with the BOC codes being revised every 10 years for use with the decennial census. The current BOC codes were used on the 2000 U.S. Census occupation and industry data, and these codes are referred to as the 2000 BOC codes. The 2000 BOC occupation and industry codes are based on, but are not identical to, the 2000 SOC and NAICS codes [OMB 2000], respectively. BOC codes are always 3-digit codes and therefore cannot provide the detailed industry coding provided by NAICS (which can code to 6 digits) nor the detailed occupation coding provided by SOC (which can also code to 6 digits). The NIOSH SENSOR-Pesticides Program recommends using the BOC codes for occupation and industry. This is because the number of workers in each of the BOC industry and occupation codes is available from Current Population Survey (CPS) data. CPS data can be used as the denominator to calculate rates of illness by industry and occupation. Although having all States use the same industry and occupation codes will facilitate the aggregation of data across States, States should choose the coding system that best suits their needs. Whichever coding system is chosen, States should use that system's most up-to-date codes.

Crosswalks are available to convert the NAICS codes into 2000 BOC industry codes, and to convert 2000 SOC codes into the 2000 BOC occupation codes. In order to convert industry and occupation data that may have been coded using older coding schemes, crosswalks are available to convert the 1990 BOC industry codes into the 2000 BOC codes and NAICS codes, and to convert the 1990 BOC occupation codes into the 2000 SOC codes and 2000 BOC occupation codes. All can be accessed at http://www.census.gov/hhes/www/ioindex.html.

Training on how to code occupation and industry is available periodically through the National Center for Health Statistics. This training usually covers all of the major occupation and industry coding systems.

4.4.1.3 EXPOSURE DESCRIPTIONS

The variables in this subject area help characterize the exposure. They describe the type of exposure (drift, direct spray, indoor air, contact, etc.), route(s) of exposure, whether the exposure was intentional, the person's activity at the time of exposure, and PPE worn by the exposed person. They also capture information about the equipment used to apply the pesticide, what the intended target of the application was, where the pesticide was being applied, and where the person was located when exposed (e.g., farm, nursery, home, school, manufacturing facility, etc.).

4.4.1.4 CHEMICAL INFORMATION

This section records information about the pesticide products associated with the exposed person's illness or injury. The system is not designed to capture information about non-pesticidal products such as fertilizers and *adjuvants*. Pesticide product information provided in SPIDER is adapted from the EPA Pesticide Product Information System (PPIS). This system can be accessed from the web at http://www.epa.gov/opppmsd1/PPISdata/index.htm. It is available in a searchable format on the Web site maintained by CDPR at http://www.cdpr.ca.gov.

States are strongly urged to collect sufficient data to permit full identification of the pesticide product whenever possible. However, at a minimum, pesticide functional class and product chemical class must be collected. This is in recognition that sometimes only minimal exposure information is available.

4.4.1.5 HEALTH EFFECTS DESCRIPTORS

This set of variables captures information about biological monitoring, medical diagnosis, pre-existing conditions, whether the person died, signs and symptoms, type of care received, and whether the person lost time from work or regular activities.

4.4.1.6 INVESTIGATION FINDINGS

These variables include enforcement agency findings, plus case investigation findings from the agency managing the surveillance program. Some variables are also specifically related to the Worker Protection Standard (WPS). The WPS variables address whether

■ the incident involved re-entry into an area, field, or greenhouse treated with pesticide

■ the worker had been informed of the re-entry interval for the treated area

(See Section 5.8.1 for more information about WPS.) In addition, a variable captures information about whether the product label was followed.

4.4.1.7 CASE CLASSIFICATION

These variables collect information about the components of the final case classification using the *Case Definition for Acute Pesticide-Related Illness and Injury Cases Reportable to the National Public Health Surveillance System* (NPHSS) described in Chapter 7 and provided in full in Appendix D. There is also a variable to record a separate case classification using either a separate State classification matrix, or to override the NIOSH classification matrix. For cases meeting the definition for reporting to the NPHSS, an additional component of case classification is a severity score of the illness/injury.

4.4.2 OPTIONAL VARIABLES

The variables in the standardized variable list that are not marked as core variables are all considered important, but are ones that some States may choose not to collect because of resource limitations. States are urged to collect as many standardized variables as possible. As already mentioned, the standardized variables include only those variables needed for national aggregation of data. Additional variables are needed for States to track and manage cases. Examples of some of these variables include personal identifiers; address and telephone number of the exposed person; name, address, and telephone number of HCP(s); laboratory sample tracking and results information for environmental and biological specimens; and information about animals (pets, livestock, and wildlife) affected by pesticide exposure. PPSP may want additional flags for particular types of cases that are of interest or concern at the

State level. Some but not all of these variables are captured in the SPIDER database.

The SPIDER system does not provide a tracking system to determine what information has been sent to or received from providers, individuals, and partner agencies. Developing a generic tracking system that would meet all States' needs is not feasible since the investigation and regulatory process is so different within each State. It is important for each PPSP to develop its own system for tracking cases. A tracking system can help to ensure that investigations are timely, that all necessary case information and medical or confidential information releases are obtained, that regulatory agency referrals and reports are received, and that appropriate feedback is given to relevant individuals, HCPs, employers, contract pesticide applicators, and partner agencies.

4.4.3 INTRODUCTION TO THE SPIDER PROGRAM

The SPIDER program is a data manager for collecting, managing, and reporting pesticide illness and injury data. Designed for NIOSH by the New York State Department of Health, the Program prepares data in the proper format for transmittal to NIOSH, and provides some pre-programmed reports used by PPSPs and NIOSH. The software was created using Microsoft Visual FoxPro, Version 5.0c, and Visual ProMatrix 5.0c. You do not need to purchase these products to run SPIDER [New York State Department of Health 1997]. Although additional reports can be created within SPIDER, more complex data analysis will require statistical analysis software (SAS).

System requirements for running SPIDER:

- Any IBM®-compatible computer with an 80486DX processor or higher.

- 32 MB RAM; 64 MB RAM recommended.

- A hard disk with 150 megabytes of free space. This will grow as cases are added.

- A 3 ½" floppy drive and a CD-ROM drive.

- VGA or higher resolution monitor running at 256 colors or more. SVGA (600 x 800) or XGA 1024 x 768) recommended.

- Microsoft Windows 95. (This is a minimum requirement. SPIDER also runs on more current systems, e.g., Windows 2000, NT4, and XP.)

- Installed Windows fonts: Arial, Courier New, and Times New Roman.

- A mouse is very helpful but not required.

This system can be installed on a local area network (LAN) for multiuser access. SPIDER is not equipped to upload cases reported electronically.

OTHER OPTIONS FOR A SURVEILLANCE DATABASE

Some States have chosen not to use SPIDER, and have developed their own data systems for collecting information about pesticide-related illness and injury. If a State decides to develop its own database system and wants the ability to easily compare data with other States, and to contribute to national data, it is important to follow the standardized variable formats. It is equally important to contact NIOSH when developing a surveillance database to ensure that your system will readily transmit the necessary data in the desired format, and that you have the current version of the standardized variables. Some States have experienced problems incorporating chemical information into their databases in a way that will permit aggregation with data from other States. NIOSH may be able to provide assistance to ensure that

chemical product data are collected and transmitted in a standardized fashion.

If a system other than SPIDER is used, the system should include documentation that describes the database, including a data dictionary, file structure, and table relationships. There should be written procedures for installation, operation, and maintenance of the system including how to backup the system.

4.5 DATA MANAGEMENT

This section provides a brief overview of the elements needed for data management. It is provided as a reminder to new PPSPs that these are issues and elements that *must* be included in any data gathering program. For in-depth information about data quality assurance, data quality control staff should refer to the broad range of published literature and training programs available on this topic.

4.5.1 GENERAL GUIDELINES FOR DATA MANAGEMENT

The importance of documenting surveillance procedures has already been emphasized. Procedures for entering reports into the data system, mechanisms to prevent duplicate entries, and management of discrepancies in information when a report is received from multiple sources should all be documented.

Having protocols for case triage and management, along with routine daily or weekly review of open cases will help ensure that data collection is complete and timely. Staff must be trained to have a clear understanding of the procedures and to strive for complete and accurate data. A clear procedure (e.g., a written protocol or an assigned coding administrator) should be in place to ensure that narrative data coding,

interpretation of medical information, and pesticide product identification are performed in a consistent manner. There should be a system to monitor the quality of data entry to ensure the results comply with acceptable error rates. Staff should receive feedback on their data entry performance.

The SPIDER system contains many automated edit checks, as well as an audit trail, error reports, and missing data reports. If an alternative system is used, it should contain edits for missing data and errors (e.g., the program should identify codes that are outside acceptable ranges and illogical date sequences). Checks for duplicate records, blank records, orphaned data, and other anomalies created by changes in relationally linked data should also be part of the system.

4.5.1.1 CONFIDENTIALITY AND SECURITY

The PPSP must develop systems for maintaining the confidentiality of hard copy and electronic records. Confidentiality procedures should be in writing to ensure that staff are clear about these procedures. Staff must understand the procedures and follow them routinely. Staff must also know whom to contact with questions. Records containing confidential information should be kept in locked file cabinets. Electronic systems should have passwords, and access to the system should be controlled.

4.5.1.2 SYSTEM BACKUPS

The administrator of the data program should establish written protocols for data system backup. Typically, there is a daily backup of data entered or edited during that working day; there should be a routine weekly or monthly backup as well. Safeguards for virus protections should be in place and routinely updated.

4.5.1.3 TRANSMITTING DATA TO NIOSH

Annually, NIOSH assembles an aggregated database using data provided by participating PPSPs in the United States. In the past, NIOSH has requested that these data be provided by May 1. This gives the States a 4-month lag period from the end of the calender year to close out cases reported during the previous year. Grants and cooperative agreements awarded by NIOSH to fund PPSPs usually require this data sharing. PPSPs that receive no funding from CDC share their data voluntarily.

For SPIDER-using States, transmittal of data is relatively simple. These States need to ensure that their data are complete and then use the *Export to NIOSH* file command to prepare a zipped data file that can be transmitted to NIOSH. States not using SPIDER should contact NIOSH during development of their database system to discuss data transmittal issues. (Call 1–800–356–4674 or see http://www.cdc.gov/niosh/pestsurv/default.html.) All personal identifiers are stripped from the data before transmission to NIOSH.

Pooling surveillance data to create an aggregated database permits the creation of knowledge to prevent and control acute pesticide-related illness and injury. The aggregated database is shared with contributing PPSPs, NCEH, and EPA. Once data are checked for quality, accuracy, and absence of personal identifiers, it is made available for public access.