Department of Pesticide Regulation

Brian R. Leahy
Director

Edmund G. Brown Jr.
Governor

MEMORANDUM

TO: Sheryl Beauvais, PhD **HSM-13009**
Senior Toxicologist
Worker Health and Safety Branch

FROM: Holly Jessop *(original signed by H. Jessop)*
Research Scientist I (Epidemiology/Biostatistics)
(916) 324-3517

DATE: November 15, 2013

SUBJECT: METHODOLOGIES FOR CALCULATING SHORT-TERM EXPOSURE USING THE 95TH PERCENTILE

The Agricultural Handler Exposure Task Force (AHETF) currently estimates the magnitude of short-term exposure using three different methods, which each yield a 95th percentile upper-bound from exposure monitoring data. Specifically, AHETF calculates the 95th percentile using:

1) An empirical distribution function with averaging, denoted $P95_S$ (i.e., SAS UNIVARIATE procedure with PCTLDEF=5).
2) A lognormal distribution, using the equation:

   Equation (1): 95th Percentile = $P95_U = (GM)*(GSD)^{1.645}$

   where GM is the geometric mean and GSD is the geometric standard deviation of the measured short-term exposures. Note that equation (1) represents the Maximum Likelihood Estimate of the percentile (Crow & Shimizu 1988), and is equivalent to (Burmaster & Hull 1997; Frank 2009):

   Equation (2): 95th Percentile = $P95_U = \exp\{\mu + 1.645*\sigma\}$

   Where μ is the arithmetic mean of the natural (base *e*) logarithms of the measured short-term exposures, and σ is the arithmetic *population* standard deviation of the same log-transformed data. Frank (2009) instructs California Department of Pesticide Regulation Worker Health and Safety (WHS) staff to implement Equation (2) in Excel as follows:

   95th Percentile = $P95_U$ = EXP(AVERAGE($\log_e$ of measured exposures) + 1.645*STDEV.P($\log_e$ of measured exposures))
3) A lognormal population, using the equation:

   Equation (3): 95th Percentile = $P95_M = (GM_Q)*(GSD_Q)^{1.645}$

   where GM_Q and GSD_Q are the geometric mean and geometric standard deviation derived from mixed models. Such modeling attempts to improve the accuracy of the 95th percentile estimate, by accounting for the fact that exposure measurements are often correlated.

Except in the case of very small N, $P95_M$ will usually provide the most accurate estimate of the magnitude of short-term exposure (because it accounts for more sources of variation), whereas $P95_S$ will usually represent the least accurate estimate (because it tends to exaggerate extreme values). However, $P95_U$ is the easiest exposure estimate to compute. In contrast, the $P95_M$ estimate requires the most complex and time-consuming computations (see Table 1).

The $P95_U$ estimate provides a compromise between accuracy and computational difficulty. Like $P95_S$, $P95_U$ is also readily computed. And, when correlations between the exposure measurements are small (a data characteristic generally striven for in observational scenarios), $P95_U$ is expected to not differ significantly from $P95_M$. However, when large correlations are present between exposure measurements (and especially when there are unequal numbers of monitoring units within different locations or days), then $P95_U$ and $P95_M$ will differ, with $P95_M$ providing the most accurate estimate.

For a number of different scenarios, AHETF has compared the results from these three different methods for calculating the 95th percentile of short-term exposures (see Tables 2 & 3), including their 95% confidence intervals (obtained via parametric bootstrapping). These results demonstrate that the point estimates for $P95_M$ and $P95_U$ can differ, due to correlations present in the data, with $P95_M$ exhibiting a much wider confidence interval. Note that $P95_M$ is often greater than $P95_U$, but can be less (e.g., AHE1001 inhalation). When clustering is minimal, the estimate $P95_U$ is approximately the same as $P95_M$ (albeit with a confidence interval that may be too narrow; e.g., AHE1002).

Because sufficient numbers of simple random samples are typically not feasible to obtain for a particular exposure scenario, the extra effort to calculate $P95_M$ is nearly always desirable as an attempt to correct for the ever-present possibility of cluster effects; AHETF explicitly states that, "In general, it is recommended that the possibility of cluster effects be considered for all statistical analyses." (Klonne & Holden 2007). For this reason, **$P95_M$ is usually considered the most preferred estimate, if available**. AHETF also explicitly states that, "For the 95th percentile, either of the lognormal-based estimates ($P95_M$ or $P95_U$) is recommended. Because it is simpler to calculate, users will probably favor $P95_U$." (Klonne & Holden 2009a & b).

Current WHS policy is to calculate and utilize $P95_U$ (Frank 2009). Presumably AHETF performs the extra work required to obtain $P95_M$ in order to avoid possible criticisms that the $P95_S$ estimate has too large uncertainty, the $P95_U$ estimate may be too inaccurate, and/or that the $P95_U$ estimate may have confidence intervals that are overly optimistic. AHETF also calculates the 95th percentile using multiple methods as part of an effort to demonstrate that a scenario dataset meets a statistical benchmark quality objective. This objective stipulates that the measure of an exposure from the AHETF monitoring program must be accurate to within 3-fold of the true value. For some scenarios this objective may best be met via the (albeit computationally more difficult) $P95_M$ estimate.

Tables

Table 1: Typical Accuracy & Computational Complexity of Short-Term Exposure 95th Percentile Calculation Methodologies

$P95_S < P95_U < P95_M$

Table 2: Dermal Short-Term Exposures (µg per day per lb active ingredient handled), Using the 95th Percentile, with 95% confidence intervals in parentheses

Report #	**$P95_S$ (95% CI)**	**$P95_U$ (95% CI)**	**$P95_M$ (95% CI)**
AHE1001	Not provided	89.8 (52.8, 149.5)	98.7 (40.8, 250.9)
AHE1002	Not provided	20506 (7947, 51024)	20506 (8063, 53342)
AHE1003	60.4 (26.7, 232.0)	74.0 (31.7, 231.8)	90.9 (33.2, 259.3)
AHE1004	37.1 (13.3, 115.4)	34.6 (15.8, 96.4)	40.5 (16.3, 104.4)

Table 3: Inhalation Short-Term Exposures (µg per day per lb active ingredient handled), Using the 95th Percentile, with 95% confidence intervals in parentheses

Report #	**$P95_S$ (95% CI)**	**$P95_U$ (95% CI)**	**$P95_M$ (95% CI)**
AHE1001	Not provided	37.8 (14.8, 92.8)	34.4 (11.1, 109.3)
AHE1002	Not provided	1269 (451, 3429)	1269 (458, 3599)
AHE1003	0.493 (0.147, 3.046)	0.660 (0.174, 3.166)	0.832 (0.195, 3.752)
AHE1004	2.21 (0.46, 3.62)	1.27 (0.56, 2.81)	1.27 (0.56, 2.91)

References

Burmaster, D.E., and Hull, D.A. (1997). Using lognormal distributions and lognormal probability plots in probabilistic risk assessments. Human and Ecological Risk Assessment: An International Journal, Vol 3, Issue 2, 235-255. DOI: 10.1080/10807039709383683

Crow, E.L., and Shimizu, K. (1988). Lognormal Distributions: Theory and Applications. New York, Marcel Dekker, Inc.

Frank, J.P. (2009), Department of Pesticide Regulation Memorandum HSM-09004: Method for calculating short-term exposure estimates.

Klonne, D.R., and Holden, L.R. (2007). Report Number AHE1001: Agricultural Handler Exposure Scenario Monograph: Mixing and Loading Dry Flowable Formulations.

Klonne, D.R., and Holden, L.R. (2008). Report Number AHE1002: Agricultural Handler Exposure Scenario Monograph: Hopper Box Seed Treatment with a Solid Formulation and Planting of Treated Seed.

Klonne, D.R., and Holden, L.R. (2009a). Report Number AHE1003: Agricultural Handler Exposure Scenario Monograph: Open Pour Mixing and Loading of Liquid Formulations.

Klonne, D.R., and Holden, L.R. (2009b). Report Number AHE1004: Agricultural Handler Exposure Scenario Monograph: Open Cab Groundboom Application of Liquid Sprays.