

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 $> 50 * \text{median rate for product and site}$
 - Many situations with too few records
 - Also ignores distribution of rates

Problems with current procedures

- Criterion 3: rate $> \text{median} + 20 * \text{median absolute deviation (MAD)}$
 - 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 $\log(\text{rate})$, which has distribution closer to normal
- Possible outlier limits (for normal distribution would expect about 3×10^{-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

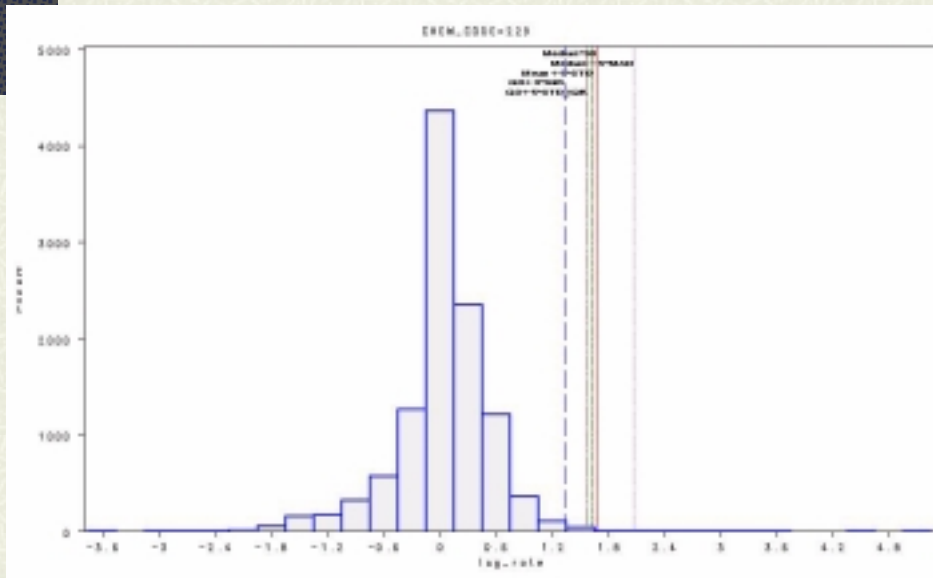
Issues with proposed outlier methods

- ▣ 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 be predominant

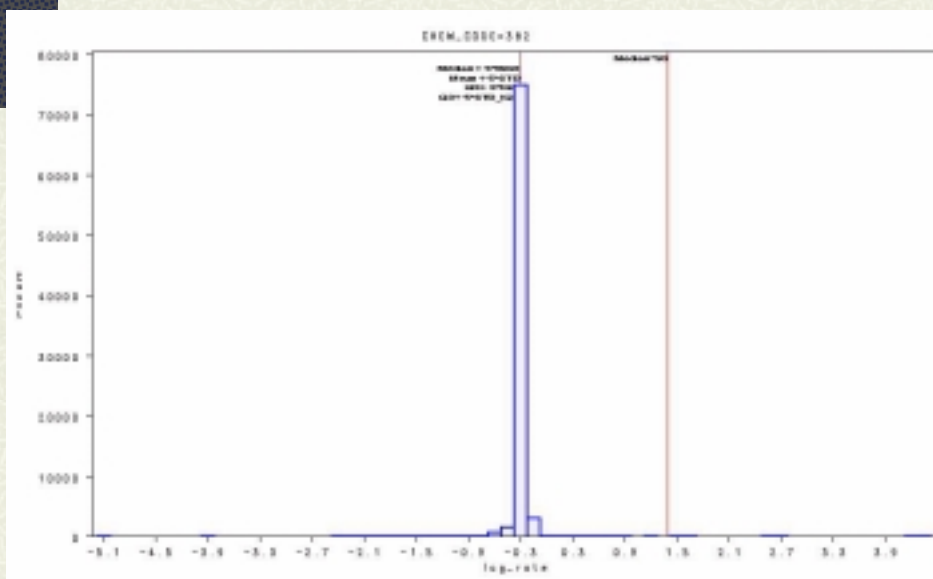
Number of outliers found in PUR 1999 to 2003

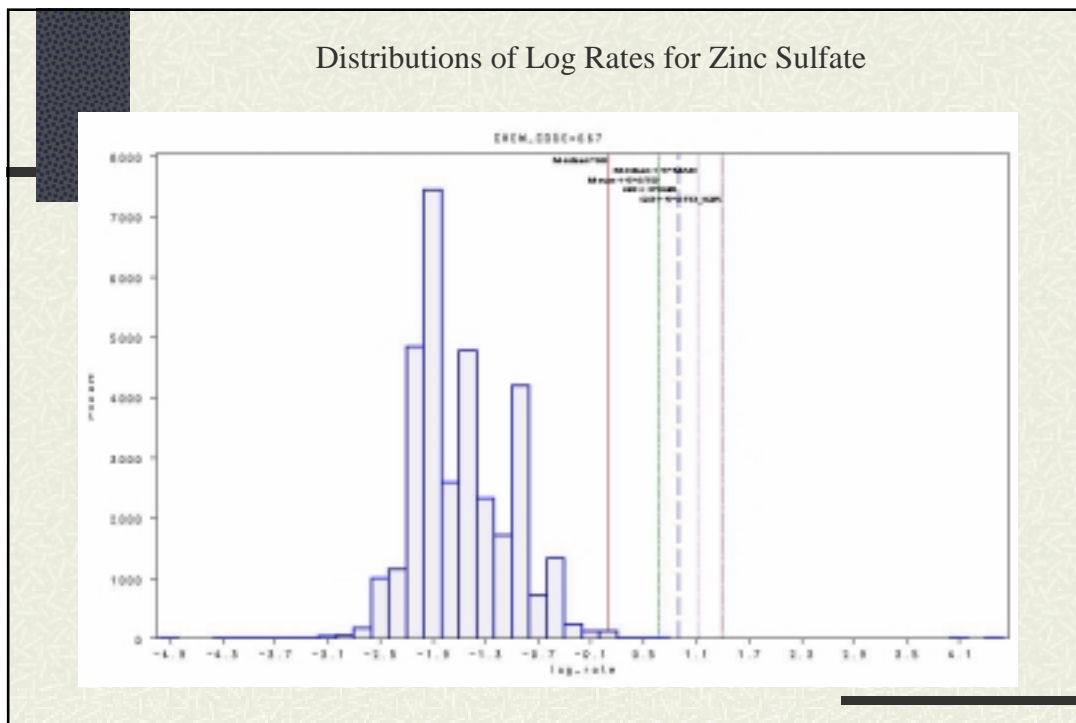
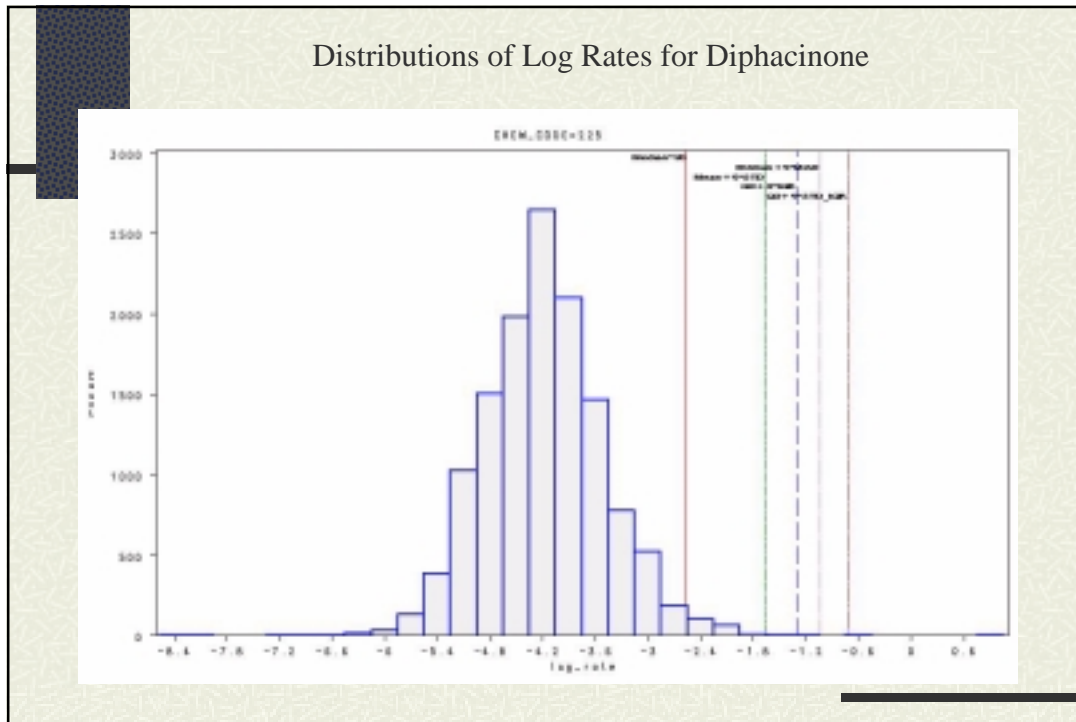
| Rate greater than... | Num Records | Relevant Records | Percent Outliers |
|------------------------------|----------------|---------------------|---------------------|
| Current criteria | | | |
| 200 lbs AI/acre | 5,427 | 9,619,914 | 0.06 |
| 1000 lbs AI/acre (fumigants) | 408 | 48,599 | 0.84 |
|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

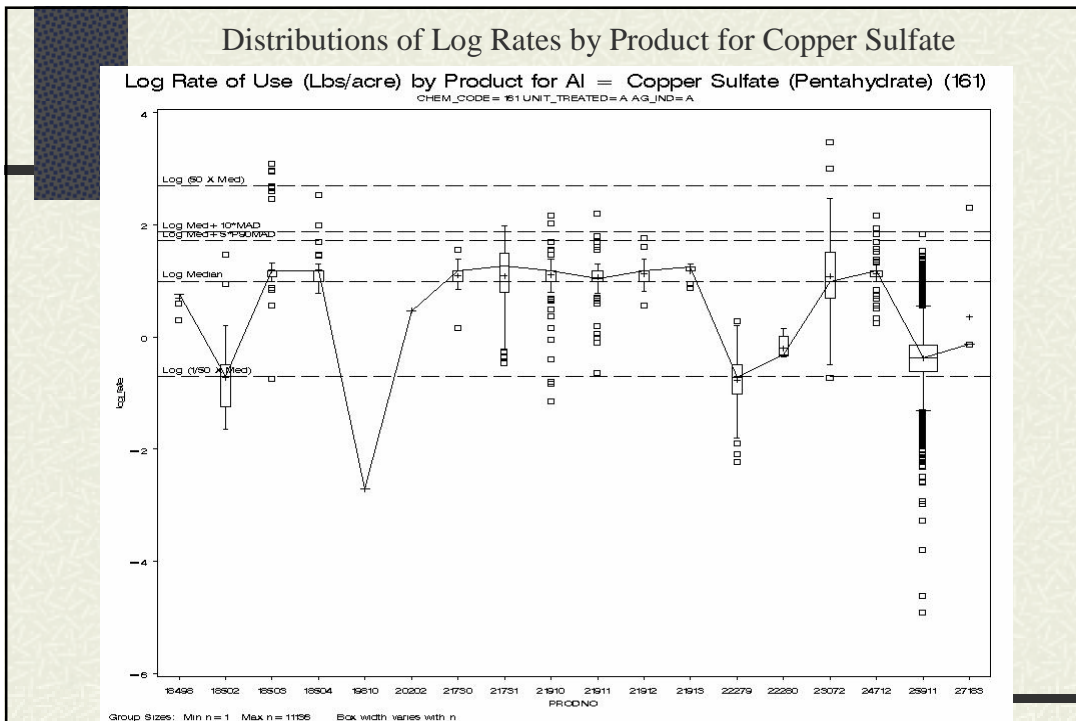
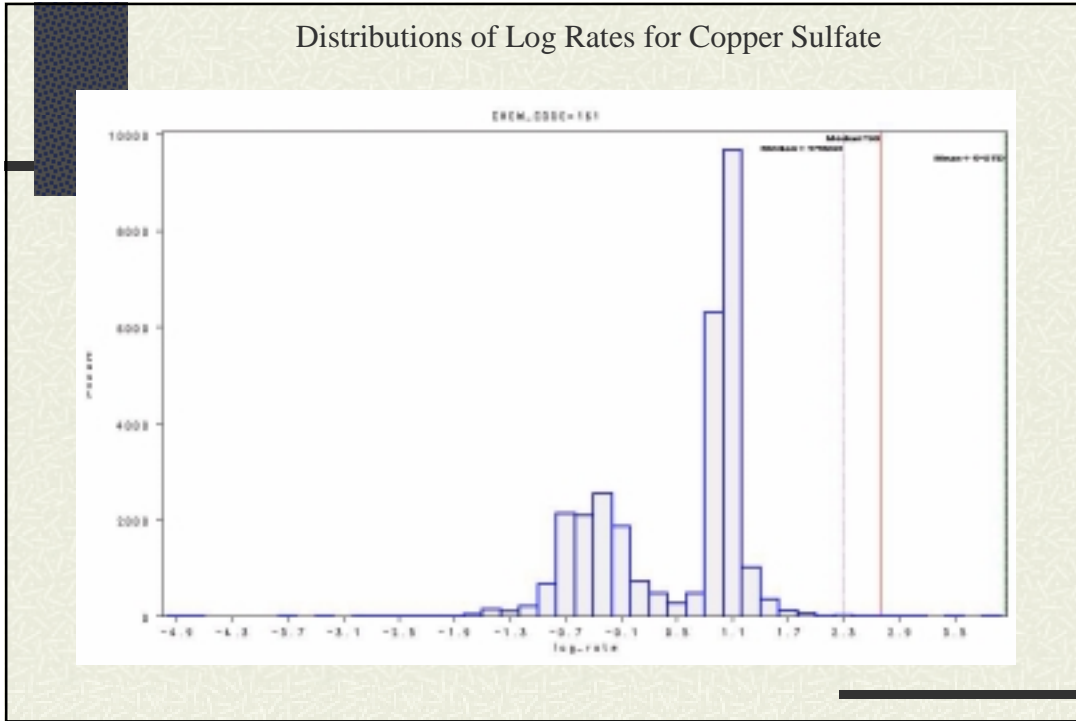
Distributions of Log Rates for Diquat Dibromide

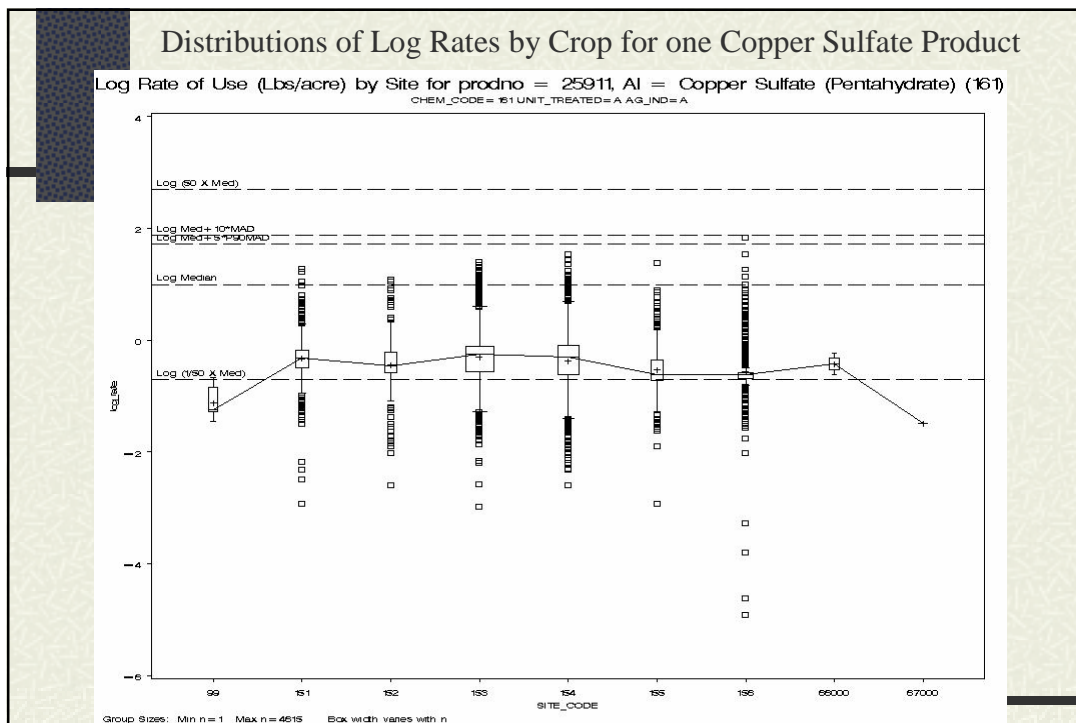
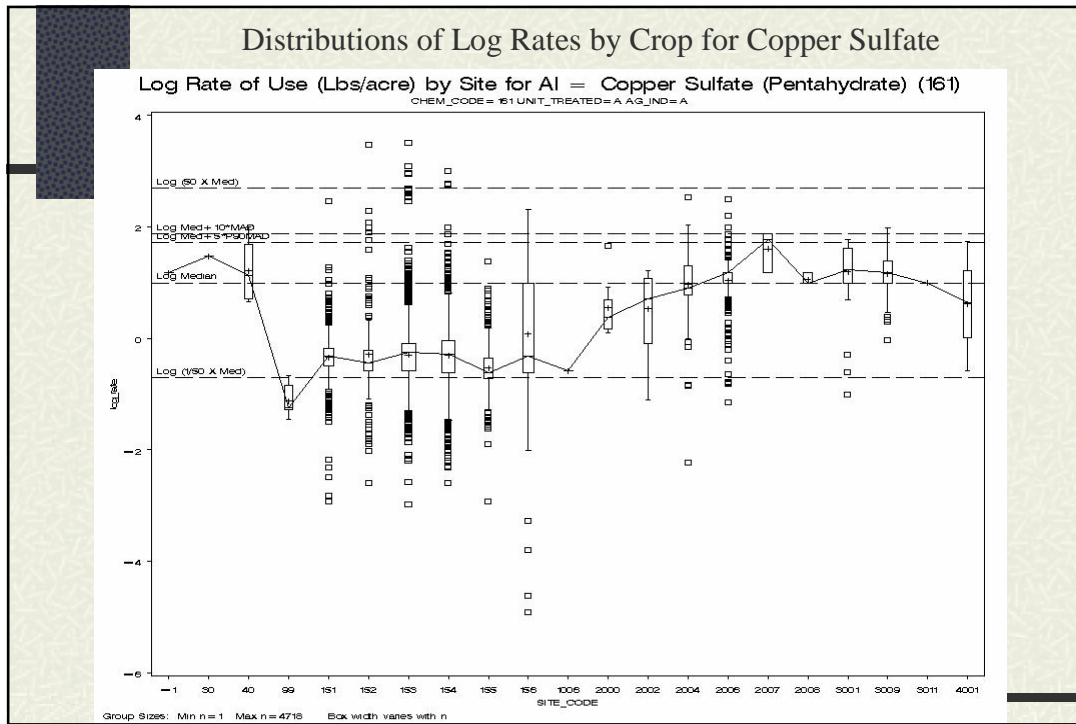


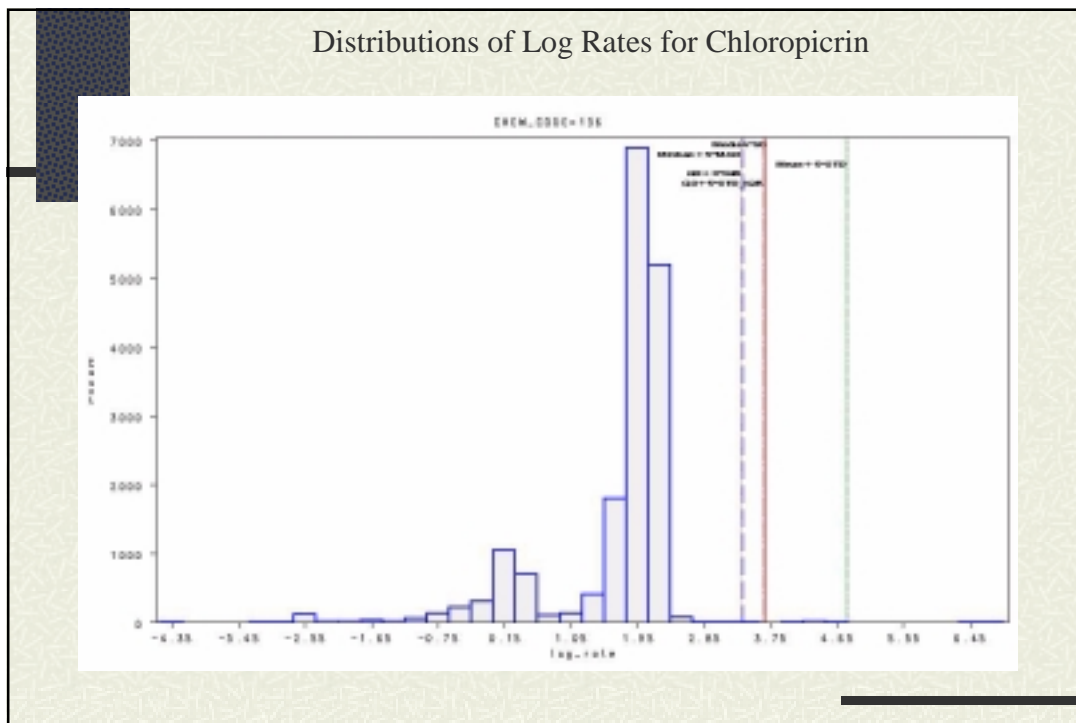
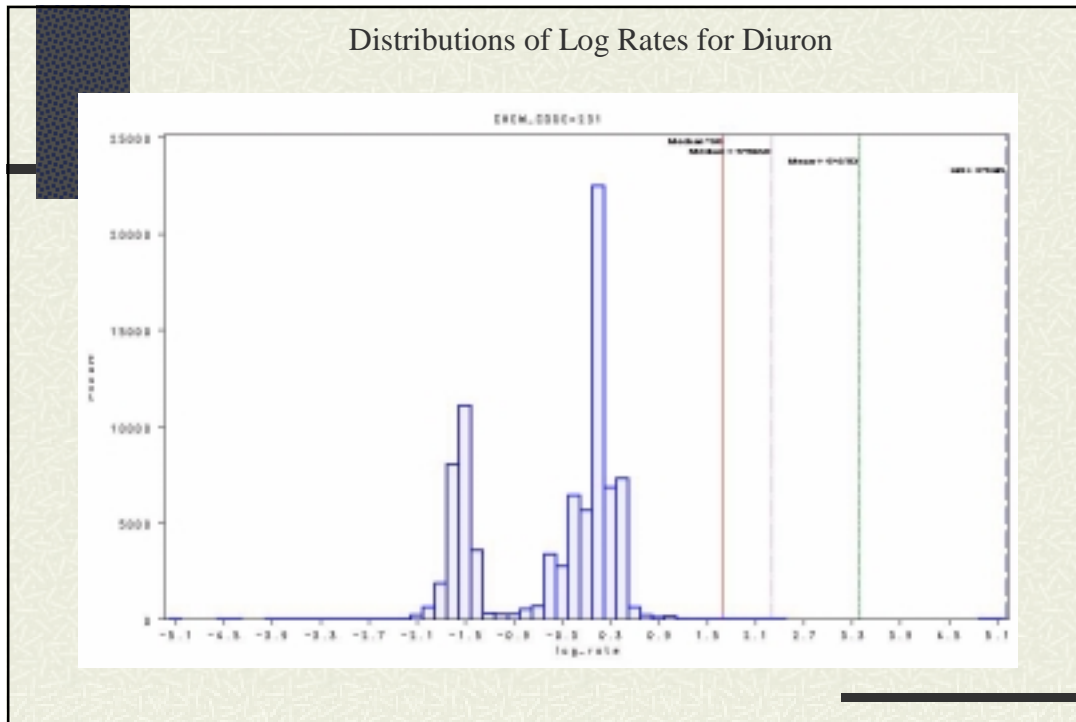
Distributions of Log Rates for Oxydemeton-methyl











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
-