APPENDIX A

Preliminary and Expanded Lists of PUR Problems and Issues

Preliminary List of PUR Problems and Issues
For Discussion at the PUR Conference, May 8, 2000

Purposes:

1. Information for DPR management and conference attendees.
2. To facilitate discussion by conference attendees of additional problems and issues faced by PUR users.
3. To provide an outline for the PUR Technical Advisory Committee's report to DPR on problems with the PUR that affect work done by committee members

Outline of Problems and Issues

1. **Data Collection**
   - A. PUR regulations and requirements
     - i. Unclear regulations. Current regulations are confusing, incomplete, or result in inaccurate values. These confusions sometimes result in errors by applicators in filling out their use reports.
     - ii. Enforcement letters are not unified into one document. Some of the procedures were more fully explained in a series of enforcement letters that were distributed to the counties after the regulations were instituted. Confusion on data entry protocol has resulted because the enforcement letters were never organized clearly into a single document (manual).
     - iii. Above issues lead to inconsistency among counties particularly for items B and C below. This makes it difficult to understand the data since the meaning of the data may vary for different counties.
   - B. General crop category vs. specific crop. Sometimes a general crop category is entered into the PUR database instead of the specific crop grown. For example, a grower may record a crop as "citrus" rather than "oranges" because the pesticide is labeled for use on "citrus." For many of our analyses we need to know the specific crop, not a general crop category.
     - i. Qualifier code. The qualify_code, which can be used to provide more specific crop information, is usually not used. Unfortunately, the few counties that do use the qualify_code do not follow the same protocol for determining its value.
     - ii. Planting sequence. The PUR data field planting_seq is usually not reported for crops that have multiple plantings in a year. There are many instances where analysts need to know how many plantings were done, the timing of the plantings, or how much pesticide was used per planting.
   - C. Site location identification. Counties are inconsistent in definition of site_loc_id and in some cases do not clearly designate a geographical location. This information is important when we need to make calculations on field level statistics. It is also necessary to arrive at total acres planted and % crop treated.

D. Acres treated are sometimes reported as the entire field even when only part of the field is treated. This is apparently the recommended practice given in an enforcement letter for strip treatments. However, this practice produces incorrect statistics for acres treated, overestimates percent acres treated, and underestimates the rate of use.
E. The data source for acres planted is variable. Data for this field are transferred from the grower's permit or from the PUR reporting form. If taken from the permit, this field is not always updated when the actual acres planted (as indicated on the PUR reporting form) is different. This field has been used to determine the number of acres planted of each crop in California. However, for some crops the acres planted from the PUR differs considerably from the County Agricultural Commissioner's (CAC) crop reports and the California Department of Food and Agriculture (CDFA) values. Why do these different sources differ so much and which values are the more accurate?
F. The number of applications is not consistently reported in monthly summary reports. Thus, it is difficult or impossible to interpret these data.
G. Reporting amount, area, or volume treated in commodity fumigation. People will sometimes use or report fumigation of an entire warehouse even if only one crate is actually treated.
H. Amount of product used. Growers sometimes will report the total amount of pesticide plus diluent used, *i.e.*, the total volume of material in the tank; they should report the amount of the pesticide product only. This results in too much product reported.

2. Data quality

A. Data validation at CAC offices is inadequate. The county data entry programs do not check for all of the errors that are checked for at DPR.
B. Poor data entry screen design. Errors in typing are difficult for data entry staff to check, often owing to poor data entry design. For example, units of measure are coded as numbers rather than the actual name of the units.
C. Illegible PUR forms.
D. Duplicate submissions or data entry. Sometimes records are entered twice or sent twice to DPR, resulting in duplicate records. Some records seem to have been "fixed," but the original record is not always deleted. This results in over reporting applications and amounts applied.
E. Lack of error corrections. Counties do not always correct errors that are sent back to them.

3. Data Access

A. Data formats. Some of the data formats used in the files that are distributed are difficult to use by some people, for example, the fixed field text files on the PUR CD-ROM.

B. Incomplete documentation. Users need access to a discussion of critical data analysis issues in order to accurately handle PUR data. Some of these problems were outlined in section 1.A, for example, the differences between counties in how some data fields are defined or used.

C. Limited summary statistical data available. There is no easy method to view different kinds of summary statistics from the PUR except that provided in a limited set of PUR summary tables.

D. Active ingredient (AI) names. Some of AI names and classifications are inconsistent with other databases or contain errors.

   i. Different databases (DPR, U.S. Environmental Protection Agency (US EPA), U.S. Department of Agriculture (USDA)) use different AI names making it difficult to compare databases. These institutions have chemical synonym databases that should be synchronized and users need to be informed of potential naming discrepancies.

   ii. DPR produces reports with very specific chemical classifications of some AIs but most researchers are interested in a more broadly defined AI classification. For example, DPR lists more than 20 different esters and salts of 2,4-D but in summary reports, an aggregate summation for all 2,4-D variants would be more useful.

   iii. The oils are particularly troublesome; oils are listed under several different names that are not clearly explained and appear to be inconsistent.

E. Crop names. DPR uses different crop names and classifications than other agencies (such as USDA and CDFA), making it difficult to compare statistics.

F. Adjuvants and inert ingredients. Adjuvants are not clearly and accurately separated from other AIs. Some products are registered as adjuvants but the Label Database is inaccurate in recording this. In other cases, an adjuvant may be considered as part of a product and would not be separately reported. Thus, information on adjuvant use in the PUR is incomplete and can't be interpreted.

G. Precision of data. Some data are reported to 4 decimal places when measurement of the particular parameter is less refined.

4. Additional data needs

A. Target pest. Lack of pest information makes it difficult to monitor pest infestations or target IPM programs. It also misses an opportunity to remind growers frequently about the importance of pest identification (a basic principle of IPM). Alternative pest control analyses (including economic impact analysis of regulatory issues) would be considerably more defensible if analysts knew how pests were being controlled from PUR data.

B. Date crop planted. This information would help users distinguish separate plantings. One example where this is important is determining the total acres planted of crops with a growing season that spans two calendar years. The acres planted, calculated from the PUR for these crops, will be overestimated when pesticide use is reported for that crop in two different years. We currently have no way of knowing whether a field had two different plantings or not. Since most crops do not span two years, we typically assume fields are different from one year to the next.

C. Date crop harvested. This information is used by US EPA and DPR's medical toxicology unit in their residue analysis programs.

D. Acres harvested. This would help users relate use to what gets into the market.

E. Crop yield

F. Additional site information
   i. Schools, day-care centers
   ii. Federal land
   iii. Rights of way
   iv. Water bodies
   v. Indian tribal lands
   vi. Homeowners
   vii. Farm animals
   viii. Institutions
   ix. Organic fields
   x. Genetically modified crops
   xi. Hospitals
   xii. Proximity to sensitive sites, e.g. schools, organic fields, etc.

G. Identify cancellations of chemical or chemical/commodity combinations.

H. Label rate.

I. Pesticide type (insecticide, herbicide, etc) or class (OP, carbamate, etc). DPR's Label Database provides this information, but the information in this database have is inaccurate and the categories are not necessarily easy to apply.

J. Identify section 18 applications.

5. Miscellaneous issues

A. Ongoing compliance monitoring. Users of the data need to know the level of reporting compliance to have a sense of the completeness of the data.

B. Errors in the Sales database. The Sales database has errors. It is used to get a general idea of possible pesticide use for which there are no reporting requirements and to get estimates of the level of reporting compliance.

C. Label Database.

i. Errors. The label database contains errors that affect data in the PUR, such as incorrect percent AI or specific gravity in pesticide products and errors in the classification of AIs.

ii. Crop names on labels. The PUR loses information on crop specificity because DPR's error checking process requires the applicator report only names of crops listed on the label. For example, if an application occurs to navel oranges and the label only lists citrus, citrus will be reported as the crop treated, not navel oranges.

iii. Maximum label rates. The label database does not contain label rates. This information is critical for error checking.

D. PUR error reporting and correction procedures.

i. Procedure for reporting errors. There is no procedure for reporting to DPR errors found in the PUR (by people in and outside of DPR).

ii. Reporting errors found and corrections made to the public. DPR has no process for informing PUR data users about recently found errors or corrections made to the database.

iii. Peer review of DPR error-checking procedures. DPR has not had their error-checking procedures reviewed by knowledgeable people from outside of DPR.

EXPANDED LIST OF PUR PROBLEMS AND ISSUES

Background: This list was formulated as a result of discussions with a technical advisory committee convened by DPR, as well as issues raised at the PUR conference.

1. DATA COLLECTION ISSUES

A. PUR regulations and requirements

1. Unclear regulations

2. Enforcement letters not unified into one document

3. Above issues lead to inconsistency among counties particularly for items B and C below

4. Do not allow counties to re-use retired grower_ids

B. General crop category vs. specific crop

1. Qualifier codes

2. Planting sequence

C. Site location ID

1. Require all reports of different applications to one field in a year to use the same grower_id [Editor: site_loc_id?] each time, so all fields can be uniquely identified).

D. Acres treated

1. Spot treatments

2. Strip/band treatments

E. Acres planted

F. Number of applications not consistently reported in monthly summary reports

G. Reporting amount, area, or volume treated in commodity fumigation

H. Amount of product used: Product vs. diluted amount.

I. PLSS reporting errors. Consider hiring an outside contractor to conduct call backs to verify information, similar to the reliability checks that epidemiologists perform.

2. DATA QUALITY ISSUES

A. Data validation at CAC offices (or field level)

B. Poor data entry screen design

C. Illegible PUR forms

D. Duplicate submissions or data entry

E. Lack of error corrections

F. In the amount used field in the PUR convert all values to either pounds or gallons (don't use measures such as pints, quarts, liters, etc.)

3. DATA ACCESS ISSUES

A. Data formats:
1. Fixed field file structure on CD-ROM is a problem for some users.
2. Provide data as an Oracle dump file.

B. Incomplete technical documentation

1. Data dictionary with informational notes
2. Example SQL and PL/SQL code

C. Limited summary statistical data available
1. Map on watershed basis using GIS.
2. Be able to sort lists on the Web site.

D. Active ingredient names

1. Inconsistent names in various databases (DPR, US EPA, USDA)
2. Specific vs. more general names (eg 2,4-D vs list of all the different esters and salts of 2,4-D)
3. Categories for the different kinds of oils
4. Include trade names

E. Crop names. Inconsistent names in various databases

F. Adjuvant and inert ingredient reporting

G. Inappropriate precision in reports

H. Data Integration. Link the PUR with other databases (e.g. DPR's illness reporting system).

1. Provide better links in the data set to information needed for modeling, e.g., environmental half lives.
2. Connect use and permit data.
3. Identify toxicity category (e.g. carcinogens, etc.)
4. Relate to residue monitoring data at the field level.
5. Relate sales data to use data.

I. Allow growers to view a history of their past use (this might be a good incentive for them to use a Web-based reporting system).

J. Release PUR data monthly

K. Do not eliminate all outliers.

4. ADDITIONAL DATA NEEDS

A. Target pests
1. Categorize target pests

B. Date crop planted

C. Date crop harvested

D. Acres harvested

E. Crop yield

F. Additional site information (change monthly summary to site specific)

1. Schools, day-care centers
2. Federal land
3. Rights of way
4. Water bodies
5. Indian tribal lands
6. Homeowners
7. Farm animals
8. Institutions
9. Organic fields
10. GM crops
11. Hospitals

12. Proximity to sensitive sites, e.g. schools, organic fields, etc.
13. Senior care centers
14. Golf courses

G. Identify cancellations of chemical or chemical/commodity combinations

H. Label rate

I. Pesticide type (insecticide, herbicide, etc) and class (OP, carbamate, etc)

J. Identify section 18 applications

K. Geography. Link PUR data to satellite imagery, aerial photography, and land use survey data sets.

L. Identify applicator (PCO) that made the application, and PCA affiliation.

M. Collect pesticide sales data by county.

N. Collect more information on structural and industrial use

O. Identify genetically modified organisms (GMOs)

5. MISCELLANEOUS ISSUES

A. Ongoing compliance monitoring

B. Errors in sales database

C. Label Database

1. Errors in percent AI and specific gravity

2. Crop names on labels

3. Maximum label rates

D. PUR error reporting and correction procedures
1. Procedure for reporting errors
2. Reporting errors found and corrections made to public
3. Peer review of DPR error-checking procedures
4. Find incentives to promote error correction.

E. Link use data with environmental media data (i.e. not only where the pesticide was released, but where it goes).

F. Notification. Use real-time reporting methods, e.g. using GIS systems, to provide advance warning of an application. Useful for schools, daycare centers, etc.

G. Other data sources. For example, what data elements in 'Doanes' sources are considered proprietary, and what makes it so valuable to US EPA?