PUR Update and Error Checking Proposal

Larry Wilhoit
Department of Pesticide Regulation

Topics of Discussion

- PUR updates
- PUR edit module
- Proposed changes to error procedures
- Conclusion
PUR Update

- PUR annual report late last year
  - Reduction of county staff due to budget cuts
  - DPR’s computer server down for 2 months
  - Found large errors that needed to be corrected
- SPURS and GIS efforts
- PUR edit module

PUR Edit Module

- Changes to PUR require SQL code
- Large back log of corrections to be made
- Now developing interface that lets authorized staff make corrections to PUR data
- Changes to structure of error tables
Proposed Changes to Error Procedures

- Form special workgroup to review and recommend improvements
- Combine loader and outlier error checking
- Include changes to label database tables
- Improve outlier procedures
- Run new procedures on all previous PUR data
- Distribute corrections to PUR on regular basis
- What is percent of non-reported use?

Current criteria for identifying outliers

- Pounds of active ingredient per acre greater than 200 (non fumigants) or 1000 (fumigants)
- Pounds of product per unit treated is greater than
  - 50 times the median rate for all applications of this product on this crop
  - The median plus 50 times the median absolute deviation (MAD)
  - A rate generated by a neural network
Evaluating Outlier Procedures

- We now have a lot more PUR data for developing criteria
- We have many corrections to the data
- We can compare reported rates with maximum label rates for some products

Problems with current procedures

- Criterion 1: rate > 200 or 1000 lbs AI/acre
  - Outlier limits may need to be adjusted
  - Applies only to records with unit treated = acres
  - Ignores distribution of rates
Problems with current procedures

 Criterion 2: rate > 50 * median rate for product and site
  - Many situations with too few records
  - Also ignores distribution of rates

Criterion 3: rate > median + 20 * median absolute deviation (MAD)
  - Makes use of distributions but has problems
  - Too few records for some products and sites
  - Many cases with median deviation = 0
Problems with current procedures

- Criterion 4: rate > limit set by neural network
  - Not easy to understand or create
  - Not based on an explicit model or knowledge
  - Currently lags too many correct records

Possible outlier criteria

- Base outlier limits on rates of each AI over last 5 years
- Use log(rate), which has distribution closer to normal
- Possible outlier limits (for normal distribution would except about $3 \times 10^{-7}$ of values greater than these limits)
  - Rate > median + 5 * STD MAD
  - Rate > trimmed mean + 5 * trimmed STD
  - Rate > 75th percentile + 3 * IQR
  - Rate > 75th percentile + 5 * STD IQR
Issues with proposed outlier methods

- Rates of use may differ considerably for different products or different sites for a given AI
- Rates may change from year to year
- Rates of use are not normally distributed
  - Some distributions are multi-modal
  - Particular rates may predominant

Number of outliers found in PUR 1999 to 2003

<table>
<thead>
<tr>
<th>Rate greater than...</th>
<th>Num Records</th>
<th>Relevant Records</th>
<th>Percent Outliers</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Current criteria</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>200 lbs AI/acre</td>
<td>5,427</td>
<td>9,619,914</td>
<td>0.06</td>
</tr>
<tr>
<td>1000 lbs AI/acre (fumigants)</td>
<td>408</td>
<td>48,599</td>
<td>0.84</td>
</tr>
<tr>
<td>50 * Median by Prod/Site</td>
<td>6,153</td>
<td>9,804,569</td>
<td>0.06</td>
</tr>
<tr>
<td>Neural Net Limit</td>
<td>25,216</td>
<td>9,780,062</td>
<td>0.26</td>
</tr>
<tr>
<td><strong>Proposed criteria:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50 * Median by AI</td>
<td>36,274</td>
<td>13,736,879</td>
<td>0.26</td>
</tr>
<tr>
<td>Median + 5 STD MAD</td>
<td>320,989</td>
<td>13,736,879</td>
<td>2.34</td>
</tr>
<tr>
<td>Mean trm + 5 * trm STD</td>
<td>69,433</td>
<td>13,736,879</td>
<td>0.51</td>
</tr>
<tr>
<td>75th pct + 3*IQR</td>
<td>159,329</td>
<td>13,736,879</td>
<td>1.16</td>
</tr>
<tr>
<td>75th pct + 5*STD IQR</td>
<td>135,865</td>
<td>13,736,879</td>
<td>0.99</td>
</tr>
</tbody>
</table>
Distributions of Log Rates for Diquat Dibromide

Distributions of Log Rates for Oxydemeton-methyl
Distributions of Log Rates for Diphacinone

Distributions of Log Rates for Zinc Sulfate
Conclusions: ways to improve error checking

- Provide more informative error tables
- Provide updates on PUR corrections
- Outliers in rate of use identified more accurately
  - Compare rates with median rate for each AI with exceptions for
    - AIs which have distinct uses
    - AIs which appear in products with more than one AI
    - Nursery crops
  - Use better measures of dispersion of rates