Appendix C. Analyses of sensitivity distributions for estimation of acute hazard concentrations to aquatic animals

Matthew Etterson
US Environmental Protection Agency
National Health and Environmental Effects Research Laboratory
Mid-Continent Ecology Division
Duluth, MN

20 December 2011

Contents

1. Introduction

Species sensitivity distributions are a common tool used for setting safe limits on chemical concentrations in surface waters (Posthuma et al. 2002, Suter 2002, Chapman et al. 2007, TenBrook et al. 2010). Although the analysis and interpretation of species sensitivity distributions varies widely, the basic methodology is quite general and can be summarized as a three-step procedure. First, results from separate toxicity tests on a given chemical using various aquatic animal species are compiled. Second, a statistical distribution to which the test results are thought to conform is chosen and fit to the data. Third, the fitted distribution is used to infer a concentration that will be protective of a desired proportion of species in an hypothetical aquatic community.

The procedure described above necessarily relies upon policy decisions, for example, concerning the proportion of species that should be protected and the necessary level of confidence with which the protective concentration is identified. This report focuses on fitting statistical distributions to the results of toxicity tests. As such its content concerns the three steps laid out in the previous paragraph, each of which can be accomplished in different ways, with different results. This report does not consider the underlying policy decisions required for application of species sensitivity distributions to regulatory decision making. The following subsections give some historical background regarding data compilation, distribution fitting, and inference from fitted sensitivity distributions, in particular as they relate to the development of a common risk assessment methodology for use by USEPAs Office of Pesticide Programs (OPP)and Office of Water (OW).

The focus of analyses in this report is on fitting SSDs to chemicals for which the minimum data are available to fulfill FIFRA requirements; however, insufficient data are available to derive water quality criteria according to the methods described in the 1985 Guidelines. The inferential endpoint for all analyses is taken to be the fifth percentile of the fitted SSD, or HC_5. However, when data are limited, HC_5 estimates may be subject to varying degrees of uncertainty. Therefore, the focus of these analyses is on the bias and uncertainty of using SSDs to derive HC_5 values at varying sample sizes.

1.1. Data

For this report, data are assumed to be the results of acute toxicity tests using aquatic animals. These are EC50s and LC50s from 96-hour continuous exposures (48 hrs for Cladocerans and a few other species; USEPA 1985, ASTM 2011). An important consideration in data compilation is the distribution of available data across species. Strictly speaking, an assumption of all the methods considered below is that the data (test results) pertain to a random sample of species from the community for which the analysis is intended to apply. Obviously this assumption is violated, in part due to constraints on our knowledge of husbandry of aquatic animals. Thus a relatively limited subset of aquatic animals makes up the greater part of all aquatic toxicity tests. Nevertheless, several important taxonomic requirements are commonly applied to the compilation and analysis of multi-taxon toxicity test results. For this report, two data subsets are of particular interest, the eight minimum data requirements (MDRs) of the 1985 Guidelines (USEPA 1985) and the three typical species used for OPP benchmark analysis. For freshwater tests the latter are rainbow trout (*Oncorhynchus mykiss*), bluegill sunfish (*Lepomis*

macrochirus) and *Daphnia magna*. For saltwater tests the three typical OPP benchmark species are sheepshead minnow (*Cyprinodon variegatus*), mysid shrimp (*Americamysis bahia*), and Eastern oyster (*Crassostrea virginica*).

1.2. Distribution fitting

Many statistical distributions have been used for fitting SSDs (*e.g*, log-normal, log-logistic, $Burr_{III}$, etc.); however, several analyses have shown that no one distribution is preferred across datasets (Newman et al. 2000, Zajdlik & Associates 2005, Chapman et al. 2007). Deciding which statistical distribution to fit to a set of data has been described as one of the most important and difficult choices in the use of species sensitivity distributions (Chapman et al. 2007). Two important decisions must be made when fitting a distribution to empirical data. The first concerns how the distribution will be fit to the data, which in this report is equivalent to the problem of parameter estimation. The second choice concerns how to assess the quality or accuracy of the fitted distribution as a general representation of the data, or goodness-of-fit. A related concern involves deciding how to choose among the fitted distributions when multiple distributions are fit to the same data. Also, for many types of distributions (*e.g.,* normal, logistic, triangular) data are transformed prior to analysis, most frequently using the common log (log_{10}) transformation. This complicates comparisons between distributions fit to transformed versus untransformed data.

Three methods commonly used for estimating the parameters of SSDs are considered below. These are maximum likelihood, moment estimators, and graphical methods (these methods are described in more detail in Methods subsections 2.2.3, 2.2.4, and 2.2.5 below). It should be noted that not all methods can be used with all distributions. It is likely that some methods will perform better than others when data are limited and when the chosen statistical distribution is a poor approximation of the underlying data distribution. Further, it is likely that rank of performance will change under different data conditions, perhaps in predictable ways.

Newman et al. (2000) recommended a non-parametric method for fitting SSDs using empirical bootstrapping (Section 2.4.2 below). However, given our current interest in the fifth percentile of acute values (see Section 1.3, below), bootstrap estimation does not seem feasible because it would require at least 20 data points to estimate the fifth percentile of the empirical distribution function. Bootstrap methods are used below to test goodness-of-fit and to estimate sampling variance.

Both Chapman et al. (2007) and TenBrook et al. (2009) emphasize the importance of visual inspection of fitted distributions against the empirical data to which they are fit. Such methods are omitted from this report because the focus of this exploration is on general patterns of fit that emerge from fitting thousands of replicate data sets. However, visual inspection of data and fit would be an important step in any particular application of SSDs to empirical data.

1.2.1. Parameter estimation

Maximum likelihood methods for SSDs were first tested by Kooijman (1987) for the logistic distribution who reported substantial bias in estimation of the scale parameter (β) with sample sizes ≤ 5. Shao (2000) described maximum likelihood estimators for the $Burr_{III}$ distribution and these estimators are implemented by the software BurrliOZ (Campbell et al. 2000). Several recent reports on the application and analysis of species sensitivity distributions have employed maximum likelihood with

other distributions (Zajdlik and Associates 2005, Chapman et al. 2007), often citing the first of the following desirable properties. First, when data fit the assumed distribution, maximum likelihood parameter estimates are the most efficient parameter estimates possible (*i.e.*, the estimates with the least sampling variance, Edwards 1992). Second, the use of maximum likelihood allows the fit of different distributions to be compared using information theoretic methods for comparing models (Burnham and Anderson 2002). Third, use of maximum likelihood allows model-averaging of estimated quantiles (Burnham and Anderson 2002). Fourth, maximum likelihood and restricted maximum likelihood (Harville 1977), allow specification of hierarchical models that may otherwise be very difficult to fit. Fifth, the development of likelihood-based estimators provides a link towards the development of Bayesian methods for SSDs. Finally, maximum likelihood methods offer a generalized means for extending the advantages of the graphical methods (below) to any continuous distribution. The latter three points are the subject of proposed future work (Section 5).

Moment estimators are a common method for fitting SSDs (Kooijman 1987, Van Straalen and Deneman 1989). In practice, they work by equating the mean and variance of a sample to the parametric mean and variance of a chosen distribution, which are functions of the parameters of that distribution. This creates two equations in two unknowns, which can then be solved for the unknown parameters. The resulting solution is an estimate of the parameters of the distribution expressed as functions of the sample mean and sample variance. Although this procedure has been described in terms of the mean and variance (the first two moments), it could be extended to higher moments as well if a distribution (*e.g.*, Burr) has more than two parameters.

Graphical methods for use in SSDs were described by Erickson and Stephan (1988). They are a subset of the general theory of order statistics (Arnold et al. 2008) and have two unique attributes that make them very attractive for use in fitting SSDs (TenBrook et al. 2008). First, once data have been ordered and the empirical percentiles obtained, the linear estimation model can be weighted toward the lower tail of the distribution, which is generally the portion of the distribution of interest for regulation and risk assessment. This can alleviate potential biases resulting from right-skewed toxicity distributions. Second and related to the first, test results that are right-censored (known only to be greater than the highest tested concentration) can often be accommodated (Erickson and Stephan 1988). Graphical methods can be used on any distribution in the location/scale family. Of the seven distributions described below (Section 2.2.13), the normal, logistic, and triangular are location/scale distributions. Thursby et al. (Appendix F) show how similar linearization methods can be extended to distributions outside the location/scale family. Section 5 (future work) contains descriptions of some potential generalizations of ordered methods that would extend their applicability to any continuous distribution.

1.2.2. Goodness-of-fit

Goodness-of-fit is a measure of how well an assumed distribution fits a set of data, given the data and the values of the estimated distributional parameters. Goodness-of-fit tests are applied after a distribution has been fit to a set of data to evaluate the extent to which the fitted distribution is a good representation of the observed data. Tests for goodness-of-fit can be divided into parametric and non-parametric methods. In either case, the test begins with the definition of a test statistic that can be reliably predicted to increase in magnitude with lack of fit (e.g., the summed discrepancy between the

empirical distribution function and the cumulative distribution function, evaluated at all data points, is used below). With parametric goodness-of-fit tests, a theoretical distribution for the test statistic can be derived, and probabilities are estimated from that theoretical distribution. In general, the derivation of the theoretical distribution of the test statistic depends on asymptotic convergence and upon the hypothesized distribution for the data. Therefore parametric goodness-of-fit tests tend not to work well at small sample sizes and generally apply to only one distribution (often the normal distribution). In contrast, non-parametric methods often work by statistical resampling methods (Efron and Tibshirani 1994) and probabilities are assessed as simple ranks of observed statistics among a set of simulated statistics. They can be applied to any continuous distribution and are valid regardless of sample size. However, both parametric and non-parametric methods lack power at small sample sizes.

Luttik and Aldenberg (1997), Aldenberg and Luttik (2002), and Newman et al. (2000) all considered parametric goodness-of-fit tests. Zajdlik & Associates (2005) recommended the Anderson-Darling test for all distributions, except the normal and log-normal distributions for which they recommended the Shapiro-Wilks test. Chapman et al. (2007) carried out extensive simulations of the power properties of goodness-of-fit tests for the normal distribution and concluded that power to detect non-normality (lack of fit) was extremely low, especially at sample sizes < 20, which correspond to both cases of particular interest in this document: $n = 3$ and $n = 8$).

Newman et al. (2000) and Shao (2000) employed non-parametric goodness-of-fit tests based on empirical bootstrap sampling (Efron and Tibshirani 1994, Manly 1997). Chapman et al. (2007) describe a parametric bootstrap procedure (also described by Efron and Tibshirani 1994), but do not apply it to goodness-of-fit testing for SSDs. Because of their utility at all sample sizes, and applicability to all continuous distributions, only bootstrap methods are considered below.

1.3. Inference from the fitted distribution

All analyses of parametric species sensitivity distributions begin by estimating the parameters of the distribution (see above). Thus the distributional parameters are a universal inferential endpoint (excluding non-parametric SSDs, Newman et al. 2000). Once the parameters are estimated, a given percentile (p) of the distribution is often chosen to represent the concentration at which (no more than) p% of species will be at risk of adverse effects, referred to herein as the HC_p. Regardless of how the distribution is fit, an HC_p is easily estimated using the quantile function for the fitted distribution. However, estimates of percentiles are subject to bias (if the distribution doesn't fit the data very well) and uncertainty (especially when the number of test results are limited). Methods for handling these aspects of distribution fitting vary widely in the SSD literature. Erickson and Stephan (1988) also pointed out that, by Jensen's inequality, an unbiased estimator for the HC_p, might be a biased estimator of the percentile (intended to be p) of species protected at the estimated HC_p if the quantile function is non-linear in p (as is generally the case).

1.3.1. Sampling variance

When distributional parameters are estimated from empirical data, the estimated parameters are subject to sampling variance. In other words, given data that conform to a specified distribution, if equal sized (but different) sets of data are drawn from the same distribution, the parameter estimates will differ with each set of data, resulting in a distribution of parameter estimates. The variances of the

distributions of parameter estimates are termed the sampling variances of the parameters. If the estimation procedure is unbiased the average parameter estimates will be arbitrarily close to the 'true' values if the procedure is repeated a sufficient number of times. However, the expected variance in these parameter estimates may be quite large and is generally inversely related to the size of the sample. This sampling variance is present in all three fitting techniques described above (maximum likelihood, moment estimators and graphical methods), though it may differ among techniques. Sampling variance of parameter estimates translates directly (though not necessarily linearly) into sampling variance around quantiles of a distribution (*i.e.*, the estimated HC_5). Common methods for estimating sampling variance around quantiles in an SSD include the delta method (Seber 1982, Shao 2000) and the bootstrap (Newman et al. 2000).

1.3.2. Confidence limits

Confidence limits for an estimated percentile are an alternative expression of sampling variance in the model parameters. These can be one-sided expressions of confidence that the true HC_p is greater than a specified concentration (Kooijman 1987, van Straalen and Deneman 1989, Aldenberg and Slob 1993) or two-sided limits related to the probability that the region defined by a lower and upper bound would contain the true HC_p (Shao 2000, Newman et al. 2000). Many methods for confidence interval development rely on either asymptotic properties of estimators or standard (mean zero and variance 1) forms of distributions. Only one-sided confidence intervals generated using extrapolation constants derived from standard forms of the logistic, normal, and triangular distributions (Aldenberg and Slob 1993, Aldenberg and Jaworska 2000, Pennington 2003) are considered below.

1.3.3. Model selection & multidistributional inference

Many researchers have discussed the important (and difficult) choice of which distribution to employ for an SSD (Newman et al. 2000, Zajdlik & Associates 2005, Chapman et al. 2007). Assessing the goodness-of-fit of a distribution provides only limited information for comparing distributions because discrepancies of fit will generally decrease monotonically with increasing number of estimated parameters. Yet an over-parameterized model may have poor predictive ability. Formal model selection criteria impose a penalty for each estimated parameter, which creates a tradeoff between parsimony and fit. The tendency for species sensitivity distributions to have two estimated parameters (though the $Burr_{III}$ distribution has three) alleviates this concern somewhat. However, model selection methods are also useful for ranking the performance of alternative distributions and for formally averaging model predictions when multiple models are fit (Burnham and Anderson 2002). Elphick (2011) used AIC to compare several candidate distributions, including log-normal, log-logistic, log-gumbel and Weibull and report similar performance.

1.4. Objectives

The primary objective of the work described in this report is to compare the performance of different distributions and estimation techniques commonly used in SSD analyses. In particular, performance is evaluated as the ability to accurately estimate the HC_p (in particular the HC_5). Accuracy is assumed to incorporate both bias and sampling variance. While a formal definition of accuracy specifies that accuracy is the sum of sampling variance and squared bias (Williams et al. 2002), the term is used more loosely here to indicate that increases in *either* bias *or* sampling variance of an estimate of

the HC_5 will decrease the accuracy of that estimate. The work described below can be divided into two parts. First, the performance of different estimation methods at different sample sizes is evaluated when the data are derived from a known distribution using simulated data with a known HC_5 (Statistical comparison 2.3.1). Second, seven distributions are fit using three fitting methods to empirical toxicity data on acetylcholinesterase inhibitors (Statistical comparisons 2.3.2 – 2.3.6) for which the true distribution and HC_5 are unknown. Presentation of the results of these analyses is organized around specific questions that will help to inform the use of species sensitivity distributions in the context of work done by the Office of Water and the Office of Pesticide Programs. These are:

1. How does sample size influence accuracy (bias and variance) of the estimated HC_5?
2. How do minimum data requirements influence accuracy of the estimated HC_5?
3. How do different estimation methods influence accuracy of the estimated HC_5?
4. Are measures of fit (goodness-of-fit, Akaike's Information Criterion) reliable indicators of performance?
5. What are actual coverage rates of 95% confidence limits for subsamples?
6. How do different candidate distributions perform relative to each other?
7. Does model-averaging across distributions improve estimates of the HC_5?
8. What future research is suggested by these results?

Specific comparisons and analyses designed to answer these questions are described in detail below.

2. Methods

2.1. Data & statistical comparisons

2.1.1. Simulation data

To evaluate the performance of distributions and fitting methods under known conditions, data were simulated from known parametric forms of seven distributions. Simulated sample sizes were 3, 8, 20, 50, and 100 mean acute values. Each simulated data set was replicated 10,000 times. Approximate parameter values for generating all simulated data are presented in Table 1.

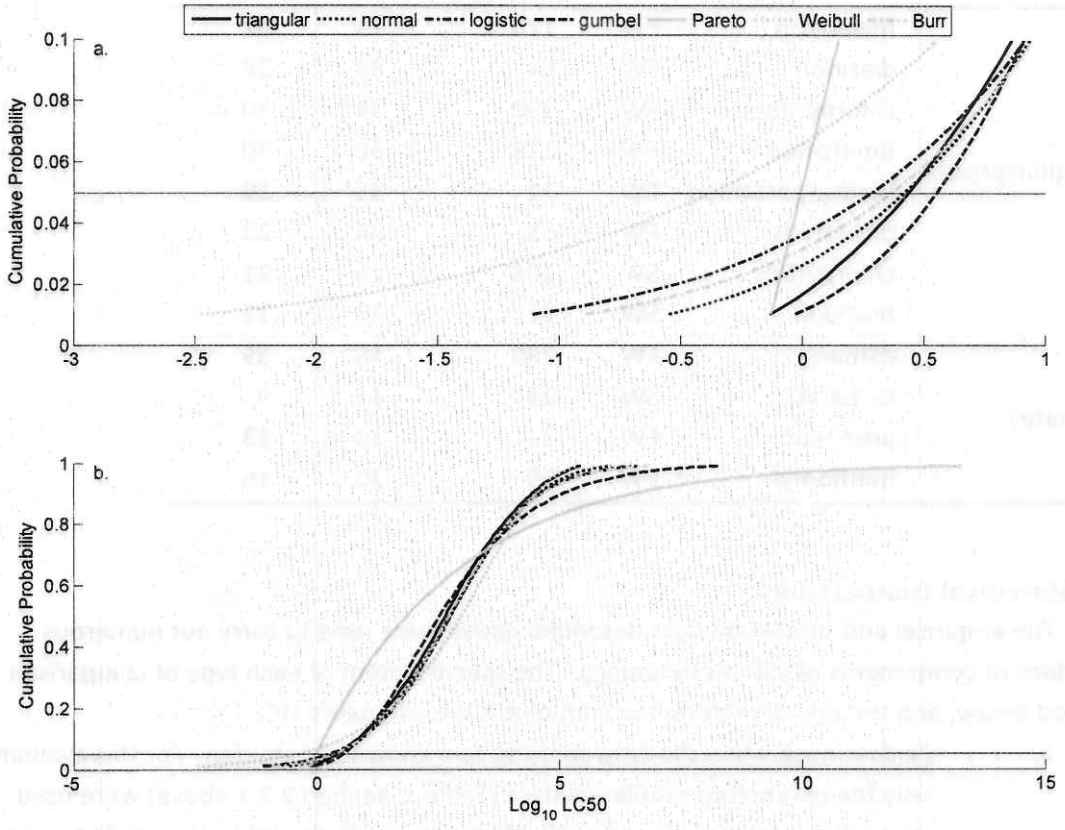
Table 1. Approximate parameters used to simulate acute toxicity data

Distribution	Parameters	Fifth percentile
normal	$\mu = 2.77, \sigma = 1.42$	2.65
logistic	$\alpha = 2.78, \beta = 0.85$	1.94
triangular	$a = -0.59, b = 5.92$	2.75
Gumbel	$\mu = 2.05, \beta = 1.36$	3.61
Weibull	$\lambda = 2940.47, k = 0.34$	0.43
Pareto	$x_m = 0.70, \alpha = 0.15$	0.99
Burr	$a = 334.71, b = 0.48, c = 1.20$	2.35

The first three distributions (normal, logistic, and triangular) are probably the most common distributions used in SSDs and have a familiar sigmoid cumulative distribution function (Fig. 1). The

Gumbel distribution is similar to the previous three, but slightly more right-skewed (Fig. 1). The Weibull and Pareto distributions were chosen to represent 'outlier' distributions that might be expected a priori to perform poorly. The Weibull is left-skewed relative to the other distributions, and the Pareto is shaped quite differently, with a monotonically declining slope to its cumulative distribution function (Fig. 1). Finally, the Burr distribution is a three parameter distribution with a sigmoid cumulative distribution function and is the primary distribution behind the BurrliOZ software (Campbell et al. 2000)

Figure 1. Plots of the cumulative distribution functions for seven distributions using the parameters from Table 1, a. in the region of the HC_5, and b. over the full distribution. Note the $\log_{10}$ scale of the horizontal axis.



2.1.2. Empirical data

Acute toxicity test results for seven organophosphate insecticides and three carbamates were obtained from Web-ICE (Raimondo et al. 2010). These chemicals were selected because their data sets fulfilled the data requirements defined under FIFRA and the 1985 Guidelines. Results are separated by freshwater versus saltwater animals. Pesticide names and sample sizes are given in Table 2. For empirical analyses, these data can be summarized at two levels, the species and the genus. When more than one toxicity test result is available for a given species, the datum for that species is taken to be the geometric mean of all such results (hereafter the species mean acute value or SMAV). When analysis is performed on genera, and more than one SMAV exists for a given genus, then the datum for that genus is taken to be the geometric mean of all such SMAVs (hereafter the genus mean acute value, or GMAV).

Summarizing the test results by genus has the effect of reducing the influence of closely related test species on the estimated HC_5. These methods for summarizing taxonomic replication have a strong historical background in the development of U.S. aquatic life water quality criteria (ALC) (USEPA 1985). However, they ignore some information when taxonomic replication is available. Alternative methods for handling replicate test results for a given taxon are suggested in Section 5 (Future analyses).

It should be noted that in separate analyses with these data, Erickson (Appendix D) has found some errors and inconsistencies. Thus, these analyses should be considered preliminary pending additional data quality review.

Table 2. Acute test results available for analysis

Class	Chemical	Water	Test Results	SMAVs	GMAVs
Organophosphates	malathion	FW	174	71	57
	diazinon	FW	84	33	28
	chlorpyrifos	FW	100	34	30
	fenitrothion	FW	113	36	30
	methyl parathion	FW	71	35	29
	dichlorvos	FW	55	25	22
	chlorpyrifos	SW	108	26	21
	fenthion	SW	22	13	12
Carbamates	carbaryl	FW	290	75	55
	carbaryl	SW	23	11	9
	propoxur	FW	21	13	13
	methomyl	FW	56	20	16

2.1.3. Statistical comparisons

The empirical and simulation data described above were used to carry out numerous evaluations of components of SSD methodology. The specific intent of each type of comparison is described below, and the primary inferential endpoint is the estimated HC_5.

1. Performance when the data conform to a known distribution. For this evaluation, only the parametric simulation data (Table 1, Section 2.2.1 above) were used. All seven distributions using all available fitting methods (Table 3) were fit to every replicate sample of simulated data.
2. Baseline empirical performance for the maximal sets of SMAVs and GMAVs. For this evaluation the empirical data (Table 2) were used. All seven distributions using all available fitting methods (Table 3) were fit to every available combination of pesticide and water type. The resulting HC_5 estimates are the reference HC_5s against which the HC_5 estimates generated under comparisons 3-6 (immediately below) are compared.
3. Comparative performance at $n = 8$ without regard to ALC MDRs. For this evaluation 1,000 subsamples of 8 GMAVs from each empirical data set (Table 2) were used.

4. Comparative performance at n = 8 when ALC MDRs are met. This evaluation was similar to (3) except that the 1,000 GMAVs were chosen so that the eight minimum data requirements (MDRs) of the 1985 guidelines (USEPA 1985) were always satisfied.
5. Comparative performance at n = 3 SMAVs without regard to OPP benchmark species. This evaluation was similar to (3) except that the 1,000 subsamples were of size n = 3.
6. Comparative performance at n = 3 for the OPP benchmark species. This evaluation was similar to (5) except that there is only one subset of data conprising the three benchmark SMAVs.

Table 3. Distributions and fitting methods used for the analyses below.

Distribution	Maximum Likelihood	Moment Estimators	Graphical Estimators
normal	yes	yes	yes
logistic	yes	yes	yes
triangular	yes	yes	yes
Gumbel	yes	yes	no
Weibull	yes	no	no
Pareto	yes	yes	no
$Burr_{III}$	yes	no	no

2.2. Statistical methods

This section is provided for completeness and to allow interested readers to evaluate the statistical methods. However, the results section has been written to be read and understood without relying on this material.

2.2.1. Random variates

Random variates were generated from a specific distribution (with parameters given in Table 1) by first drawing a vector of uniform random variates of the desired sample size. The uniform random variates were then used as input to the quantile function for the distribution together with the specified parameter values. The resulting output is a vector of random variates from the specified distribution. Quantile functions for all distributions used are presented in Section 2.2.15, below.

2.2.2. The HC_5

For the purposes of this report, the inferential endpoint for all analyses is the HC_5. Regardless of sample size and method of estimation, the HC_5 is estimated as the output of the quantile function for the specified distribution and estimated parameters evaluated at 0.05. Quantile functions for all distributions used are presented in Section 2.2.15, below.

2.2.3. Maximum likelihood estimation

Log-likelihood equations were developed for each of the seven distributions as the natural logarithm of the probability density function (f) for the distribution. The resulting log-likelihood was summed over all data points:

$$L(\boldsymbol{\theta}|\mathbf{X}) \propto \sum_{i=1}^{n} \ln\left(f(x_i|\boldsymbol{\theta})\right)$$

Maximum likelihood estimates (MLEs) of parameters were obtained by numerically searching for the parameter values that maximized the log-likelihood over all data points (equivalently minimized the negative log-likelihood). MLEs were always estimated by numerical search, even when closed-form solutions to the MLEs exist. This was done to subject MLEs for all distributions to the same (albeit minor) uncertainty associated with numerical estimation. Likelihood expressions are presented for each distribution in Section 2.2.15.

2.2.4. Moment estimation

Moment estimators were derived, wherever possible, by setting the expected distributional mean and variance equal to the sample mean and variance, and solving for the distributional parameters. For example, let $\bar{x}$ and s^2 represent the sample mean and variance, respectively (regardless of distribution). The mean and variance of a logistic distribution are α and $\frac{\pi^2}{3}\beta^2$. Setting $\alpha = \bar{x}$ and $s^2 = \frac{\pi^2}{3}\beta^2$ and solving for α and β results in the two moment estimators $\hat{\alpha} = \bar{x}$ and $\hat{\beta} = \frac{s}{\pi}\sqrt{3}$, where the circumflex over the parameter symbols indicates that they are estimated quantities. For the Weibull distribution, moment estimators could not be derived because the equations for the mean and variance contain gamma functions of the Weibull parameters. For the Burr distribution, moment estimators were not derived because the Burr distribution has three parameters, requiring three equations, but an equation for the third moment of the Burr distribution was not immediately available. Moment estimators for the remaining five distributions (normal, logistic, triangular, Gumbel, and Pareto) are presented in Section 2.2.15, below.

2.2.5. Graphical estimation

Graphical estimation works by exploiting a functional relationship between the parametric quantiles of a standard form of a distribution (parameters chosen so that the distribution has mean 0 and variance 1) to the quantiles of the empirical distribution function. This allows the mean and variance of the sample to be estimated, from which moment estimators may be used, in turn, to estimate the distributional parameters. This procedure is described in detail by Erickson and Stephan (1988). This procedure is only applied to distributions in the location-scale family, which includes the normal, logistic, and triangular distributions among the candidate distributions considered in this report. Note that, Thursby et al. (Appendix F) apply related linearization methods to some distributions outside the location/scale family. Parameters for standard forms of normal, logistic, and triangular distributions are presented in Section 2.2.15, below. For this report, all data points were weighted equally to minimize the number of possible explanations for differences in performance between graphical estimates and the other two methods (maximum likelihood and moment estimators).

2.2.6. Transformations

When the normal, logistic, triangular, or Gumbel distribution was used, the data were first common-log transformed ($\log_{10}$). When the Weibull, Pareto, or Burr distribution was used, the data were untransformed. This complicates comparisons among distributions, especially using maximum likelihood and AIC. To solve this problem, the likelihoods for the normal, logistic, triangular, and Gumbel distributions were reformulated as follows. First, let:

$$y = \log_{10}(x).$$

Therefore the cumulative distribution functions for the four distributions using $\log_{10}$-transformed data are of the form: $F(y|\boldsymbol{\theta})$. Thus, the probability density functions for the untransformed data (x) can be calculated using the product rule.

$$f(x|\boldsymbol{\theta}) = \frac{d}{dx}F(y|\boldsymbol{\theta}) = \frac{d}{dy}F(y|\boldsymbol{\theta})\frac{dy}{dx} = f(y|\boldsymbol{\theta})\frac{1}{x\ln(10)}$$

In the above equation, the expressions $f(y|\boldsymbol{\theta})$ are the probability densities for the $\log_{10}$-transformed data for the respective distributions. When the likelihood is maximized over the transformed data the transformation factor $\frac{1}{x\ln(10)}$ can be ignored for the purposes of obtaining the MLEs because the transformation factor does not contain the parameters of interest. However, to compare distributions using AIC the factor must be included so that the AIC statistics are on the same scale. Thus, for distributions on $\log_{10}$-transformed data:

$$L(\boldsymbol{\theta}|\mathbf{X}) \propto \sum_{i=1}^{n} \ln\left(\frac{1}{x_i \ln(10)} f(y|\boldsymbol{\theta})\right) = L(\boldsymbol{\theta}|\mathbf{Y}) - \ln(10)\sum_{i=1}^{n} y_i - n\ln(\ln(10))$$

2.2.7. Bias

Bias was estimated as the expected value of the difference between the estimated HC_5 and the known value. Specifically, bias was estimated as:

$$\text{bias}(\hat{HC}_5) = E(\hat{HC}_5 - HC_5)$$

In the above equation, the notation $\hat{HC}_5$ and HC_5 denote the estimated and known values of the HC_5, respectively. These were calculated differently depending on whether the distributions were applied to transformed versus untransformed data, respectively. When a distribution was fit to untransformed data (Weibull, Pareto, Burr), the known HC_5 is:

$$HC_5 = F^{-1}(0.05|\boldsymbol{\theta})$$

Above, F^{-1} is the quantile function for the distribution and $\boldsymbol{\theta}$ is a vector of known parameters. When a distribution is fit to transformed data (normal, logistic, triangular, Gumbel), the known HC_5 is:

$$HC_5 = 10^{F^{-1}(0.05|\boldsymbol{\theta})}$$

Similarly, when a distribution was fit to untransformed data the expected value of the HC_5 was calculated as the simple average of the fifth percentiles across all iterations of the simulation:

$$E\left(\hat{HC}_5\right) = \frac{1}{10,000}\sum_{r=1}^{10,000} F^{-1}\left(0.05 \mid \hat{\boldsymbol{\theta}}_r\right) = E\left(F^{-1}\left(0.05 \mid \hat{\boldsymbol{\theta}}_r\right)\right)$$

However, when a distribution on transformed data was fit, the resulting set of estimated values of the HC_5 is non-normally distributed and the arithmetic mean of the estimated HC_5s is a biased estimator of the expected HC_5. In this case the expected HC_5 was estimated as the geometric mean of HC_5 estimates, or:

$$E\left(\hat{HC}_5\right) = \sqrt[10,000]{\prod_{r=1}^{10,000} 10^{F^{-1}\left(0.05 \mid \hat{\boldsymbol{\theta}}_r\right)}} = \left(10^{\sum_{r=1}^{10,000} F^{-1}\left(0.05 \mid \hat{\boldsymbol{\theta}}_r\right)}\right)^{\frac{1}{10,000}} = 10^{E\left(F^{-1}\left(0.05 \mid \hat{\boldsymbol{\theta}}_r\right)\right)}$$

When discussing bias in the results the term bias is used in two ways, referring both to bias and its absolute magnitude (i.e., |bias|), where the meaning is clear from the context.

Bias cannot be assessed for analyses of empirical data because the true HC_5s are not known. Yet it is still of interest to compare the mean performance of the estimators when data requirements are imposed and when they are not. To do so, an HC_5 was estimated for each chemical using 'all available data'. To examine the effect of the ALC MDRs 'all available data' were taken to be the set of all GMAVs for a given chemical. To examine the effect of the typical three OPP species 'all available data' were taken to be the set of all SMAVs for a given chemical. Thus all calculations proceeded as described above, but with the HC_5 calculated using all available data substituted for the known HC_5.

For easier interpretation, bias is not presented as defined above. Rather it is presented as the ratio of the estimated HC_5 to the known HC_5: $E\left(\hat{HC}_5\right) / HC_5$. In figures, this is further presented on the $\log_{10}$ scale, in which case the y-axis tick marks represent over- or under-estimation by each successive order of magnitude. In tables, the untransformed ratio is given.

2.2.8. Bootstrap sampling

Bootstrap resampling was used to assess goodness-of-fit and sampling variance (next two sections). With both empirical and parametric bootstrap sampling the methodology depends on generating replicate sets of data based on the reference data being analyzed (below, for demonstration purposes, 100 replicates were used). The reference data is the actual data set being analyzed and in this report it may be the simulation data (Table 1, Section 2.1.1 above) or it may be a particular set of empirical data (Table 2, Section 2.1.2). The process begins after a particular distribution is fitted to the reference data. The new sets of data are the same size as the original data and are generated by drawing (with replacement) new samples from the original data (empirical bootstrap) or by drawing from the fitted distribution (parametric bootstrap). To these new data sets, the same distribution is fit, generating a distribution of fits under the hypothesis that the data fit the specified distribution with the estimated parameter values for the original data.

2.2.9. Goodness-of-fit

Goodness-of-fit was assessed using both empirical and parametric bootstrap resampling (Section 2.2.8. above). The summed (over all data points) squared distance between the empirical

distribution function (EDF) and the cumulative distribution function (CDF) for the fitted distribution was used as a summary measure of fit. Large values of this statistic indicate poorer fit, whereas small values indicate better fit. Using the bootstrapping procedures 100 values of the fit statistic were generated to which the observed fit statistic from the reference data was compared. The proportion of these bootstrapped fit statistics that were greater than or equal to the observed fit statistic for the reference data was interpreted as the P-value for lack of fit (small P-values indicate significant lack of fit). This level of replication (100 iterations) will give poor resolution on the P-value for any single data set, but will still allow precise estimation of average behavior of the goodness-of-fit algorithm over 10,000 or 1,000 samples of simulation or empirical data, respectively.

2.2.10. Sampling variance

Sampling variance for distributional parameters and HC_5 estimates was estimated using both empirical and parametric bootstrap methods to generate 100 values of each statistic. Sampling variance for a given estimated value was estimated as the variance among the 100 bootstrap samples for the same statistic.

2.2.11. Confidence limits

One-sided 95% confidence limits for the HC_5 for the logistic, normal, and triangular distributions were generated using extrapolation constants provided by Aldenberg and Slob (1993), Aldenberg and Jaworska (2000) and Pennington (2003). These are intended to be expressions of 95% confidence that the true HC_5 is greater than the calculated confidence limit. The formula for generating the confidence limit is:

$$L = \overline{x} - ks$$

In the above equation, L is the lower 95% confidence limit, $\overline{x}$ is the sample mean, s is the sample standard deviation, and k is an extrapolation constant that is specific to both the distribution fitted (logistic, normal, or triangular) and sample-size. These confidence limits were generated only for $n = 3$ and $n = 8$.

2.2.12. Model selection

Whenever multiple distributions were fit using maximum likelihood to the same set of data (whether simulated or empirical) Akaike's Information Criterion (with correction for sample size: AIC_c) was used as a relative measure of the performance of distributions. The formula for AIC_c is:

$$\mathrm{AIC_c} = -2L + 2K\left(\frac{n}{n-K-1}\right)$$

In the above equation, L is the maximized log-likelihood function, n is the sample size, and K is the number of parameters estimated in fitting the distribution. The second term on the right hand side of the above equation is a penalty term. It increases the AIC_c statistic with each additional parameter estimated. Because the quotient within the parentheses is zero or negative whenever $n \leq K + 1$, AIC_c cannot be applied to such cases. Unfortunately this includes all cases when $n = 3$, corresponding to the comparisons of analyses of random sets of 3 SMAVS to those of the typical OPP species.

2.2.13. Model averaging

Model-averaged HC_5 values were calculated as weighted averages of the HC_5 values from each individual distribution fit to the same data set. The weights used were Akaike weights (w_i), as defined by Burnham and Anderson (2002:75). The formula for Akaike weights is:

$$w_i = \frac{\exp\left(-\frac{1}{2}\Delta_i\right)}{\sum_{j=1}^{7}\exp\left(-\frac{1}{2}\Delta_j\right)}$$

In the above equation, $\Delta_i = \mathrm{AIC_c}\left(\text{distribution i}\right) - \min\left(\mathrm{AIC_c}\right)$ and the summation limit of 7 indicates that seven distributional fits were compared. Model-averaged estimates of the HC_5 were calculated as:

$$\overline{\mathrm{HC}_5} = \sum_{j=1}^{7} w_j \mathrm{HC}_{5j}$$

In the above equation, the HC_{5j} is the estimate of the HC_5 from the j^{th} distribution considered.

Sampling variance of the $\overline{\mathrm{HC}_5}$ was estimated using equation 4.9 of Burnham and Anderson (2002:162):

$$\mathrm{var}\left(\overline{\mathrm{HC}_5}\right) = \sum_{j=1}^{7} w_i \sqrt{\mathrm{var}\left(\mathrm{HC}_{5j}\right) + \left(\mathrm{HC}_{5j} - \overline{\mathrm{HC}_5}\right)^2}$$

2.2.14. Software & distributions

All analyses described below were accomplished using subroutines written in Matlab (hereafter, called "the SSD toolbox"). Use of the SSD toolbox requires basic Matlab software and the Matlab Optimization toolbox (Mathworks 2010). The SSD toolbox contains functions for fitting seven distributions (normal, logistic, triangular, Gumbel, Weibull, Pareto, and $\mathrm{Burr_{III}}$) using three methods (maximum likelihood, moment estimation, and graphical estimation), though not all methods can be used with each distribution (Table 3). The SSD toolbox also contains subroutines for performing empirical and parametric bootstrap sampling for testing goodness-of-fit, estimating sampling variance, and performing information-theoretic model selection and multi-model inference. Table 4 gives some standard statistical notation used in describing the distributions. Table 5 gives a list of distribution functions available in the SSD toolbox.

Table 4. Statistical notation for the description of distributions tested as candidates for use in estimating the HC_p

symbol	description
n	sample size
$\bar{x}$	sample mean: $\frac{1}{n}\sum_{i=1}^{n} x_i$
s	sample standard deviation: $\frac{1}{n-1}\sum_{i=1}^{N}(x_i-\bar{x})^2$
$\exp(x)$	exponential function (e^x)
$\mathbf{X}$	column-vector of untransformed data (mean acute values)
$\mathbf{Y}$	column-vector of $\log_{10}$-transformed data (mean acute values)
$\boldsymbol{\theta}$	column-vector of parameters for any given distribution
$f(x\|\boldsymbol{\theta})$	probability density at x (or y) conditional on $\boldsymbol{\theta}$
$F(x\|\boldsymbol{\theta})$	cumulative distribution function at x (or y) conditional on $\boldsymbol{\theta}$
$F^{-1}(x\|\boldsymbol{\theta})$	quantile function at x (or y) conditional on $\boldsymbol{\theta}$
$L(\boldsymbol{\theta}\|\mathbf{X})$ or L	log-likelihood for $\boldsymbol{\theta}$ conditional on $\mathbf{X}$ (or $\mathbf{Y}$)

Table 5. Names of functions in the SSD toolbox

distribution	pdf	cdf	quantile	likelihood	moments	random variates
normal	[1]normpdf	[1]normcdf	[1]norminv	normlik	normmom	normrnd
logistic	logipdf	logicdf	logiinv	logilik	logimom	logirnd
[2]triangular	triapdf	triacdf	triainv	trialik	trimom	triarnd
Gumbel	gumpdf	gumcdf	guminv	gumlik	gummom	gumrnd
Pareto	parpdf	parcdf	parinv	parlik	parmom	parrnd
[3]Weibull	wblpdf	wblcdf	wblinv	wbllik	n/a	wblrnd
Burr	burpdf	burcdf	burinv	burlik	n/a	burrnd

[1]functions included in standard Matlab (pdf = probability density function, cdf = cumulative distribution function).

[2]Triangular functions written for the symmetric triangular distribution only

[3]Weibull functions are also available in Matlab's statistics toolbox

n/a = not applicable

2.2.15. Equations for Distributions

Normal distribution

Parameters: $\boldsymbol{\theta} = [\mu; \sigma]$

μ (location)

σ (scale)

Transformation: $y_i = \log_{10}(x_i)$

$$f(y \mid \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right)$$

$$F(y \mid \mu, \sigma) = \int_{-\infty}^{y} f(z)dz$$

$$F^{-1}(p) = \Phi(p)$$

$$L(\boldsymbol{\theta} \mid \mathbf{Y}) = -\frac{n}{2}\ln(2\pi) - n\ln(\sigma) + \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mu)^2$$

$$\frac{dL}{d\mu} = -\frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i - \mu)$$

$$\frac{dL}{d\sigma} = -\frac{1}{\sigma}\left(n + \frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i - \mu)^2\right)$$

Neither the cdf (F) nor quantile function (F^{-1}) have explicit forms. However, both can be readily approximated to arbitrary precision in most mathematical software.

Mean: μ

Variance: σ^2

The standard normal distribution has mean 0 and unit variance when:

$\mu = 0$

$\sigma = 1$

Maximum Likelihood Estimators:

$$\hat{\mu} = \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

$$\hat{\sigma} = s\frac{n-1}{n}$$

Moment Estimators:

$\hat{\mu} = \bar{y}$

$\hat{\sigma} = s$

Logistic distribution

Parameters: $\boldsymbol{\theta} = [\alpha; \beta]$

α (location)

β (scale)

Transformation: $y_i = \log_{10}(x_i)$

$$f(y:\alpha,\beta) = \frac{\exp\left(-(y-\alpha)/\beta\right)}{\beta\left(1+\exp\left(-(y-\alpha)/\beta\right)\right)^2}$$

$$F(y:\alpha,\beta) = \frac{1}{1+\exp\left(-(y-\alpha)/\beta\right)}$$

$$F^{-1}(p) = \alpha + \beta \ln\left(\frac{p}{1-p}\right)$$

Let:

$$r_i = y_i - \alpha$$

$$m_i = \exp\left(-\frac{r_i}{\beta}\right)$$

$$L(\boldsymbol{\theta}\,|\,\mathbf{Y}) = \frac{1}{\beta}\sum_{i=1}^{n}(r_i) - n\ln(\beta) - 2\sum_{i=1}^{n}\ln(1+m_i)$$

$$\frac{dL}{d\alpha} = \frac{n}{\beta} - \frac{2}{\beta}\sum_{i=1}^{n}\left(\frac{m_i}{1+m_i}\right)$$

$$\frac{dL}{d\beta} = \frac{1}{\beta^2}\sum_{i=1}^{n} r_i - \frac{n}{\beta} - \frac{2}{\beta^2}\sum_{i=1}^{n}\frac{r_i m_i}{1+m_i}$$

Mean: α

Variance: $\frac{\pi^2}{3}\beta^2$

The standard logistic distribution has mean 0 and variance 1 when:

$$\alpha = 0\,,\ \beta = \frac{\sqrt{3}}{\pi}$$

Maximum Likelihood Estimators:

Closed form estimators do not exist, but MLEs are easily obtained numerically.

Moment Estimators:

$\hat{\alpha} = \bar{y}$ and $\hat{\beta} = s\dfrac{\sqrt{3}}{\pi}$

Triangular distribution (symmetric)

Parameters: $\boldsymbol{\theta} = [a;b]$, a (minimum), b (maximum)

Transformation: $y_i = \log_{10}(x_i)$

If: $a \le y \le \dfrac{a+b}{2}$:

$$f(y \mid a,b) = \frac{4(y-a)}{(b-a)^2}$$

$$F(y \mid a,b) = \frac{2(y-a)^2}{(b-a)^2}$$

$$L(\boldsymbol{\theta} \mid \mathbf{Y}) = -2\ln(b-a) + \ln(4) + \sum_{i=1}^{n}(y_i - a)$$

If: $p \le 0.5$

$$F^{-1}(p) = a + \sqrt{\frac{p(b-a)^2}{2}}$$

If: $\dfrac{a+b}{2} < y \le b$:

$$f(y \mid a,b) = \frac{4(b-y)}{(b-a)^2}$$

$$F(y \mid a,b) = 1 - \frac{2(y-b)^2}{(b-a)^2}$$

$$L(\boldsymbol{\theta} \mid \mathbf{Y}) = -2\ln(b-a) + \ln(4) + \sum_{i=1}^{n}(b-y$$

If: $p > 0.5$

$$F^{-1}(p) = b + \sqrt{\frac{(1-p)(b-a)^2}{2}}$$

Mean:

$$\frac{a+b}{2}$$

Variance

$$\frac{(b-a)^2}{24}$$

The standard symmetric triangular distribution has mean zero and variance 1 when:

$a = -\sqrt{6}$, and $b = \sqrt{6}$

The following moment estimators can also be used:

$\hat{a} = 2\bar{y} - \hat{b}$ and $\hat{b} = \bar{y} + s\sqrt{6}$

Gumbel (Gompertz) distribution

Parameters: $\boldsymbol{\theta} = [\mu, \beta]$

μ (location)

β (scale)

Transformation: $y_i = \log_{10}(x_i)$

$$F(y) = \exp\left(-\exp\left(\frac{(\mu - y)}{\beta}\right)\right)$$

$$f(y) = \frac{1}{\beta}\exp\left(\frac{(\mu - y)}{\beta} - \exp\left(\frac{(\mu - y)}{\beta}\right)\right)$$

$$F^{-1}(p) = \mu - \beta\ln(-\ln(p))$$

Let:

$$z_i = \frac{\mu - y_i}{\beta}$$

Then:

$$L(\boldsymbol{\theta} \mid \mathbf{Y}) = -n\ln(\beta) + \sum_{i=1}^{n} z_i - \sum_{i=1}^{n}\exp(z_i)$$

Note:

$$\frac{dL}{dz_i} = -1 + \exp(z_i), \quad \frac{dz_i}{d\mu} = \frac{1}{\beta} \quad \text{and} \quad \frac{dz_i}{d\beta} = -\frac{\mu - y_i}{\beta^2}$$

$$\frac{dL}{d\mu} = \frac{1}{\beta}\sum_{i=1}^{n}(1 - \exp(z_i)) \text{ and } \frac{dL}{d\beta} = -\frac{n-1}{\beta}\sum_{i=1}^{n} z_i(1 - \exp(z_i))$$

Mean:

$\mu + \beta\gamma$, where γ = Euler-Mascheroni constant.

Variance

$$\frac{\beta^2\pi^2}{6}$$

The gompertz distribution has mean 0 and variance 1 when :

$$\mu = -\gamma\beta$$

$$\beta = \frac{\sqrt{6}}{\pi}$$

Moment Estimators:

$$\hat{\beta} = \frac{s}{\pi}\sqrt{6} \quad \text{and} \quad \hat{\mu} = \bar{y} - \hat{\beta}\gamma$$

Weibull distribution

Parameters: $\boldsymbol{\theta} = [\lambda; k]$

λ (scale)

k (shape)

Transformation: none

$$F(x) = 1 - \exp\left(-\left(\frac{x}{\lambda}\right)^k\right)$$

$$f(x) = \frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} \exp\left(-\left(\frac{x}{\lambda}\right)^k\right)$$

$$F^{-1}(p) = \lambda\left(-\ln(1-p)\right)^{1/k}$$

$$L(\boldsymbol{\theta} \mid \mathbf{X}) = n\ln(k) - kn\ln(\lambda) + (k-1)\sum_{i=1}^{n}\ln(x_i) - \sum_{i=1}^{n}\left(\frac{x_i}{\lambda}\right)^k$$

$$\frac{dL}{d\lambda} = -\frac{k}{\lambda}\left(n - \sum_{i=1}^{n}\left(\frac{x_i}{\lambda}\right)^k\right)$$

$$\frac{dL}{dk} = \frac{n}{k} - n\ln(\lambda) + \sum_{i=1}^{n}\ln(x_i) - \sum_{i=1}^{n}\left(\left(\frac{x_i}{\lambda}\right)^k \ln\left(\frac{x_i}{\lambda}\right)\right)$$

The mean and variance of the Weibull distribution are:

$$\lambda\Gamma\left(1+\frac{1}{k}\right)$$

$$\lambda^2\Gamma\left(1+\frac{2}{k}\right) - \left(\lambda\Gamma\left(1+\frac{1}{k}\right)\right)^2$$

The gamma functions in the above two equations prevent closed-form moment estimators from being derived.

Pareto distribution

Parameters: $\boldsymbol{\theta} = [x_m; \alpha]$

x_m (scale)
α (shape)

Transformation: none

$$F(x) = 1 - \left(\frac{x_m}{x}\right)^{\alpha}$$

$$f(x) = \frac{\alpha x_m^{\alpha}}{x^{\alpha+1}}$$

$$L(\boldsymbol{\theta} \mid \mathbf{X}) = n\left(\ln(\alpha) + \alpha \ln(x_m)\right) - (\alpha+1)\sum_{i=1}^{n} \ln(x_i)$$

$$\frac{dL}{dx_m} = \frac{\alpha n}{x_m}$$

$$\frac{dL}{d\alpha} = n\left(\frac{1}{\alpha} + \ln(x_m)\right) - \sum_{i=1}^{n} \ln(x_i)$$

Note that the derivative dL / dx_m is necessarily negative because x_m , $\alpha > 0$. Therefore the negative log-likelihood continues to get smaller with larger values of x_m and so standard numerical optimization of the negative log-likelihood will result in $\hat{x}_m \to \infty$. The solution to this is to assume that x_m is known and equal to the minimum observed value. In this case, there is also a simple MLE for α, which can be found by setting $dL / d\alpha = 0$. Thus the MLEs are:

$$\hat{x}_m = \min(x_i)$$

$$\hat{\alpha} = \frac{n}{\sum_{i=1}^{n} \ln\left(\frac{x_i}{\hat{x}_m}\right)}$$

The following moment estimators can also be used:

$$x_m = \frac{\bar{x}(\alpha - 1)}{\alpha}$$

$$\alpha = 1 + \sqrt{\frac{\bar{x}^2}{s^2} + 1}$$

Burr$_{III}$ distribution

(Note: this is the Burr$_{III}$ distribution from Shao 2000, not the Burr$_{XII}$ distribution from Shao 2003)

Parameters: $\boldsymbol{\theta} = [b; c; k]$

$$F(x) = \frac{1}{\left[1+\left(\frac{b}{x}\right)^{c}\right]^{k}}$$

$$f(x) = \frac{kc}{b}\frac{\left(\frac{b}{x}\right)^{c+1}}{\left[1+\left(\frac{b}{x}\right)^{c}\right]^{k+1}}.$$

Note that Shao (2000, Eq. 8) incorrectly gives the pdf as: $f(x) = \frac{kc}{b}\frac{\left(\frac{b}{x}\right)^{c+1}}{\left[1+\left(\frac{b}{x}\right)^{c+1}\right]^{k+1}}$

$$L(\boldsymbol{\theta} \mid \mathbf{X}) = n\ln(c) + n\ln(k) + cn\ln(b) - (c+1)\sum_{i=1}^{n}\ln(x_i) - (k+1)\sum_{i=1}^{n}\ln\left(1+\left(\frac{b}{x_i}\right)^{c}\right)$$

Let:

$$z_i = 1+\left(\frac{b}{x_i}\right)^{c}$$

Then:

$$\frac{dz_i}{db} = \frac{c}{b}\left(\frac{b}{x_i}\right)^{c} \text{ and}$$

$$\frac{dz_i}{dc} = \ln\left(\frac{b}{x_i}\right)\left(\frac{b}{x_i}\right)^{c}$$

$$\frac{dL}{db} = \frac{cn}{b} - (k+1)\sum_{i=1}^{n}\frac{1}{z_i}\frac{dz_i}{db}$$

$$\frac{dL}{dc} = \frac{n}{c} + n\ln(b) - \sum_{i=1}^{n}\ln(x_i) - (k+1)\sum_{i=1}^{n}\frac{1}{z_i}\frac{dz_i}{dc}$$

$$\frac{dL}{dk} = \frac{n}{k} - \sum_{i=1}^{n}\ln(z_i)$$

3. Results & Discussion

Results below are organized around seven of the eight questions outlined in Objectives (Section 1.4). The eighth question, concerning future analyses, is described in Section 5. For brevity only a subset of results that are illustrative of general patterns are presented. Full results are available in section 8, Supplemental Results. As described above, increases in either bias or sampling variance reduce the accuracy of an estimate of the HC_5. Thus patterns of both bias and sampling variance in relation to each question are presented.

Bias is depicted as ratio of the estimated HC_5 to the known HC_5 in two ways. In figures the ratio is plotted with the vertical axis on the $\log_{10}$ scale. In tables, the $\log_{10}$ value of the ratio is given. Thus, in figures, the y-value gives the average relative over- or under-estimate at any given sample size, method and distribution. In tables positive values of the ratio depict a tendency to overestimate the HC_5 (positive bias) and negative values of the ratio depict a tendency to underestimate the HC_5 (negative bias). Calculating bias necessarily implies that the true value of the HC_5 is known. However, this is only the case for the simulated data. For empirical analyses, the HC_5 calculated using all available data is substituted for the known HC_5.

3.1. How does sample size influence bias and variance of the estimated HC_5?

3.1.1. Bias

For all analyses of simulation data, bias in the estimated HC_5 decreased with increasing sample size, regardless of the method used for fitting the distribution (Fig. 2, Table 6). In general, the four distributions fit to $\log_{10}$-transformed data were less biased than the three untransformed distributions. Given the magnitude of sampling variance (see immediately below) it is possible that some of the observed bias may also be comprised of some residual sampling variance. At sample sizes of $n = 3$ and $n = 8$, bias in distributions fit using maximum likelihood and moment estimators was large and always positive.

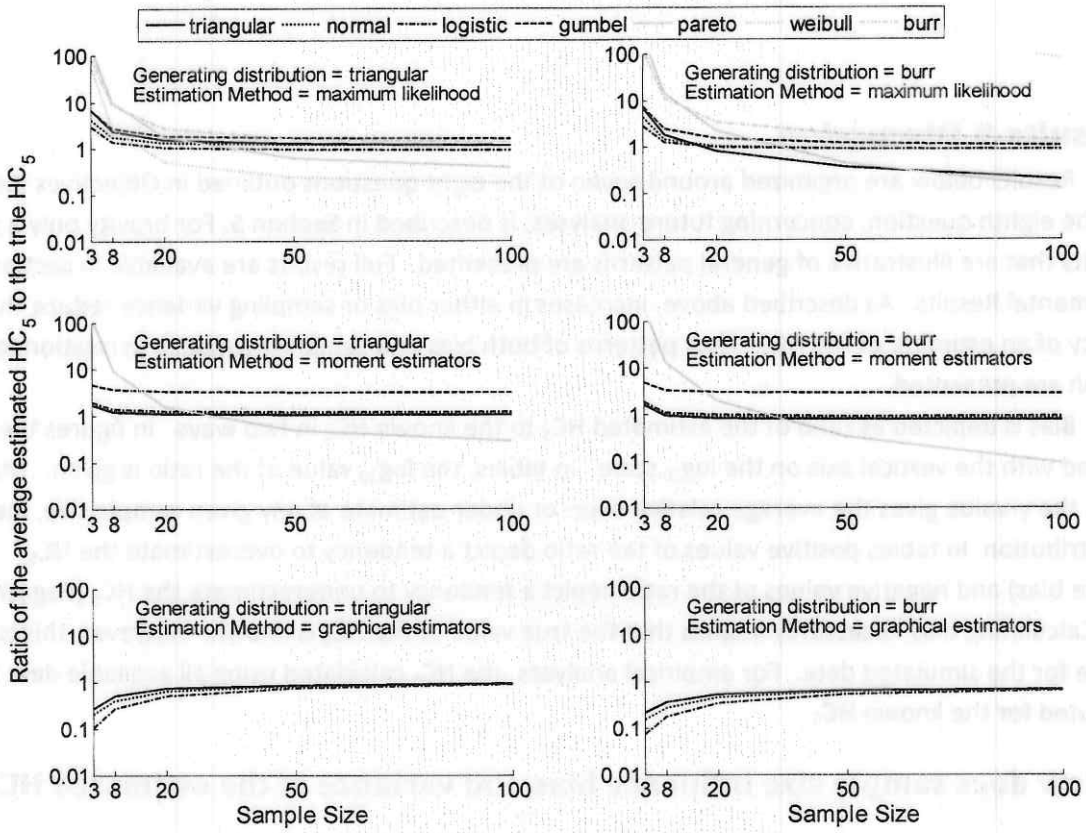


Figure 2. Ratio of the average estimated HC_5 to the true HC_5 for simulated data plotted as a function of sample size for simulated log-triangular data and simulated Burr data.

Table 6. Log_{10} ratios of the average estimated HC_5 to the true HC_5 for seven distributions fit to five sample sizes of simulated log-triangular toxicity data using **maximum likelihood**.

	Sample Size				
Distribution	3	8	20	50	100
triangular	0.813	0.373	0.177	0.085	0.050
normal	0.617	0.243	0.118	0.071	0.056
logistic	0.462	0.138	0.018	-0.031	-0.046
Gumbel	0.821	0.421	0.269	0.210	0.193
Pareto	2.172	0.996	0.245	-0.218	-0.427
Weibull	1.785	0.431	-0.297	-0.635	-0.752
Burr	1.944	0.924	0.426	0.203	0.122

Table 7. Log_{10} ratios of the average estimated HC_5 to the true HC_5 for five distributions fit to five sample sizes of simulated log-triangular toxicity data using **moment estimators**.

	Sample Size				
Distribution	3	8	20	50	100
triangular	0.219	0.067	0.024	0.009	0.005
normal	0.255	0.106	0.063	0.049	0.045
logistic	0.281	0.134	0.092	0.077	0.073
Gumbel	0.662	0.544	0.510	0.498	0.495
Pareto	2.160	0.952	0.167	-0.324	-0.546

Table 8. Log_{10} ratios of the average estimated HC_5 to the true HC_5 for three distributions fit to five sample sizes of simulated log-triangular toxicity data using **graphical methods**.

	Sample Size				
Distribution	3	8	20	50	100
triangular	-0.572	-0.296	-0.141	-0.064	-0.033
normal	-0.697	-0.402	-0.201	-0.085	-0.033
logistic	-0.987	-0.576	-0.298	-0.132	-0.055

Table 9. Log_{10} ratios of the average estimated HC_5 to the true HC_5 for seven distributions fit to five sample sizes of simulated Burr toxicity data using **maximum likelihood**.

	Sample Size				
Distribution	3	8	20	50	100
triangular	0.849	0.261	-0.099	-0.440	-0.718
normal	0.633	0.163	0.011	-0.049	-0.079
logistic	0.453	0.090	-0.019	-0.057	-0.079
Gumbel	0.872	0.383	0.183	0.064	0.002
Pareto	2.412	1.161	0.352	-0.369	-0.907
Weibull	1.947	3.102	3.397	2.793	1.877
Burr	2.116	1.044	0.534	0.255	0.131

Table 10. Log_{10} ratios of the average estimated HC_5 to the true HC_5 for five distributions fit to five sample sizes of simulated Burr toxicity data using **moment estimators**.

	Sample Size				
Distribution	3	8	20	50	100
triangular	0.194	-0.036	-0.097	-0.120	-0.137
normal	0.233	0.008	-0.052	-0.074	-0.091
logistic	0.262	0.039	-0.019	-0.042	-0.058
Gumbel	0.683	0.504	0.459	0.440	0.426
Pareto	2.394	1.114	0.273	-0.481	-1.046

Table 11. Log_{10} ratios of the average estimated HC_5 to the true HC_5 for three distributions fit to five sample sizes of simulated Burr toxicity data using **graphical estimators**.

	Sample Size				
Distribution	3	8	20	50	100
triangular	-0.681	-0.448	-0.285	-0.203	-0.181
normal	-0.818	-0.567	-0.353	-0.227	-0.181
logistic	-1.138	-0.765	-0.465	-0.281	-0.206

3.1.2. Variance

Variance in the estimated HC_5 decreased steeply with sample size up to about $n = 50$ in analyses of simulated data (Fig. 3, note log_{10}-scale of vertical axis). As with bias, the normal, logistic, and triangular distributions performed very similarly across sample sizes and fitting methods. At large sample sizes, distributions on untransformed data often showed smaller sampling variances, especially the Pareto distribution (Fig. 3). The Weibull distribution fit to $Burr_{III}$-distributed data using maximum likelihood showed an anomalous and enormous increase in sampling variance between $n = 3$ and $n = 100$ (not shown), suggesting instability in the numerical algorithms for estimating parameters of that distribution. This pattern of instability in the Weibull was repeated across other generating distributions as well (see Section 8, Supplementary Results).

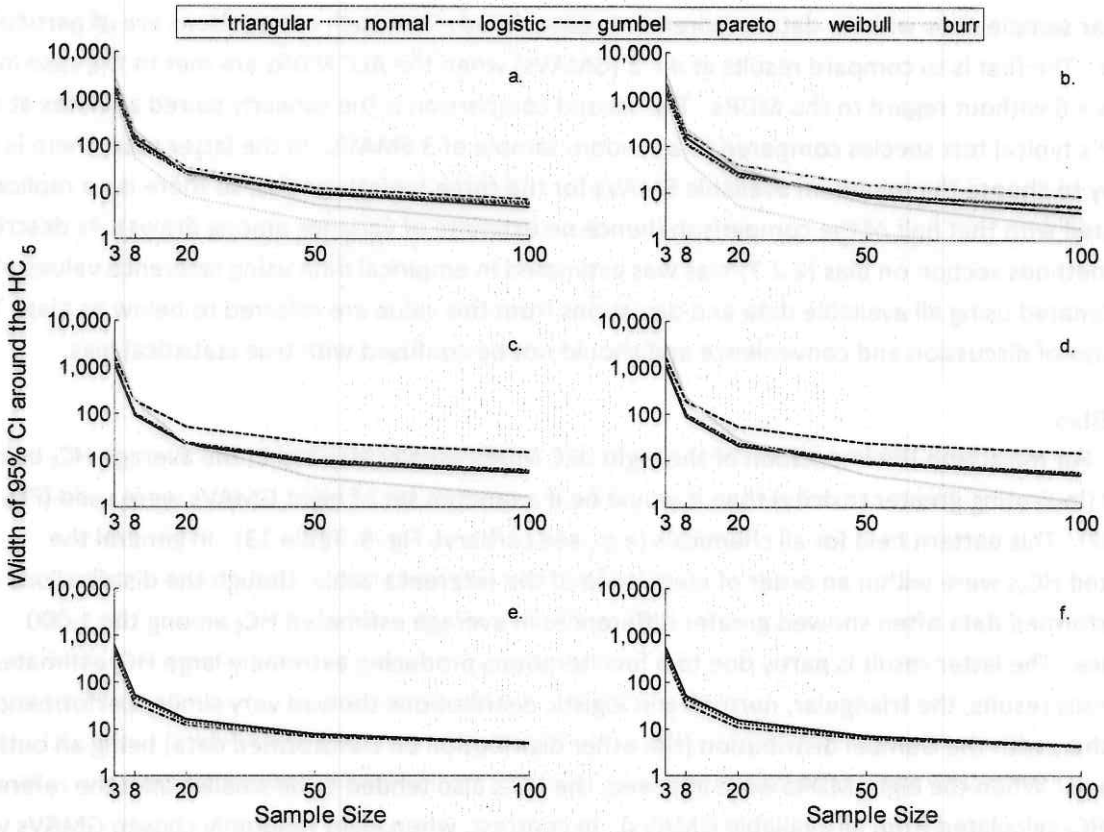


Figure 3. Width of 95% confidence interval in concentration units for 10,000 replicates of simulated data plotted as a function of sample size. Simulated data are log-triangular (column 1: panels a, c, and e) and $Burr_{III}$ (column 2: panels b, d, and f). Distributions fit using maximum likelihood (row 1: panels a and b), moment estimators (row 2: panels c and d) and graphical methods (row 3: panels e and f).

In summary, over all estimation methods, both bias and variance decreased with increasing sample size. The slope of this relationship was quite steep up to about $n = 20$, suggesting that each additional data point results in substantially improved performance at limited sample sizes. Maximum likelihood estimates exhibited the greatest variability, moment estimates were generally the least biased, and graphical methods tended to produce the most conservative estimates. Even at the smallest sample sizes all methods produced estimates that were often, on average, within an order of magnitude of the true HC_5. However, this result should be interpreted with caution, because at limited sample sizes the estimated HC_5s may range over two or three orders of magnitude when data requirements are not imposed. The precision of estimated HC_5s improves considerably with imposition of data requirements (see below).

3.2. How do minimum data requirements influence bias and variance of the estimated HC_5?

Results of the simulation exercises do not inform the effects of the imposition of data requirements because the randomly chosen data points from theoretical distributions did not possess taxonomic identities. This section describes results of the imposition of data requirements to analyses

at similar sample sizes with no data requirements considered. Two such comparisons are of particular interest. The first is to compare results at n = 8 (GMAVs) when the ALC MDRs are met to the case in which n = 8 without regard to the MDRs. The second comparison is the similarly paired analyses at n = 3 for OPP's typical test species compared to a random sample of 3 SMAVs. In the latter case, there is only one way to choose the minimum available SMAVs for the three typical species, so there is no replication associated with that half of the comparison (hence no estimate of variance among draws). As described in the methods section on bias (2.2.7) bias was estimated in empirical data using reference values of the HC_5 estimated using all available data and deviations from this value are referred to below as bias. This is for ease of discussion and convenience and should not be confused with true statistical bias.

3.2.1. Bias

For malathion the imposition of the eight ALC MDRs usually resulted in the average HC_5 being smaller (indicating greater toxicity) than it would be if a random set of eight GMAVs were used (Fig. 4, Table 12). This pattern held for all chemicals (*e.g.,* see carbaryl, Fig. 5, Table 13). In general the estimated HC_5s were within an order of magnitude of the reference value, though the distributions on untransformed data often showed greater differences in average estimated HC_5 among the 1,000 iterations. The latter result is partly due to a few iterations producing extremely large HC_5 estimates. As in previous results, the triangular, normal, and logistic distributions showed very similar performance to each other, with the Gumbel distribution (the other distribution on transformed data) being an outlier in this group. When the eight MDRs were imposed, the HC_5s also tended to be smaller than the reference value (HC_5 calculated from all available GMAVs). In contrast, when eight randomly chosen GMAVs were used, the estimated HC_5 was often larger than the reference value (Figs. 4, 5). An exception to this pattern occurred with graphical methods (Fig. 3c).

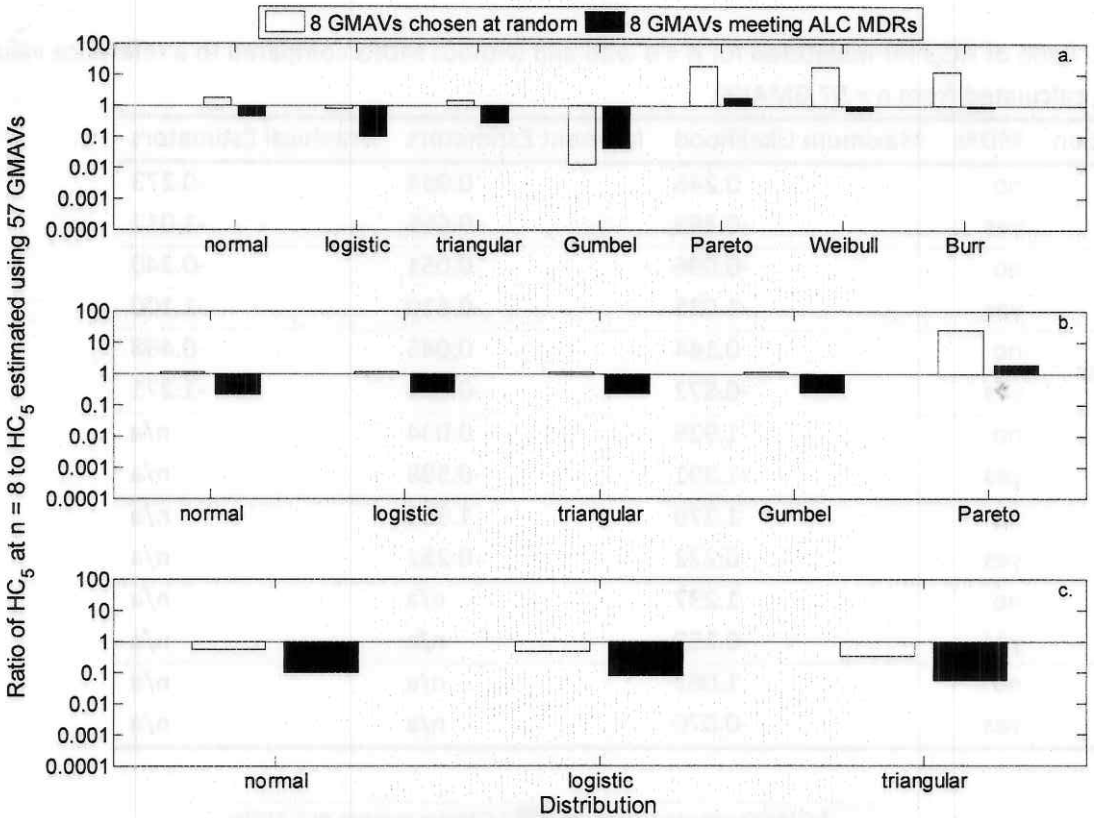


Figure 4. Average ratio of estimated HC_5s for **malathion** from 1,000 draws of empirical data (n = 8 GMAVs) to the HC_5 estimated from all available GMAVs (n = 57). Panels represent a. maximum likelihood estimates, b. moment estimates, c. graphical estimates. Note $\log_{10}$ scale of Y-axis.

Table 12. Ratio of HC_5s for **malathion** for n = 8 with and without MDRs compared to a reference value (*i.e.*, HC_5 calculated from n = 57 GMAVs)

Distribution	MDRs	Maximum Likelihood	Moment Estimators	Graphical Estimators
normal	no	0.246	0.054	-0.273
	yes	-0.369	-0.666	-1.012
logistic	no	-0.096	0.051	-0.340
	yes	-1.025	-0.610	-1.100
triangular	no	0.144	0.045	-0.488
	yes	-0.572	-0.634	-1.271
Gumbel	no	-1.929	0.034	n/a
	yes	-1.391	-0.598	n/a
Pareto	no	1.279	1.383	n/a
	yes	0.222	0.252	n/a
Weibull	no	1.237	n/a	n/a
	yes	-0.159	n/a	n/a
Burr	no	1.083	n/a	n/a
	yes	-0.070	n/a	n/a

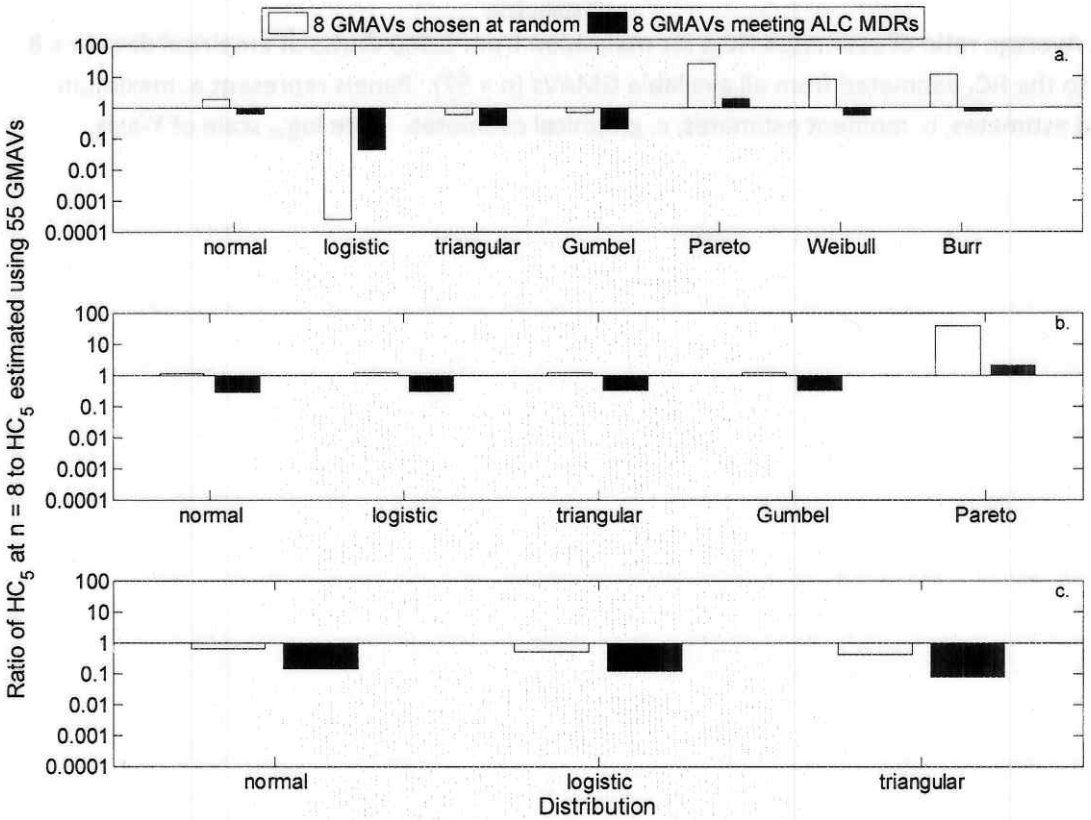


Figure 5. Average ratio of estimated HC_5s for **carbaryl** from 1,000 draws of empirical data (n = 8 GMAVs)

to the HC_5 estimated from all available freshwater GMAVs (n = 55). Panels represent a. maximum likelihood estimates, b. moment estimates, c. graphical estimates. Note log_{10} scale of Y-axis.

Table 13. Ratio of HC_5s for **carbaryl** for n = 8 with and without MDRs compared to a reference value (*i.e.*, HC_5 calculated from n = 55 GMAVs)

Distribution	MDRs	Maximum Likelihood	Moment Estimators	Graphical Estimators
normal	no	0.254	0.033	-0.214
	yes	-0.193	-0.537	-0.872
logistic	no	-3.604	0.068	-0.298
	yes	-1.370	-0.518	-0.935
triangular	no	-0.249	0.065	-0.376
	yes	-0.592	-0.491	-1.099
Gumbel	no	-0.161	0.078	n/a
	yes	-0.702	-0.479	n/a
Pareto	no	1.396	1.570	n/a
	yes	0.269	0.293	n/a
Weibull	no	0.892	n/a	n/a
	yes	-0.251	n/a	n/a
Burr	no	1.049	n/a	n/a
	yes	-0.129	n/a	n/a

Analysis of the three typical OPP test species also on average resulted in smaller HC_5s (indicating greater toxicity) compared to the analysis of random subsets of 3 SMAVs (Figs. 6, 7 Tables 14,15). The pattern was similar to that resulting from imposition of the 8 ALC MDRs, except that bias ratios were much larger for the distributions on untransformed data, ranging close to two orders of magnitude. Similarity of performance among the distributions on untransformed data is again evident (Figs. 6, 7). Values of the HC_5s estimated using all available SMAVs are given in Section 8. Similarly, figures showing comparison between random sets of size 3 and the three typical OPP species are given for all chemicals considered in Section 8.

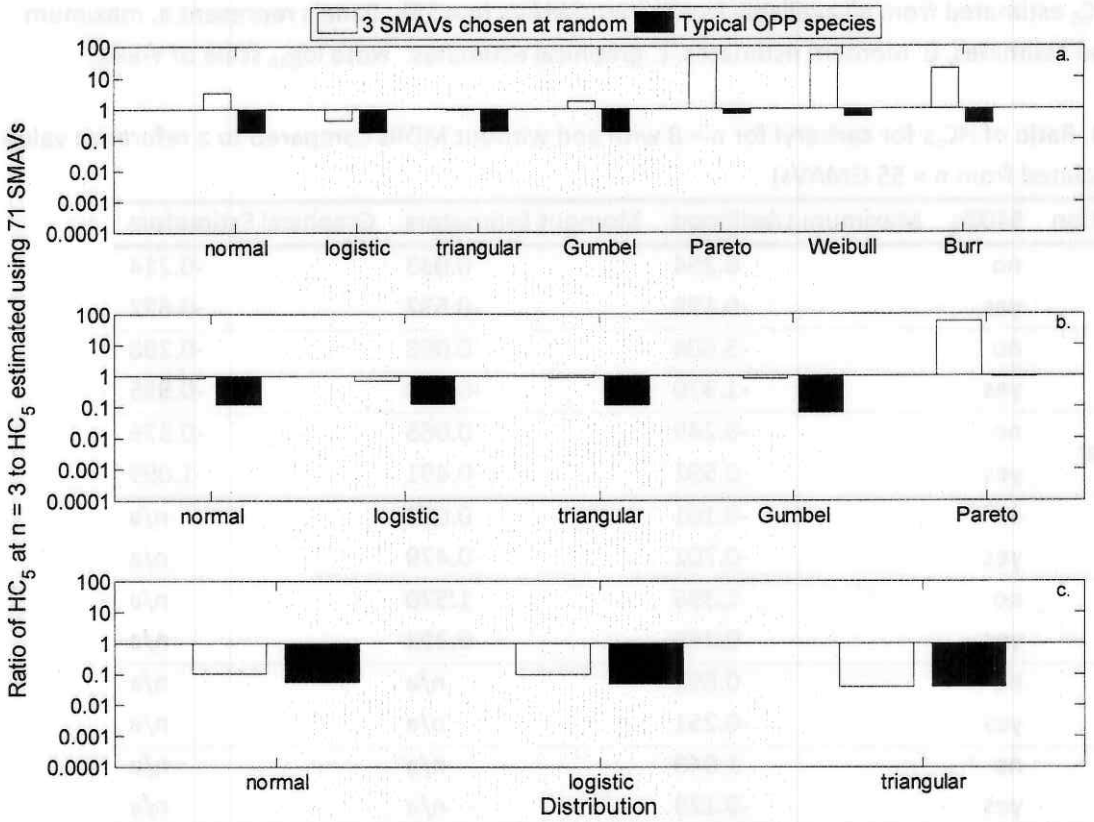


Figure 6. Average ratio of estimated HC_5s for **malathion** from 1,000 draws of empirical data (n = 3 SMAVs) to the HC_5 estimated from all available SMAVs (n = 71). Panels represent a. maximum likelihood estimates, b. moment estimates, c. graphical estimates. Note $\log_{10}$ scale of Y-axis.

Table 14. Ratio of HC_5s for **malathion** for n = 3 with and without MDRs compared to a reference value (*i.e.*, HC_5 calculated from n = 71 SMAVs).

Distribution	MDRs	Maximum Likelihood	Moment Estimators	Graphical Estimators
normal	no	0.502	-0.079	-0.983
	yes	-0.767	-0.910	-1.271
logistic	no	-0.405	-0.175	-1.028
	yes	-0.757	-0.933	-1.321
triangular	no	-0.055	-0.075	-1.412
	yes	-0.745	-0.949	-1.434
Gumbel	no	0.228	-0.098	n/a
	yes	-0.763	-1.191	n/a
Pareto	no	1.701	1.744	n/a
	yes	-0.159	-0.053	n/a
Weibull	no	1.694	n/a	n/a
	yes	-0.275	n/a	n/a
Burr	no	1.280	n/a	n/a
	yes	-0.495	n/a	n/a

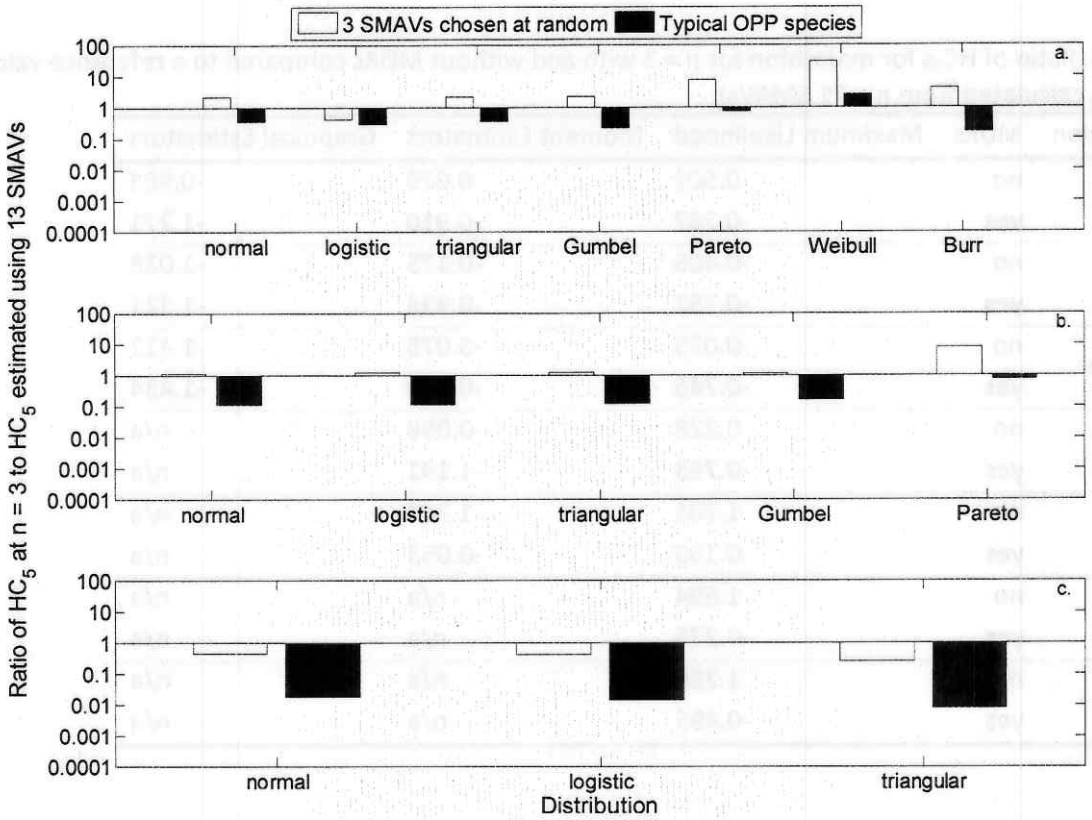


Figure 7. Average ratio of estimated HC_5s for **propoxur** from 1,000 draws of empirical data (n = 3 SMAVs) to the HC_5 estimated from all available SMAVs (n = 13). Panels represent a. maximum likelihood estimates, b. moment estimates, c. graphical estimates. Note $\log_{10}$ scale of Y-axis.

Table 15. Ratio of HC_5s for **propoxur** for n = 3 with and without MDRs compared to a reference value (*i.e.*, HC_5 calculated from n = 13 SMAVs).

Distribution	MDRs	Maximum Likelihood	Moment Estimators	Graphical Estimators
normal	no	0.321	0.059	-0.347
	yes	-0.441	-0.953	-1.759
logistic	no	-0.403	0.076	-0.403
	yes	-0.550	-0.940	-1.850
triangular	no	0.343	0.087	-0.596
	yes	-0.459	-0.931	-2.110
Gumbel	no	0.322	0.051	n/a
	yes	-0.675	-0.793	n/a
Pareto	no	0.849	0.905	n/a
	yes	-0.139	-0.135	n/a
Weibull	no	1.458	n/a	n/a
	yes	0.393	n/a	n/a
Burr	no	0.614	n/a	n/a
	yes	-0.800	n/a	n/a

3.2.2. Variance

Imposition of the 8 ALC MDRs greatly reduced the variance in 1,000 replicate draws of empirical data (Fig. 8, note $\log_{10}$ scale of vertical axis). In general, 95% confidence intervals around estimated HC_5s were at least twice as wide when the MDRs were not imposed. Thus, while the imposition of MDRs often produced, on average, HC_5s that differed only modestly from similar sample sizes with no data requirements, particular results from given iterations differed substantially. Random subsets of size 3 (SMAVs) produced even larger variance estimates than random subsets of size 8 (Fig. 9, again note $\log_{10}$ scale of vertical axis). This pattern holds across all AChE inhibitors, but the variances are not always as large as those observed for malathion (*e.g.*, propoxur, Fig. 10). This result reflects two important differences when data requirements are imposed. First, there are fewer combinations of test results that satisfy the requirements, and thus less variation in the resulting estimated HC_5s. Second, and more importantly, the imposition of data requirements forces the inclusion of sensitive species in the analyses. Thus subsets of data including only test results with high LC50s are excluded from consideration.

In summary, minimum data requirements greatly improved the quality of HC_5 estimates, regardless of the method used for estimation. Although MDRs often produced more biased estimates of the HC_5, they virtually always did so in a conservative way – by producing HC_5 estimates that were, on average, lower than the reference values (HC_5s estimated using all available data). Data requirements at n = 8 also greatly reduced the variation among HC_5 estimates. The latter result is due partly to the forced inclusion of both sensitive and insensitive taxa in every data sample and partly due to the reduced number of possible data combinations.

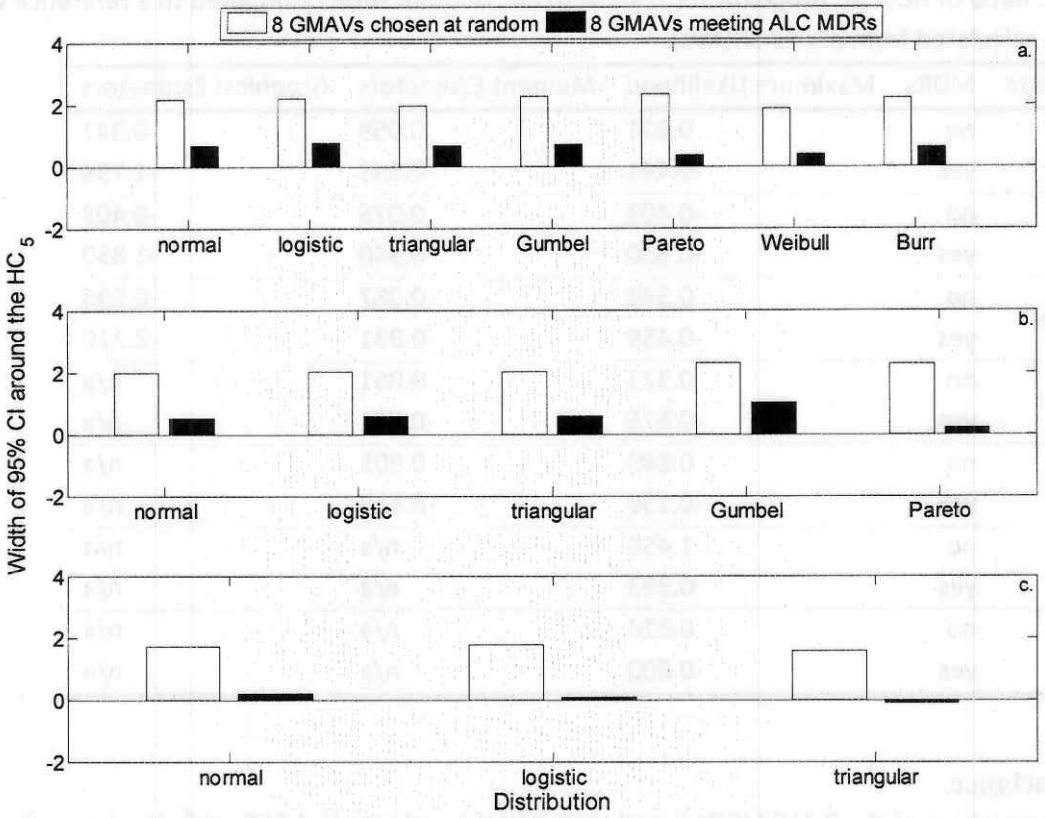


Figure 8. Width of 95% confidence interval for **malathion** HC_5 estimates from 1,000 draws of 8 randomly chosen GMAVs compared to when the 8 ALC MDRs are imposed. Y-axis is plotted on $\log_{10}$ scale. The three panels correspond to a. maximum likelihood, b. moment estimators, and c. graphical methods, respectively.

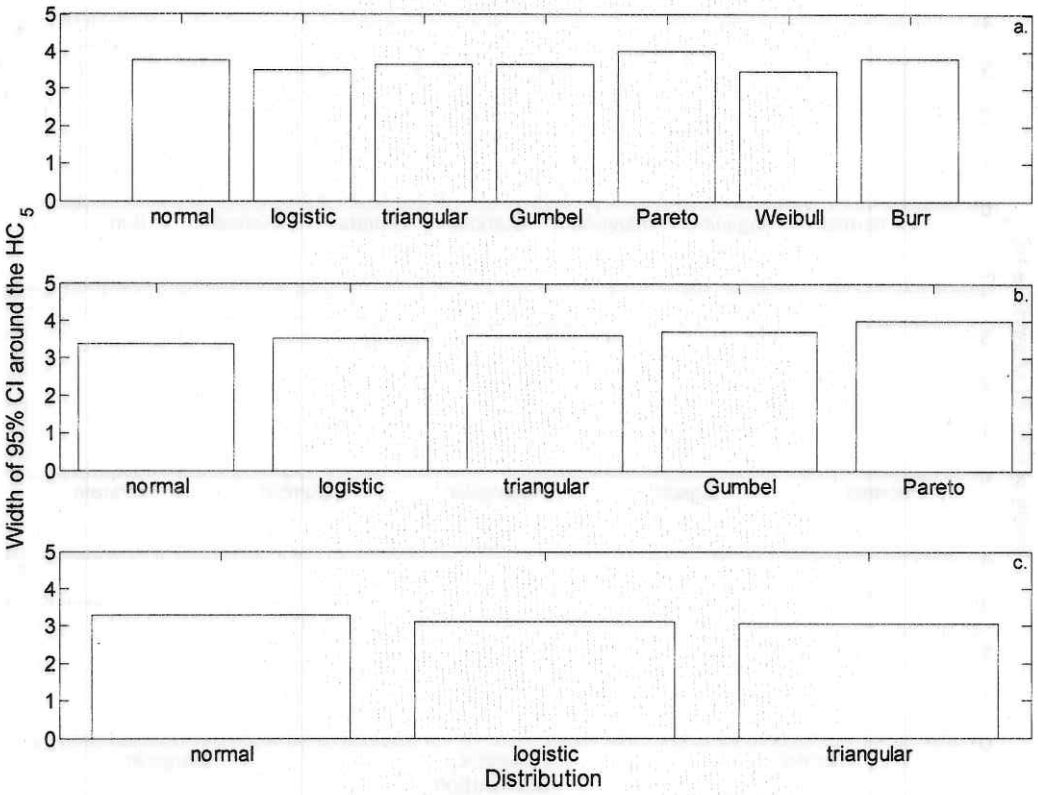


Figure 9. Width of 95% confidence interval for **malathion** HC_5 estimates from 1,000 draws of 3 randomly chosen SMAVs. Y-axis is plotted on $\log_{10}$ scale. The three panels correspond to a. maximum likelihood, b. moment estimators, and c. graphical methods, respectively.

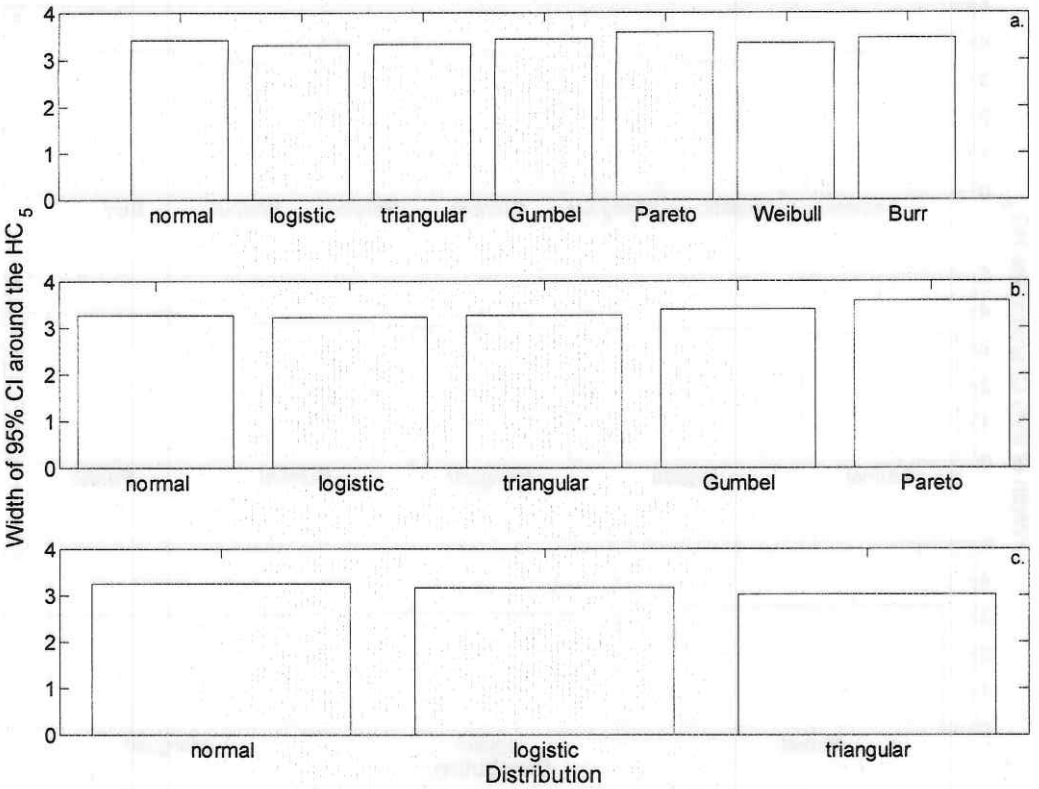


Figure 10. Variance in **propoxur** HC_5 estimates from 1,000 draws of 3 randomly chosen SMAVs. Y-axis is plotted on $\log_{10}$ scale. The three panels correspond to a. maximum likelihood, b. moment estimators, and c. graphical methods, respectively.

3.3. How do different estimation methods influence bias and variance of the estimated HC_5?

3.3.1. Bias

In analyses of simulated data, the most notable effect of fitting method was that the graphical methods tended to produce the lowest (most conservative) estimates of the HC_5, especially in the range of sample sizes of concern for this project ($3 \leq n \leq 8$, Fig. 2). However, moment estimators generally produced the least biased (closest to true or reference HC_5) estimates in this range of sample sizes. Differences among methods diminished as sample size increased. With both moment estimators and graphical methods, bias was often still quite small, even at the smaller sample sizes. Within the range of sample sizes of interest maximum likelihood consistently overestimated the true HC_5. The reason for this overestimation is partly due to the method (maximum likelihood) consistently underestimating the variance of the distribution being fit. When coupled with unbiased estimates of the mean, the underestimate of the variance has the effect of drawing the HC_5 in towards the mean (*i.e.*, introducing a positive bias in the estimated HC_5).

Analyses of empirical data followed similar patterns as analyses of simulated data (above) with respect to estimation method. In analyses of repeated draws of empirical data for malathion at $n = 8$ GMAVs, moment estimators tended to produce closer estimates of the HC_5 to the reference values (analysis of all available GMAVs) when the ALC MDRs were met (Figs. 6, 7). In analyses of random

samples of size n = 3 SMAVs graphical methods also produced, on average, smaller (more conservative) estimates of the HC_5 compared to reference values (Fig. 9, n = 71 SMAVs), than did moment estimators. Maximum likelihood consistently produced larger estimates than the reference values (empirical data).

3.3.2. Variance

Variance was very large, especially with limited data and similar among estimation methods. Variance among estimated HC_5s tended to be slightly smaller with graphical methods, but was still very large when data was limited (Fig. 3). Similar patterns were observed in analyses of empirical data.

3.4. Are measures of fit (AIC, goodness-of-fit) reliable indicators of performance?

3.4.1. Akaike's Information Criterion

In analyses of AIC_c weights for all distributions that were fit to the simulation data for the normal distribution, AIC_c showed little ability to discriminate among distributions when data was limited (Table 16 - 22). With simulated log-normal data, even at n = 50, AIC_c failed to assign the most weight to the normal distribution, though by n = 100, the normal distribution was identified as the best distribution, albeit with only 39% of total weight on average (Table 16). This pattern generally held across distributions (Tables 17-22), though AIC_c showed better discriminatory ability with the Weibull and Pareto distributions (Table 20, 21). Note that at n = 3 AIC_c cannot be calculated for the distributions considered here.

Table 16. Average AIC_c weight for each distribution fit to 10,000 samples of log-**normal** data.

n	**normal**	logistic	triangular	Gumbel	Weibull	Pareto	Burr
8	**0.15**	0.12	0.21	0.12	0.16	0.24	0.01
20	**0.19**	0.15	0.30	0.11	0.14	0.06	0.05
50	**0.25**	0.19	0.34	0.06	0.07	0.00	0.08
100	**0.39**	0.21	0.25	0.02	0.03	0.00	0.10

Table 17. Average AIC_c weight for each distribution fit to 10,000 samples of log-**logistic** data.

n	normal	**logistic**	triangular	Gumbel	Weibull	Pareto	Burr
8	0.14	**0.13**	0.20	0.12	0.17	0.23	0.01
20	0.18	**0.19**	0.24	0.12	0.16	0.06	0.06
50	0.22	**0.32**	0.17	0.07	0.09	0.00	0.13
100	0.24	**0.46**	0.06	0.02	0.03	0.00	0.20

Table 18. Average AIC_c weight for each distribution fit to 10,000 samples of log-**triangular** data.

n	normal	logistic	**triangular**	Gumbel	Weibull	Pareto	Burr
8	0.15	0.12	**0.21**	0.12	0.16	0.24	0.01
20	0.18	0.12	**0.36**	0.10	0.13	0.06	0.04
50	0.17	0.07	**0.63**	0.04	0.05	0.00	0.03
100	0.11	0.02	**0.84**	0.01	0.01	0.00	0.01

Table 19. Average AIC_c weight for each distribution fit to 10,000 samples of log-**Gumbel** data.

n	normal	logistic	triangular	**Gumbel**	Weibull	Pareto	Burr
8	0.11	0.10	0.16	**0.14**	0.09	0.40	0.01
20	0.11	0.10	0.16	**0.25**	0.02	0.30	0.07
50	0.06	0.06	0.08	**0.55**	0.00	0.07	0.18
100	0.02	0.02	0.01	**0.69**	0.00	0.00	0.26

Table 20. Average AIC_c weight for each distribution fit to 10,000 samples of **Weibull** data.

n	normal	logistic	triangular	Gumbel	**Weibull**	Pareto	Burr
8	0.15	0.14	0.21	0.08	**0.28**	0.13	0.01
20	0.13	0.13	0.19	0.02	**0.49**	0.01	0.03
50	0.07	0.07	0.08	0.00	**0.76**	0.00	0.02
100	0.03	0.02	0.01	0.00	**0.93**	0.00	0.01

Table21. Average AIC_c weight for each distribution fit to 10,000 samples of **Pareto** data.

n	normal	logistic	triangular	Gumbel	Weibull	**Pareto**	Burr
8	0.06	0.05	0.08	0.11	0.03	**0.66**	0.01
20	0.01	0.01	0.02	0.06	0.00	**0.89**	0.01
50	0.00	0.00	0.00	0.01	0.00	**0.99**	0.00
100	0.00	0.00	0.00	0.00	0.00	**1.00**	0.00

Table 22. Average AIC_c weight for each distribution fit to 10,000 samples of **$Burr_{III}$** data.

n	normal	logistic	triangular	Gumbel	Weibull	Pareto	**Burr**
8	0.14	0.12	0.20	0.12	0.16	0.25	**0.01**
20	0.17	0.18	0.24	0.14	0.12	0.08	**0.06**
50	0.21	0.31	0.17	0.11	0.05	0.00	**0.14**
100	0.24	0.42	0.06	0.04	0.01	0.00	**0.23**

3.4.2. Goodness-of-fit

Results of parametric goodness-of-fit tests are presented in a series of tables below. In all cases, the tables give P-values for the probability of detecting lack-of-fit in a fitted distribution with a desired Type I error rate of $\alpha = 0.05$. Columns represent attempts to fit a given distribution to data generated under the distribution specified for each row. P-values represent the average frequency with which a distribution is rejected (at $\alpha = 0.05$) by a bootstrap goodness-of-fit test. Thus, when the generating distribution (row) and fitted distribution (column) are identical, the rejection rate should be close to the nominal rate (0.05). When the generating and fitted distributions are different, the P-values represent the power of the test to detect lack of fit. Ideally they should be elevated over the nominal rate. Each table is specific to a sample size and fitting method. Only the results of parametric bootstrap tests are presented; goodness-of-fit tests based on the empirical bootstrap never rejected alternative distributions at rates above the nominal rate.

At $n = 3$ parametric goodness-of-fit tests showed approximately nominal rejection rates at $\alpha = 0.05$ (diagonal entries Tables 23 - 25). However, they also showed little power to reject distributions from which the data were not generated (off-diagonal P-values in Tables 23 - 25 rarely larger than

nominal rejection rate: 0.05). At slightly larger samples (n = 8), off-diagonal rejection rates were elevated over the nominal rate, but still reflected a low probability of rejecting alternative distributions as candidates (Tables 26 - 28). At large sample sizes (n = 100), power was high (Tables 29 - 31) and alternative distributions would be rejected more often than not. Power tables for intermediate sample sizes (n = 20 and n = 50) are presented in the supplementary results (Section 8). Power did not depend upon the method used to fit the distribution. However, rejection rates for moment estimators for the Pareto distribution suggest that they rarely produce reasonable parameter estimates.

Table 23. Power of parametric bootstrap goodness-of-fit tests for maximum likelihood at n = 3.

		Fitted Distribution						
		normal	logistic	triangular	Gumbel	Pareto	Weibull	Burr
Generating Distribution	normal	**0.047**	0.047	0.053	0.040	0.027	0.040	0.040
	logistic	0.053	**0.040**	0.027	0.027	0.000	0.053	0.027
	triangular	0.073	0.073	**0.053**	0.073	0.047	0.040	0.060
	Gumbel	0.087	0.100	0.073	**0.073**	0.053	0.067	0.040
	Pareto	0.040	0.067	0.060	0.033	**0.053**	0.067	0.013
	Weibull	0.040	0.047	0.073	0.027	0.020	**0.047**	0.040
	Burr	0.053	0.060	0.060	0.020	0.020	0.053	**0.033**

Table 24. Power of parametric bootstrap goodness-of-fit tests for moment estimators at n = 3.

		Fitted Distribution				
		normal	logistic	triangular	Gumbel	Pareto
Generating Distribution	normal	**0.047**	0.053	0.060	0.053	0.513
	logistic	0.040	**0.027**	0.040	0.127	0.520
	triangular	0.053	0.073	**0.067**	0.060	0.567
	Gumbel	0.087	0.073	0.093	**0.073**	0.460
	Pareto	0.067	0.067	0.047	0.027	**0.567**
	Weibull	0.047	0.060	0.060	0.073	0.573
	Burr	0.060	0.073	0.053	0.067	0.540

Table 25. Power of parametric bootstrap goodness-of-fit tests for graphical estimates at n = 3.

		Fitted Distribution		
		normal	logistic	triangular
Generating Distribution	normal	**0.053**	0.040	0.060
	logistic	0.033	**0.040**	0.047
	triangular	0.060	0.080	**0.060**
	Gumbel	0.093	0.093	0.100
	Pareto	0.073	0.047	0.067
	Weibull	0.047	0.053	0.060
	Burr	0.053	0.053	0.067

Table 26. Power of parametric bootstrap goodness-of-fit tests for maximum likelihood at $n = 8$.

		Fitted Distribution						
		normal	logistic	triangular	gumbel	Pareto	Weibull	Burr
Generating Distribution	normal	**0.107**	0.107	0.080	0.073	0.033	0.067	0.087
	logistic	0.033	**0.040**	0.027	0.060	0.067	0.040	0.033
	triangular	0.047	0.093	**0.067**	0.073	0.047	0.093	0.047
	Gumbel	0.067	0.060	0.093	**0.020**	0.033	0.087	0.047
	Pareto	0.300	0.187	0.313	0.020	**0.013**	0.120	0.147
	Weibull	0.127	0.080	0.133	0.087	0.200	**0.047**	0.127
	Burr	0.067	0.107	0.120	0.047	0.060	0.093	**0.060**

Table 27. Power of parametric bootstrap goodness-of-fit tests for moment estimators at $n = 8$.

		Fitted Distribution				
		normal	logistic	triangular	Gumbel	Pareto
Generating Distribution	normal	**0.093**	0.093	0.073	0.100	0.980
	logistic	0.040	**0.027**	0.047	0.140	0.980
	triangular	0.060	0.047	**0.067**	0.133	0.960
	Gumbel	0.087	0.073	0.113	**0.033**	0.973
	Pareto	0.367	0.273	0.353	0.047	**0.980**
	Weibull	0.113	0.107	0.133	0.253	0.980
	Burr	0.113	0.093	0.107	0.140	0.960

Table 28. Power of parametric bootstrap goodness-of-fit tests for graphical estimates at $n = 8$.

		Fitted Distribution		
		normal	logistic	triangular
Generating Distribution	normal	**0.040**	0.040	0.040
	logistic	0.053	**0.027**	0.073
	triangular	0.067	0.040	**0.073**
	Gumbel	0.107	0.073	0.147
	Pareto	0.300	0.240	0.313
	Weibull	0.127	0.100	0.127
	Burr	0.100	0.073	0.093

Table 29. Power of parametric bootstrap goodness-of-fit tests for maximum likelihood at $n = 100$.

		Fitted Distribution						
		normal	logistic	triangular	gumbel	Pareto	Weibull	Burr
Generating Distribution	normal	**0.067**	0.147	0.453	0.567	1.000	0.753	0.127
	logistic	0.167	**0.073**	0.860	0.547	1.000	0.820	0.120
	triangular	0.093	0.227	**0.093**	0.627	1.000	0.733	0.173
	Gumbel	0.807	0.627	1.000	**0.033**	1.000	0.613	0.067
	Pareto	1.000	1.000	1.000	0.107	**0.053**	0.260	0.993
	Weibull	0.773	0.600	1.000	0.220	1.000	**0.040**	0.853
	Burr	0.200	0.047	0.880	0.487	1.000	0.867	**0.053**

Table 30. Power of parametric bootstrap goodness-of-fit tests for moment estimates at n = 100.

		Fitted Distribution				
		normal	logistic	triangular	Gumbel	Pareto
Generating Distribution	normal	**0.073**	0.133	0.147	0.793	1.000
	logistic	0.167	**0.053**	0.440	0.853	1.000
	triangular	0.087	0.260	**0.073**	0.893	1.000
	Gumbel	0.827	0.800	0.887	**0.053**	1.000
	Pareto	1.000	1.000	1.000	0.887	**1.000**
	Weibull	0.807	0.760	0.880	1.000	1.000
	Burr	0.207	0.080	0.427	0.660	1.000

Table 31. Power of parametric bootstrap goodness-of-fit tests for graphical estimates at n = 100.

		Fitted Distribution		
		normal	logistic	triangular
Generating Distribution	normal	**0.080**	0.047	0.180
	logistic	0.293	**0.047**	0.520
	triangular	0.060	0.060	**0.080**
	Gumbel	0.853	0.760	0.913
	Pareto	1.000	1.000	1.000
	Weibull	0.807	0.727	0.920
	Burr	0.273	0.113	0.487

3.5. What are actual coverage rates of 95% confidence limits for subsamples?

With simulated data, the actual coverage rates of 95% lower confidence limits on the HC_5 were very close to nominal coverage rates at n = 3 and n = 8 (Tables 32 and 33). Actual coverage rates were slightly closer to nominal rates when the data were fit to the same distribution under which it was generated, but the effect of switching distributions was quite minor (Tables 32 and 33).

Table 32. Frequency with which the true HC_5 is greater than lower 95% confidence limit based on published extrapolation constants at N = 3.

		Fitted Distribution		
		normal	logistic	triangular
Generating Distribution	normal	0.948	0.953	0.948
	logistic	0.943	0.951	0.944
	triangular	0.951	0.956	0.952

Table 33. Frequency with which the true HC_5 is greater than lower 95% confidence limit based on published extrapolation constants at N = 8.

		Fitted Distribution		
		normal	logistic	triangular
Generating Distribution	normal	0.948	0.960	0.942
	logistic	0.937	0.952	0.929
	triangular	0.952	0.966	0.945

In contrast to simulated data, actual coverage rates generally differed from nominal rates for empirical data, when the true HC_5 was taken to be the value estimated using all available data. When minimum data requirements were imposed, actual coverage rates always exceeded nominal rates (in fact actual coverage rates were often 100%). When minimum data requirements were not imposed coverage rates were often much lower than nominal rates (Tables 34, 37, 39).

Table 34. Frequency with which the reference HC_5 is greater than lower 95% confidence limit based on published extrapolation constants at N = 8, but not necessarily considering ALC MDRs for **malathion**.

		Fitted Distribution		
		normal	logistic	triangular
Generating Distribution	normal	0.696	0.727	0.657
	logistic	0.823	0.849	0.778
	triangular	0.704	0.651	0.640

Table 35. Frequency with which the reference HC_5 is greater than lower 95% confidence limit based on published extrapolation constants at N = 8, and ALC MDRs are met for **malathion**.

		Fitted Distribution		
		normal	logistic	triangular
Generating Distribution	normal	1.000	0.999	0.998
	logistic	1.000	1.000	1.000
	triangular	0.999	0.997	0.989

Table 36. Frequency with which the reference HC_5 is greater than lower 95% confidence limit based on published extrapolation constants at N = 3, but not necessarily typical OPP species for **malathion**.

		Fitted Distribution		
		normal	logistic	triangular
Generating Distribution	normal	0.978	0.958	0.960
	logistic	0.969	0.980	0.967
	triangular	0.971	0.960	0.967

Table 37. Frequency with which the reference HC_5 is greater than lower 95% confidence limit based on published extrapolation constants at N = 3, but not necessarily considering ALC MDRs for **carbaryl**.

		Fitted Distribution		
		normal	logistic	triangular
Generating Distribution	normal	0.712	0.674	0.635
	logistic	0.848	0.811	0.738
	triangular	0.619	0.637	0.620

Table 38. Frequency with which the reference HC_5 is greater than lower 95% confidence limit based on published extrapolation constants at N = 8, and ALC MDRs are met for **carbaryl**.

		Fitted Distribution		
		normal	logistic	triangular
Generating Distribution	normal	0.971	0.969	0.935
	logistic	1.000	0.999	0.992
	triangular	0.921	0.932	0.926

Table 39. Frequency with which the reference HC_5 is greater than lower 95% confidence limit based on published extrapolation constants at N = 3, but not necessarily typical OPP species for **carbaryl**.

		Fitted Distribution		
		normal	logistic	triangular
Generating Distribution	normal	0.783	0.787	0.757
	logistic	0.863	0.846	0.837
	triangular	0.545	0.775	0.733

3.6. How do different candidate distributions perform relative to each other?

Several patterns of variation among distributions emerged in the previous sections, which are restated here. First, the normal, logistic, and triangular distributions tended to perform very similarly to each other over the long run (10,000 simulated data sets, or 1,000 random draws of empirical data). This pattern generally reflected a qualitative difference between distributions fit to $\log_{10}$-transformed data and those fit to untransformed data. The Gumbel distribution ($\log_{10}$-transformed data) was more similar to the other distributions on transformed data, but was an outlier in that group. Distributions on untransformed data tended to be much more biased than distributions on transformed data especially at smaller sample sizes. At all sample sizes considered, it was often the case that the least biased distribution was not the same distribution as the generating distribution.

The most conspicuous effect of distribution concerned the tendency for very large positive biases in the HC_5 at limited sample sizes in the simulation data when distributions were fit to untransformed data, especially when maximum likelihood was used for parameter estimation. Thus, at with limited data, fitting distributions to untransformed data tends to result in over-estimates of the HC_5. A similar pattern was observed relative to the reference values when random subsets of data of

size n = 8 (GMAVs) or n = 3 (SMAVs) were analyzed. Part of this result is likely due to uncertainty associated with the numerical fitting algorithms applied to the untransformed data (with very large variance).

The ambiguous results observed above concerning distributional performance in analyses of the simulation data were also evident in analyses of the empirical data. Regardless of whether the ALC MDRs were imposed, AIC_c weights across distributions were remarkably evenly distributed for analyses of 12 AChE inhibitors (Tables 40, 41). The Burr distribution tended to get slightly less weight than the other six distributions probably due to the fact that it requires three estimated parameters rather than two. The evenness of distributional performance did not depend on whether or not the ALC MDRs were imposed (Tables 40, 41).

Table 40. Average AIC_c weights for GMAVs when n = 8 and ALC MDRs are met

	water	normal	logistic	triangular	Gumbel	Pareto	Weibull	Burr
malathion	FW	0.15	0.16	0.16	0.14	0.17	0.13	0.08
diazinon	FW	0.15	0.13	0.17	0.14	0.21	0.16	0.03
chlorpyrifos	FW	0.16	0.13	0.16	0.16	0.21	0.13	0.07
fenitrothion	FW	0.15	0.15	0.19	0.14	0.15	0.17	0.05
methyl parathion	FW	0.16	0.13	0.16	0.13	0.21	0.15	0.06
dichlorvos	FW	0.14	0.15	0.19	0.12	0.13	0.18	0.09
chlorpyrifos	SW	0.15	0.14	0.13	0.16	0.22	0.11	0.10
fenthion	SW	0.16	0.12	0.21	0.09	0.12	0.23	0.06
carbaryl	FW	0.15	0.13	0.17	0.14	0.19	0.18	0.04
carbaryl	SW	0.15	0.11	0.19	0.11	0.23	0.21	0.01
methomyl	FW	0.15	0.12	0.21	0.12	0.25	0.15	0.01
propoxur	FW	0.16	0.13	0.20	0.19	0.23	0.07	0.01

Table 41. Average AIC_c weights for GMAVs when n = 8 without regard to ALC MDRs

	water	normal	logistic	triangular	Gumbel	Pareto	Weibull	Burr
malathion	FW	0.15	0.16	0.17	0.13	0.17	0.14	0.09
diazinon	FW	0.14	0.14	0.16	0.13	0.16	0.16	0.10
chlorpyrifos	FW	0.16	0.14	0.15	0.13	0.19	0.14	0.09
fenitrothion	FW	0.14	0.13	0.18	0.13	0.15	0.16	0.11
methyl parathion	FW	0.13	0.15	0.16	0.14	0.17	0.15	0.10
dichlorvos	FW	0.16	0.15	0.16	0.13	0.16	0.16	0.09
chlorpyrifos	SW	0.14	0.12	0.15	0.15	0.19	0.12	0.12
fenthion	SW	0.16	0.12	0.17	0.15	0.14	0.17	0.08
carbaryl	FW	0.16	0.15	0.14	0.16	0.15	0.17	0.06
carbaryl	SW	0.16	0.11	0.20	0.10	0.22	0.20	0.01
methomyl	FW	0.16	0.12	0.20	0.13	0.17	0.19	0.03
propoxur	FW	0.16	0.15	0.16	0.14	0.14	0.16	0.07

In contrast to the ambiguous results above AIC_c did show good discriminatory power when applied to maximal sets of GMAVs (Table 42). At these sample sizes (ranging from 12 to 57, Table 2) AIC_c placed the greatest weight (on average) on the triangular and Weibull distributions, though there was also considerable variation among chemicals as to which distribution performed best (Table 42). Note however, that this result is due largely to the lack of replication inherent in this comparison, whereas previous results represented average AIC weights over 1,000 subsamples.

Table 42. Summary of AIC_c weights by chemical and distribution for all GMAVs.

Chemical	Water	normal	logistic	triangular	Gumbel	Pareto	Weibull	Burr
malathion	FW	0.10	0.02	0.84	0.00	0.00	0.03	0.01
diazinon	FW	0.07	0.02	0.16	0.00	0.00	0.74	0.01
chlorpyrifos	FW	0.11	0.03	0.62	0.11	0.09	0.02	0.03
fenitrothion	FW	0.01	0.01	0.00	0.00	0.00	0.97	0.01
methyl parathion	FW	0.05	0.01	0.09	0.00	0.00	0.84	0.01
dichlorvos	FW	0.12	0.09	0.37	0.01	0.00	0.39	0.03
chlorpyrifos	SW	0.04	0.04	0.02	0.26	0.59	0.00	0.05
fenthion	SW	0.15	0.09	0.18	0.02	0.01	0.48	0.07
carbaryl	FW	0.00	0.00	0.00	0.00	0.00	1.00	0.00
carbaryl	SW	0.15	0.10	0.22	0.06	0.12	0.33	0.01
methomyl	FW	0.15	0.08	0.28	0.03	0.01	0.41	0.03
propoxur	FW	0.24	0.17	0.29	0.15	0.03	0.09	0.03
Average		0.10	0.05	0.26	0.05	0.07	0.44	0.02

3.7. Does model-averaging across distributions improve estimates of the HC_5?

In analyses of the parametric simulation data, model-averaging inflated bias, often substantially (Fig. 11, compare to Fig. 2). However, model-averaging had little effect on variance in estimates of the HC_5 from the 10,000 replicates of the parametric simulation data (Fig. 12). As above, both bias and variance, declined sharply with sample size.

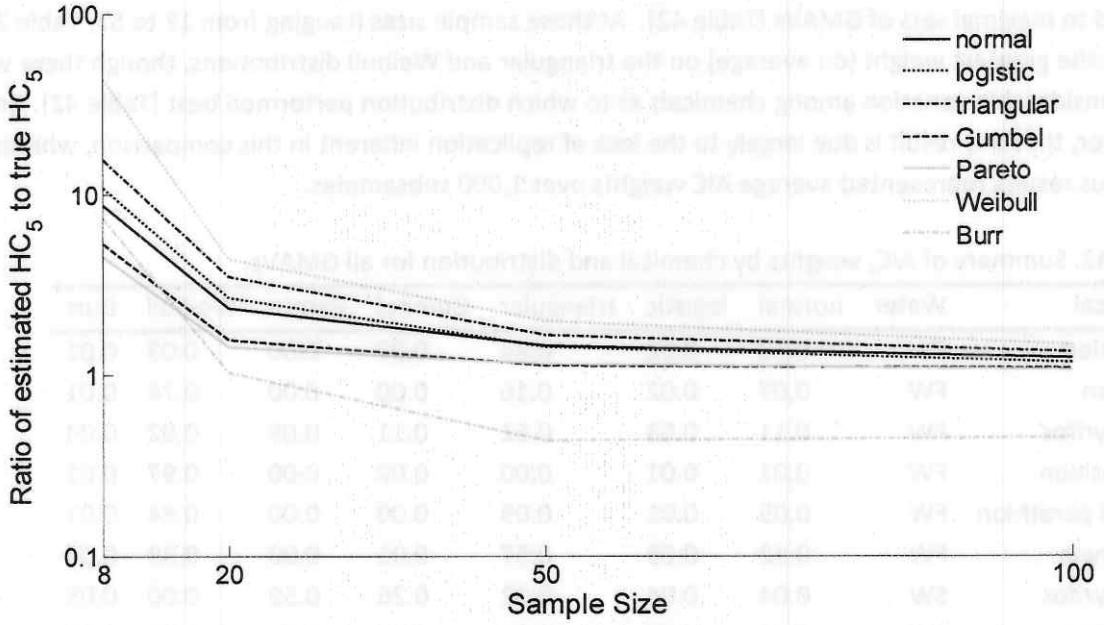


Figure 11. Bias in 10,000 replicates of model-averaged HC_5 estimates. Legend refers to generating distribution.

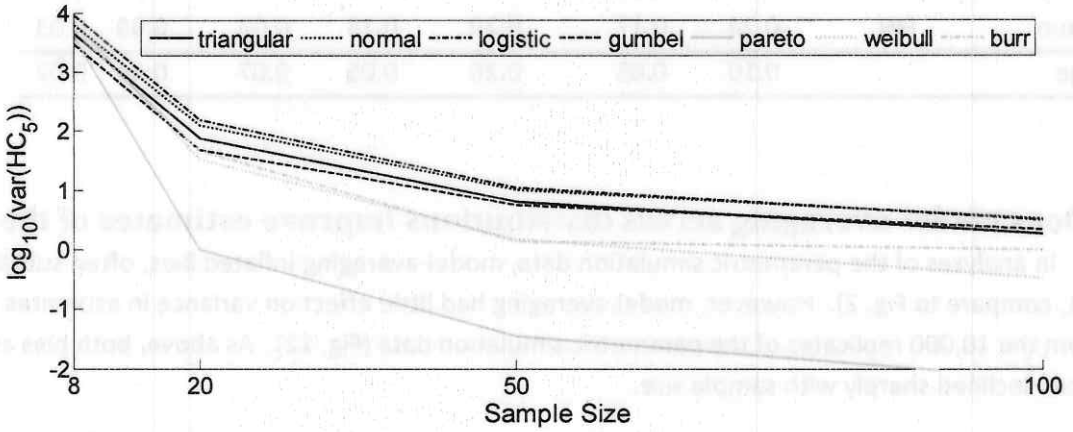


Figure 12. $\log_{10}$(variance) in 10,000 replicates of model-averaged HC_5 estimates. Legend refers to generating distribution.

In contrast to the results on simulated data, AIC_c model averaging results on HC_5s estimated from empirical data were probably the most interesting and surprising of all. The most surprising pattern was the ability of model-averaging to insulate the estimated HC_5 from effects of poor fitting and stochastic subsampling. For example, Table 43 shows the average estimated HC_5s under each distribution for 1,000 sets of 8 GMAVs subsampled from the malathion data without regard to the ALC MDRs. In particular, the first row of this table gives the estimated HC_5s for maximum likelihood analyses and the last column gives the AIC_c model-averaged value. Note that this value is much lower than the averages of all the distributions, and within an order of magnitude of the value estimated from the

maximal set of GMAVs (Section 8). A similar effect was observed for datasets of size n = 8 with the ALC MDRS met (Table 44). Further, this pattern was repeated across chemicals (*e.g.*, fenitrothion, Table 45). This result suggests that formal model-averaging across distributions can help insulate against outlier values (though data sets that contain only large values cannot produce a low HC_5).

Table 43. Mean estimated **malathion** HC_5s when n = 8 GMAVs regardless of MDRs

Fitting Method	normal	logistic	triangular	Gumbel	Weibull	Pareto	Burr	**average**
Maximum Likelihood	16.2	18.4	22.9	21.8	8.5	30.2	23.3	**4.7**
Moment Estimators	15.4	16.6	14.6	30.5		26.5		
Graphical Methods	6.0	4.1	7.6	15.5				

Table 44. Mean estimated **malathion** HC_5s when n = 8 GMAVs and ALC MDRs are met

Fitting Method	normal	logistic	triangular	Gumbel	Weibull	Pareto	Burr	**average**
Maximum Likelihood	1.7	1.3	2.2	2.4	0.4	2.6	2.1	**1.4**
Moment Estimators	1.1	1.2	1.0	3.9		2.1		
Graphical Methods	0.3	0.2	0.4	1.2				

Table 45. Mean estimated **fenitrothion** HC_5s when n = 8 GMAVs regardless of MDRs

Fitting Method	normal	logistic	triangular	Gumbel	Weibull	Pareto	Burr	**average**
Maximum Likelihood	21.35	34.66	24.75	22.11	14.21	25.38	19.09	**0.51**
Moment Estimators	29.57	16.54	23.18	20.30	17.35			
Graphical Methods	16.79	15.18	11.55	11.75				

4. Implications of these results for the use of SSDs in estimating the HC_5

For brevity, the above results have been stated with relatively little interpretation. Below the primary conclusions with respect to the estimation of HC_5s are presented as bulleted points. As above, the sections are organized around the first seven questions from the Objectives section (1.4 above). The eighth question, concerning future work, is addressed in Section 5 below.

4.1. Implications of sample size results for estimating the HC_5

1. When SSDs are fit to samples within the range of concern for this project ($3 \le n \le 8$), bias is likely to be large. However, when distributions are fit to $\log_{10}$-transformed data, bias was often less than an order of magnitude. Within the range of interest, direction of bias depended primarily on which estimation method was used (see 4.3 below).
2. When SSDs are fit to sample sizes less than 50 variance is likely to be very large.
3. The above two points apply to all three fitting methods, but are most worrisome with maximum likelihood.
4. To control large biases and variances with limited data, additional constraints must be imposed on the analyses (*e.g.*, minimum data requirements – see next section).

4.2. Implications of MDRs for estimating the HC_5

1. Imposition of ALC MDRs and limitation of analysis to $n = 8$ often introduces a small negative bias (*i.e.*, within an order of magnitude or less) compared to the HC_5 derived from a maximal set of GMAVs for a given chemical.
2. Analysis restricted to the three typical OPP species also introduces a small negative bias compared to the HC_5 derived from a maximal set of GMAVs for a given chemical.
3. Imposition of ALC MDRs greatly reduces the variance in the HC_5 estimated from random subsets of data. This occurs for two reasons. First, data samples are forced to include both sensitive and insensitive species. Second, there are fewer potential data combinations.
4. The latter result is partly due to a reduction in the number of possible data combinations.
5. Although the imposition of MDRs often produces, on average, HC_5s that differed only modestly from similar sample sizes with no data requirements, particular results from given iterations differed substantially.

4.3. Implications of choice of estimation method for estimating the HC_5

1. Within the sample-size range of interest for this project ($3 \leq n \leq 8$), maximum likelihood methods appear to be positively biased.
2. Positive bias with maximum likelihood can often be quite large, but is minimized with distributions on untransformed data.
3. Moment estimators were generally the least biased method.
4. Graphical methods, on average, produced the lowest values of the HC_5.
5. Variance in the estimated HC_5 did not differ much between methods.

4.4. Implications of performance of fit measures for estimating the HC_5

1. AIC_c is not likely to be useful for discriminating among competing distributions at the sample sizes of interest for this project ($3 \leq n \leq 8$).
2. Goodness-of-fit tests are not likely to be useful for rejecting alternative distributions the sample sizes of interest for this project.
3. Both AIC_c and parametric bootstrap goodness-of-fit tests would be quite useful at larger sample sizes.
4. Empirical bootstrap sampling showed little utility at any sample size.
5. Goodness-of-fit tests applied to the Pareto distribution fit using moment estimators suggest that those estimators are not accurate.

4.5. Implications of the performance of 95% confidence limits for estimating the HC_5

1. Extrapolation constants produce confidence limits at nominal coverage rates when data are simulated from a known distribution.
2. In empirical analyses, the HC_5 estimated using all available data was lower than the 95% confidence limit based on subsamples of $n = 3$ and $n = 8$ with no data requirements more often than expected. True coverage rates were between 60% and 90%. Thus, this procedure might result in a lower confidence that the number of species at risk was 5% or less. Conversely, it

could be argued that the reference value (the HC_5 estimated using all available data) is itself a biased estimator of the true HC_5 if the sample of tested species is biased towards insensitive species.

3. In contrast to (2) above, the HC_5 estimated using all available data was greater than the 95% confidence limit based on subsamples of $n = 3$ and $n = 8$ when data requirements were imposed more often than expected. True coverage rates were very close to 1. Thus, this procedure might result in over-protection.

4.6. Implications of choice of distribution for estimating the HC_5

1. In general, distributions on transformed data performed better than those on untransformed data.
2. Of the distributions on transformed data, the triangular distribution performed best across all 9 organophosphates.
3. Of the distributions on untransformed data, the $Burr_{III}$ distribution seemed to perform best.
4. Regardless of points 1 and 2 above, there remains considerable uncertainty about which distribution is 'best' and the results do not support the exclusive use of any one distribution.

4.7. Implications of multidistributional inference for estimating the HC_5

1. Calculating the average HC_5 across distributions for simulated data did not help produce a better estimate of the HC_5.
2. In contrast to (1), calculating the AIC_c model-averaged HC_5 for empirical data sets greatly reduced bias in estimates of the HC_5, even when, on average, every distribution was biased.
3. The reasons for the different conclusions resulting from simulated versus empirical data are not clear and should be studied further before a distribution-averaging approach is adopted.

5. What future research is suggested by these results?

The results of analyses of both simulation and empirical analyses indicate several endemic problems with fitting species sensitivity distributions. While all fitting methods tend to be unbiased, especially with large sample sizes, large variability hampers our ability to make non-trivial statements about the location of the HC_5 with limited data. Further, no single distribution reliably performs best even for analyses of random subsets of test results on the same chemical. Another problem (not explored in the above results) is that goodness-of-fit of a full distribution may not be a reliable indicator of goodness-of-fit in the region of the HC_5. The latter two points seem to argue in favor of different fitting methods, with the goodness-of-fit issue suggesting graphical methods, which can be weighted toward the region of the HC_5, whereas the variable performance of distributions argues for the use of maximum likelihood whereby distributional performance can be easily compared and formal averaging techniques may be used across distributions. The following section describes how various extensions of maximum likelihood can be brought to bear on all of these issues.

5.1. Hierarchical models

A hierarchical model, in modern statistical usage, is one in which a model parameter is specified to be (in turn) the prediction of some further model. For example, de Zwart (2002) developed a hierarchical model of the logistic distribution parameter β upon chemical mode of action. He also provided a table of estimated β values for use in fitting log-logistic SSDs, if a chemical's mode of action is known (de Zwart 2002: Table 8.4). In that case all available data for a given chemical may be used to estimate α (the other logistic distribution parameter) and the value of β can be assumed known. Similarly, interspecies correlation extrapolation (ICE) models exploit correlations among species test results across previously tested species to predict the toxicity for untested species based on values for tested species (Raimondo et al. 2010). These predicted values can then be used as surrogate test results in the development of SSDs (Appendix B). These methods are vulnerable to the criticism that they do not properly take into account the uncertainty associated with the lower-level model predictions. Furthermore, preliminary analyses of de Zwart's (2002) method using different data suggest considerable disagreement among estimates of β, especially with respect to the AChEI MOA (see Section 8.9 in Supplementary Results). Thus caution is warranted in consideration of adoption of this method. Modern hierarchical model fitting methods attempt to simultaneously fit all hierarchically linked models and to properly account for the associated model variance. Hierarchical models, in the context of SSDs, could be specified in many ways, which are briefly outlined below.

The most important limitation to the use of SSDs is sample size, which could lead to large biases, substantial uncertainty, and uninformative measures of fit. One way to bring more data to the table would be to investigate methods for "borrowing" information on toxicity from related chemicals, for example using information on mode of action (de Zwart 2002) or adverse outcome pathway (AOP, Ankley et al. 2010). Such a model might look very much like de Zwart's (2002) regression model, but the regression would be fit simultaneously with the SSD, using numerical optimization methods and maximum likelihood. The resulting model might be able to exploit some of the many benefits of maximum likelihood estimation (Section 1.2.1), including the ability to use formal model selection methods (Section 1.3.3), without suffering the large biases and variance associated with limited data, if the pool of available data for a given mode of action is large. Furthermore, the resulting model could be updated as new test results become available.

Another common feature in modern hierarchical models is the specification of random effects (*i.e.*, a model parameter is itself assumed to be a random variate from some specified distribution). An obvious application for this idea in SSD methodology would be to treat multiple test results for a given taxon as random draws from a possible distribution of test values for that taxon. This would help with concerns about how to summarize multiple test results for a given taxon (for example, whether to use the arithmetic mean versus geometric mean versus lowest test result). Furthermore, it would implicitly incorporate the natural statistical notion that multiple results for a taxon are a data strength, rather than a weakness (as variable results within taxa are commonly viewed, Etterson pers. obs.).

The above two ideas are reinforced by the observed result in this report that SSD analysis with limited data benefits enormously from the imposition of minimum data requirements (see Section 4.2 above and associated sub-sections in Results). Thus different taxonomic groups, both at the species and genus level, exhibit markedly different sensitivities to a given chemical. It is natural to assume that these

sensitivities are often phylogenetically conserved and this assumption is reinforced from studies showing similar sensitivities among related taxa for a given mode of action (de Zwart 2002). In this case, it may be the case that related species share their sensitivity by sharing an adverse outcome pathway triggered by the chemical. Knowledge of relationships among species (http://tolweb.org) and methods for modeling the distribution of character traits (toxicity test results in our context) on phylogenetic trees (see for example Huelsenbeck 2003) have increased greatly since the publication of the 1985 Guidelines (USEPA 1985). Use of phlyogenetic structure in SSD analyses is likely to be more useful with highly specific modes of action (such as acetylcholinesterase inhibition analyzed in this report) compared to more general modes of action such as narcosis. Development of the methods described in this paragraph would require a substantial commitment of time and research, but with a potentially large benefit to our understanding of the distribution of toxicity across taxa.

5.2. Order statistics

Graphical methods for SSDs are an example of a common technique in the branch of statistics referred to as the theory of order statistics. Their primary advantages, as reviewed above, are the ability to weight the fit in the region of the HC_5 (or any other quantile of interest) and the ability to handle unbounded test results. Their primary disadvantages are that they are limited to distributions in the location/scale family (in our analyses these are normal, logistic, and triangular) and they do not permit the use of formal model selection techniques. However, given an ordered set of toxicity test results it is possible to develop relatively simple expressions for the likelihood (under any continuous distribution) for the ordered data that are flexible in the amount of weight applied to any given region of the likelihood. This theory is briefly described below.

Suppose that n test results are available for a given chemical and let x_i represent the i^{th} ordered EC50 ($1 \le i \le n$), where the results are ordered from lowest to highest. Further, let $f(x)$ and $F(x)$ represent the probability density and cumulative density functions for the particular distribution from which the (unordered) data are assumed to arise. Then the pdf for the ordered data (f_o), weighted toward the first i order statistics is:

$$f_o(\mathbf{X}|\boldsymbol{\theta}) \propto \left(1-F(x_i)\right)^{n-i} \prod_{j=1}^{i} f(x_j)$$

In particular, suppose n = 100 and it is considered desirable to give particular weight to the first ten order statistics corresponding to the lower tenth quantile of the empirical cdf. These are $x_1, x_2, x_3, \ldots x_{10}$. The remaining 90 order statistics are assumed to fall in the upper 90th quantile of the distribution. The joint probability density, $f_o(x)$, for the order statistics is proportional to:

$$f_o(\mathbf{X}|\boldsymbol{\theta}) \propto \left(1-F(x_{10})\right)^{90} \prod_{j=1}^{10} f(x_j)$$

Like the graphical methods described by Erickson and Stephan (1988), the pdf above uses all available data, but weights the fit more strongly to the lower 10% of the data. The upper 90 order statistics are simply assumed to fall within the upper 90th quantile of the distribution, but their specific value is not used. Obviously, any unbounded data that fall within this upper quantile may be accommodated by simply counting them in the exponent to $1 - F(x_{10})$.

5.3. Bayesian methods

Several features of Bayesian methods make them attractive for use in the analysis of SSDs. First, in attempts to understand the toxicity of a chemical, there may be data available about the toxicity of related chemicals, which could serve as valuable prior information in statistical analyses. Or, data from multiple chemicals could be pooled and analyzed using hierarchical models that include either fixed or random effects of a specific chemical. The latter methods, especially using some of the complicated hierarchical structures described immediately above, are much easier to fit using Bayesian methods. Finally, it is often desirable to make some kind of probabilistic statement about the location of the HC_5. However, under a classical statistical framework, the HC_5 is fixed (but unknown) resulting in cumbersome interpretations of confidence intervals and the like. Under a Bayesian framework the HC_5 would be viewed as a random variable with an estimable posterior distribution from which probabilistic statements could be formulated.

6. Acknowledgments

This report benefitted greatly from comments by R. Erickson, M. Frankenberry, K. Garber, C. Russom, and C. Stephan. K. Garber conducted and contributed the preliminary analysis on MOA-specific parameters for the logistic distribution (Section 8.9).

7. Literature cited

Aldenberg, T. and W. Slob. 1993. Confidence limits for hazardous concentrations based on logistically distributed NOEC toxicity data. Ecotoxicology and Environmental Safety 25:48-63.

Aldenberg, T. and J.S. Jaworska. 2000. Uncertainty of the hazardous concentration and fraction affected for normal species sensitivity distributions. Ecotoxicology and Environmental Safety 46:1-18.

Ankley, G.T, R.S. Bennett, R.J. Erickson, D.J. Hoff, M.W. Horning, R.D. Johnson, D.R. Mount, J.W. Nichols, C.L. Russom, P.K. Schmieder, J.A. Serrano, J.E.Tietge, and D.L. Villeneuve. 2010. Adverse outcome pathways: a conceptual framework to support ecotoxicology research and risk assessment. Environmental Toxicology and Chemistry 29(3):730–741.

Arnold, B.C., N. Balakrishnan, and H.H. Nagaraja. 2008. A first course in order statistics. Society for Industrial and Applied Mathmatics. Philadelphia, PA, USA.

ASTM. 2011. Standard guide for conducting acute toxicity tests on test materials with fishes, macroinvertebrates, and amphibians. Designation E729-96. Pp. 67-88 in Annual Book of ASTM Standards Volume 11.06. ASTM International. West Conshohocken, PA, USA.

Burnham, K.P. and D.R. Anderson. 2002. Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach, 2nd Ed.

Campbell, E., M. Palmer, Q. Shao, and D. Wilson. 2000. BurrliOZ. CSIRO, Mathematical and Information Sciences. www.cmis.csiro.au/envir/burrlioz.

Chapman, P.F., M. Reed, A. Hart, et al. (2007). Methods of Uncertainty Analysis. Work Package 4 in EUFRAM: Concerted action to develop a European framework for probabilistic risk assessment of the environmental impacts of pesticides. Version 3, March 2007. http://www.eufram.com/, accessed 8 August 2011.

de Zwart, D. 2002. Observed regularities in SSDs for aquatic species. *In* Posthuma, L. and G.W. Suter (Eds.), The use of Species Sensitivity distributions in Ecotoxicology. SETAC Press, Pensacola, FL.

Edwards, A.W.F. 1992. Likelihood: Expanded edition. Johns Hopkins University Press. Baltimore, MD, USA.

Efron, B. and R.J. Tibshirani. 1994. An Introduction to the Bootstrap. Chapman and Hall/CRC. Boca Raton, FL. USA.

Elphick, J.R.F., K.D. Bergh, and H.C. Bailey. 2011. Chronic toxicity of chloride to freshwater species: effects of hardness and implications for water quality guidelines. Environmental Toxicology and Chemistry 30:239-246.

Erickson, R.J. and C.E. Stephan. 1988. Calculation of the final acute value for water quality criteria for aquatic organisms. U.S. Environmental Protection Agency. Duluth, MN. EPA/600/3-88-018.

Etterson, M.A. and L.R. Nagy. 2008. Is mean squared error a consistent indicator of model accuracy for spatially structured demographic models. Ecological Modelling 211:202-208.

Garber, K. S. Raimondo, and P. TenBrook. 2010. Exploration of methods for characterizing effects of chemical stressors to aquatic animals. Working paper draft 11/2/2010.

Harville, D.A. 1977. Maximum likelihood approaches to variance component estimation and to related problems. Journal of the American Statistical Association 72:320-340.

Host, G.E., R.R. Regal, and C.E. Stephan. 1995. Analyses of acute and chronic data for aquatic life. U.S. Environmental Protection Agency.

Huelsenbeck, J.P., R. Nielsen, and J.P. Bollback. 2003. Stochastic mapping of morphological characters. Systematic Biology 52(2):131–158.

Kooijman, S.A.L.M. 1987. A safety factor for LC_{50} values allowing for differences in sensitivity among species. Water Research 21:269-276.

Manly, B.F.J. 1997. Randomization, Bootstrap, and Monte Carlo Methods in Biology, 2nd Ed. Chapman & Hall/CRC, Boca Raton, FL.

Mathworks. 2010. Matlab; the language of technical computing. Mathworks, Inc. Natick, MA. USA.

Newman, M.C., D.R. Ownby, L.C.A. Mézin, D.C. Powell, T.R.L Christensen, S.B. Lerberg, and B.-A. Anderson. 2000. Applying species sensitivity distributions in ecological risk assessment: assumptions of distribution type and sufficient numbers of species. Environmental Toxicology and Chemistry 19:508-515.

Pennington, D.W. 2003. Extrapolating ecotoxicological measures from small data sets. Ecotoxicology and Environmental Safety 56:238-250.

Posthuma, L, T.P. Traas, and G.W. Suter. 2002. General introduction to species sensitivity distributions. Pp. 3-10 *in* L. Posthuma, G.W. Suter, and T.P. Traas (Eds.) Species Sensitivity Distributions in Ecotoxicology. Lewis. Boca Raton, FL. USA.

Raimondo, S., D.N. Vivian, and M.G. Barron. 2010. Web-based Interspecies Correlation Estimation (Web-ICE) for Acute Toxicity: User Manual. Version 3.1. EPA/600/R-10/004. Office of Research and Development, U. S. Environmental Protection Agency. Gulf Breeze, FL.

Seber, G.A.F. 2002. The estimation of animal abundance and related parameters, 2nd Ed. Reprint of 1982 edition. Blackburn Press. Caldwell, NJ. USA.

Shao, Q. 2000. Estimation for hazardous concentrations based on NOEC toxicity data: an alternative approach. Environmetrics 11:583-595.

Shao, Q. 2003. Notes on maximum likelihood estimation for the three-parameter Burr XII distribution. Computational Statistics & Data Analysis 45:675-687.

U.S. EPA. 1985. Guidelines for deriving numerical national water quality criteria for the protection of aquatic organisms and their uses. U.S. Environmental Protection Agency. Duluth, MN PB85-227049 or EPA 822/R-85-100.

Suter, G.W. 1998. Comments on the interpretation of distributions in "Overview of recent developments in ecological risk assessment." Risk Analysis 18:3-4.

Suter, G.W. 2002. North American history of species sensitivity distributions. Pp. 11-18 *in* L. Posthuma, G.W. Suter, and T.P. Traas (Eds.) Species Sensitivity Distributions in Ecotoxicology. Lewis. Boca Raton, FL. USA.

Stephan, C.E. 2011. Clarification of the guidance given in the 1985 guidelines concerning evaluation of results of toxicity tests with aquatic organisms. Sixth Draft. 5 October 2011.

TenBrook, P. L., R.S. Tjeerdema, P. Hann, and J. Karkoski. 2008. Reviews of Environmental Contamination and Toxicology 199, DOI: 10.1007/978-0-387-09808-1.

TenBrook PL, Palumbo AJ, Fojut TL, Hann P, Karkoski J, Tjeerdema RS. 2010. The University of California-Davis Methodology for Deriving Aquatic Life Pesticide Water Quality Criteria. Reviews of Environmental Contamination and Toxicology. Volume 209.

van Straalen, N.M and C.A.J. Denneman. 1989. Ecotoxicological evaluation of soil quality criteria. Ecotoxicology and Environmental Safety 18:241-251.

van Straalen, N.M. 2002. Threshold models for species sensitivity distributions applied to aquatic risk assessment for zinc. Environmental Toxicology and Chemistry 11:167-172.

Williams, B.K, J.D. Nichols, and M.J. Conroy. 2002. Analysis and Management of Animal Populations. Academic Press. San Diego, CA. 817 Pp.

Zajdlik & Associates. 2005. Statistical analysis of the SSD approach for development of Canadian Water Quality Guidelines. Canadian Council of Ministers of the Environment. Project #354-2005.

8. Supplementary results

This section provides full results of SSD analyses not presented in text.

8.1. HC_5s estimated using all available GMAVs

Table S8.1.1. Estimated HC_5s for **malathion** using all GMAVs for freshwater taxa (n = 57)

Fitting Method	normal	logistic	triangular	Gumbel	Pareto	Weibull	Burr	average
Maximum Likelihood	3.318	2.554	3.819	4.321	1.556	0.588	2.330	3.639
Moment Estimators	3.165	3.394	2.871	9.527	1.181			
Graphical Methods	2.355	2.126	2.453					

Table S8.1.2. Estimated HC_5s for **diazinon** using all GMAVs for freshwater taxa (n = 28)

Fitting Method	normal	logistic	triangular	Gumbel	Pareto	Weibull	Burr	average
Maximum Likelihood	0.912	0.780	1.248	0.978	0.546	0.607	0.481	0.733
Moment Estimators	0.820	0.886	0.736	2.777	0.408			
Graphical Methods	0.468	0.380	0.527					

Table S8.1.3. Estimated HC_5s for **chlorpyrifos** using all GMAVs for freshwater taxa (n = 30)

Fitting Method	normal	logistic	triangular	Gumbel	Pareto	Weibull	Burr	average
Maximum Likelihood	0.045	0.026	0.078	0.091	0.086	0.005	0.090	0.074
Moment Estimators	0.041	0.044	0.037	0.130	0.070			
Graphical Methods	0.025	0.021	0.028					

Table S8.1.4. Estimated HC_5s for **fenitrothion** using all GMAVs for freshwater taxa (n = 30)

Fitting Method	normal	logistic	triangular	Gumbel	Pareto	Weibull	Burr	average
Maximum Likelihood	0.820	1.015	0.219	0.527	0.034	1.023	0.573	1.016
Moment Estimators	0.743	0.802	0.668	2.475	0.022			
Graphical Methods	0.439	0.362	0.491					

Table S8.1.5. Estimated HC_5s for **methyl parathion** using all GMAVs for freshwater taxa (n = 29)

Fitting Method	normal	logistic	triangular	Gumbel	Pareto	Weibull	Burr	average
Maximum Likelihood	1.070	0.864	1.548	1.222	0.529	0.824	0.556	0.904
Moment Estimators	0.964	1.043	0.863	3.339	0.389			
Graphical Methods	0.552	0.449	0.621					

Table S8.1.6. Estimated HC_5s for **dichlorvos** using all GMAVs for freshwater taxa (n = 22)

Fitting Method	normal	logistic	triangular	Gumbel	Pareto	Weibull	Burr	average
Maximum Likelihood	0.262	0.334	0.273	0.226	0.152	0.076	0.188	0.197
Moment Estimators	0.221	0.244	0.193	1.010	0.105			
Graphical Methods	0.096	0.070	0.116					

Table S8.1.7. Estimated HC_5s for **chlorpyrifos** using all GMAVs for saltwater taxa (n = 21)

Fitting Method	normal	logistic	triangular	Gumbel	Pareto	Weibull	Burr	average
Maximum Likelihood	0.016	0.012	0.011	0.052	0.057	0.000	0.048	0.051
Moment Estimators	0.014	0.015	0.013	0.045	0.048			
Graphical Methods	0.007	0.006	0.009					

Table S8.1.8. Estimated HC_5s for **fenthion** using all GMAVs for saltwater taxa (n = 12)

Fitting Method	normal	logistic	triangular	Gumbel	Pareto	Weibull	Burr	average
Fitting Method	normal	logistic	triangular	Gumbel	Pareto	Weibull	Burr	average
Maximum Likelihood	0.123	0.105	0.133	0.055	0.069	0.092	0.051	0.101
Moment Estimators	0.095	0.103	0.084	0.345	0.050			

Table S8.1.9. Estimated HC_5s for **carbaryl** using all GMAVs for freshwater taxa (n = 55)

Fitting Method	normal	logistic	triangular	Gumbel	Pareto	Weibull	Burr	average
Maximum Likelihood	9.840	13.966	9.096	7.759	3.311	9.707	8.487	9.707
Moment Estimators	9.442	10.017	8.693	24.025	2.578			
Graphical Methods	7.298	6.671	7.572					

Table S8.1.10. Estimated HC_5s for **carbaryl** using all GMAVs for saltwater taxa (n = 9)

Fitting Method	normal	logistic	triangular	Gumbel	Pareto	Weibull	Burr	average
Maximum Likelihood	4.767	4.265	8.756	5.410	7.821	3.318	3.028	5.538
Moment Estimators	3.552	3.799	3.233	10.266	6.505			
Graphical Methods	1.148	0.775	1.460					

Table S8.1.11. Estimated HC_5s for **propoxur** using all GMAVs for freshwater taxa (n = 13)

Fitting Method	normal	logistic	triangular	Gumbel	Pareto	Weibull	Burr	average
Maximum Likelihood	19.953	16.262	27.387	28.107	18.660	3.038	22.657	21.276
Moment Estimators	16.966	17.907	15.732	39.804	15.789			
Graphical Methods	8.480	6.605	9.910					

Table S8.1.12. Estimated HC_5s for **methomyl** using all GMAVs for freshwater taxa (n = 16)

Fitting Method	normal	logistic	triangular	Gumbel	Pareto	Weibull	Burr	average
Maximum Likelihood	72.219	69.515	80.708	73.932	67.928	55.055	55.982	66.886
Moment Estimators	67.535	69.425	64.980	104.406	63.608			
Graphical Methods	49.831	44.587	53.390					

8.2. HC_5s estimated using all available SMAVs

Table S8.2.1. Estimated HC_5s for **malathion** using all SMAVs for freshwater taxa (N = 71)

Fitting Method	normal	logistic	triangular	Gumbel	Pareto	Weibull	Burr	average
Maximum Likelihood	2.652	1.942	2.745	3.611	0.987	0.425	2.345	2.711
Moment Estimators	2.552	2.740	2.312	7.824	0.736			
Graphical Methods	1.981	1.823	2.028					

Table S8.2.2. Estimated HC_5s for **diazinon** using all SMAVs for freshwater taxa (N = 33)

Fitting Method	normal	logistic	triangular	Gumbel	Pareto	Weibull	Burr	average
Maximum Likelihood	0.858	0.729	1.224	0.946	0.544	0.468	0.447	0.661
Moment Estimators	0.784	0.847	0.704	2.644	0.408			
Graphical Methods	0.478	0.398	0.529					

Table S8.2.3. Estimated HC_5s for **chlorpyrifos** using all SMAVs for freshwater taxa (N = 34)

Fitting Method	normal	logistic	triangular	Gumbel	Pareto	Weibull	Burr	average
Maximum Likelihood	0.030	0.017	0.059	0.064	0.046	0.003	0.063	0.055
Moment Estimators	0.028	0.030	0.025	0.089	0.037			
Graphical Methods	0.017	0.015	0.019					

Table S8.2.4. Estimated HC_5s for **fenitrothion** using all SMAVs for freshwater taxa (N = 36)

Fitting Method	normal	logistic	triangular	Gumbel	Pareto	Weibull	Burr	average
Maximum Likelihood	0.725	0.721	0.219	0.552	0.034	0.711	0.417	0.708
Moment Estimators	0.668	0.721	0.601	2.222	0.022			
Graphical Methods	0.423	0.357	0.463					

Table S8.2.5. Estimated HC_5s for **methyl parathion** using all SMAVs for freshwater taxa (N = 35)

Fitting Method	normal	logistic	triangular	Gumbel	Pareto	Weibull	Burr	average
Maximum Likelihood	1.284	1.090	1.742	1.416	0.530	1.056	0.697	1.095
Moment Estimators	1.180	1.274	1.061	3.948	0.389			
Graphical Methods	0.738	0.621	0.810					

Table S8.2.6. Estimated HC_5s for **dichlorvos** using all SMAVs for freshwater taxa (N = 25)

Fitting Method	normal	logistic	triangular	Gumbel	Pareto	Weibull	Burr	average
Maximum Likelihood	0.090	0.076	0.153	0.116	0.106	0.015	0.043	0.111
Moment Estimators	0.077	0.085	0.066	0.406	0.074			
Graphical Methods	0.033	0.024	0.040					

Table S8.2.7. Estimated HC_5s for **chlorpyrifos** using all SMAVs for saltwater taxa (N = 26)

Fitting Method	normal	logistic	triangular	Gumbel	Pareto	Weibull	Burr	average
Maximum Likelihood	0.023	0.020	0.013	0.000	0.057	0.000	0.066	0.050
Moment Estimators	0.021	0.023	0.019	0.060	0.048			
Graphical Methods	0.013	0.010	0.014					

Table S8.2.8. Estimated HC_5s for **fenthion** using all SMAVs for saltwater taxa (N = 13)

Fitting Method	normal	logistic	triangular	Gumbel	Pareto	Weibull	Burr	average
Maximum Likelihood	0.042	0.030	0.055	0.051	0.035	0.023	0.011	0.035
Moment Estimators	0.032	0.035	0.029	0.132	0.025			
Graphical Methods	0.010	0.007	0.013					

Table S8.2.9. Estimated HC_5s for **carbaryl** using all SMAVs for freshwater taxa (N = 75)

Fitting Method	normal	logistic	triangular	Gumbel	Pareto	Weibull	Burr	average
Maximum Likelihood	7.778	14.610	0.242	3.079	0.020	10.006	8.185	10.006
Moment Estimators	7.535	8.016	6.911	20.005	0.012			
Graphical Methods	6.096	5.694	6.202					

Table S8.2.10. Estimated HC_5s for **carbaryl** using all SMAVs for saltwater taxa (N = 11)

Fitting Method	normal	logistic	triangular	Gumbel	Pareto	Weibull	Burr	average
Maximum Likelihood	9.009	11.896	12.545	7.871	7.971	10.065	7.030	10.412
Moment Estimators	7.203	7.670	6.597	19.427	6.505			
Graphical Methods	2.894	2.093	3.538					

Table S8.2.11. Estimated HC_5s for **propoxur** using all SMAVs for freshwater taxa (N = 13)

Fitting Method	normal	logistic	triangular	Gumbel	Pareto	Weibull	Burr	average
Maximum Likelihood	19.953	16.262	27.387	28.107	18.660	3.038	22.657	21.276
Moment Estimators	16.966	17.907	15.732	39.804	15.789			
Graphical Methods	8.480	6.605	9.910					

Table S8.2.12. Estimated HC_5s for **methomyl** using all SMAVs for freshwater taxa (N = 20)

Fitting Method	normal	logistic	triangular	Gumbel	Pareto	Weibull	Burr	average
Maximum Likelihood	46.739	51.295	36.482	41.573	19.909	33.557	39.461	37.875
Moment Estimators	43.920	45.354	41.993	72.910	17.567			
Graphical Methods	32.543	29.128	34.818					

8.3. Comparison plots of analyses at n = 8 (ALC MDRs imposed versus not imposed)

Notes: Figures in this series for **malathion** and **carbaryl** freshwater tests are Figures 3 and 4 in the main text. Panels represent a. maximum likelihood estimates, b. moment estimates, c. graphical estimates. Note log_{10} scale of Y-axes.

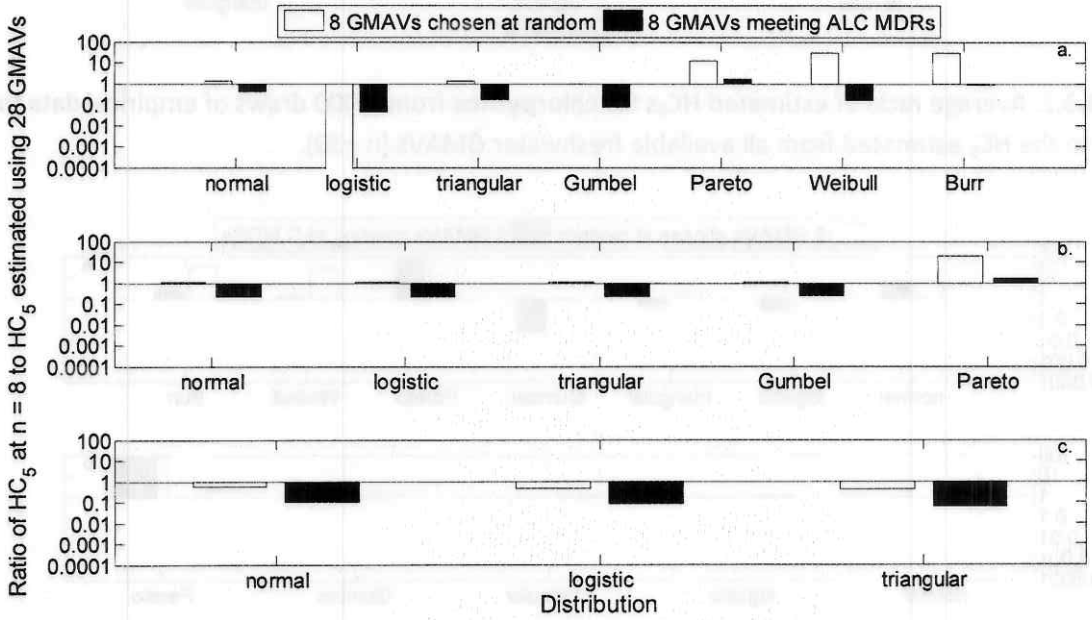


Figure S8.3.1. Average ratio of estimated HC_5s for **diazinon** from 1,000 draws of empirical data (n = 8 GMAVs) to the HC_5 estimated from all available freshwater GMAVs (n = 28).

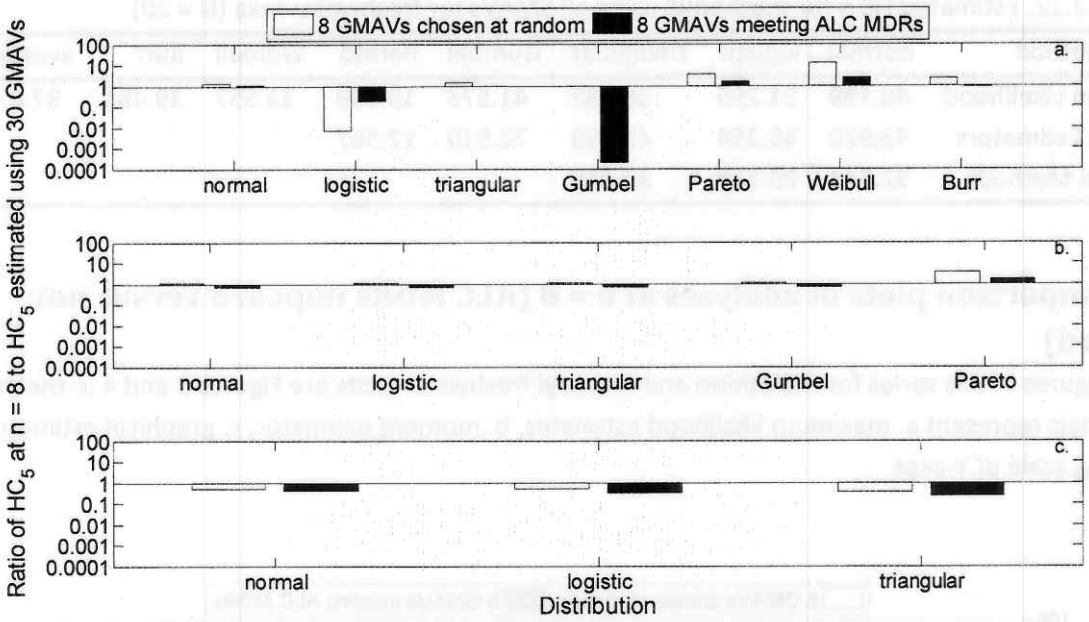


Figure S8.3.2. Average ratio of estimated HC_5s for **chlorpyrifos** from 1,000 draws of empirical data (n = 8 GMAVs) to the HC_5 estimated from all available freshwater GMAVs (n =30).

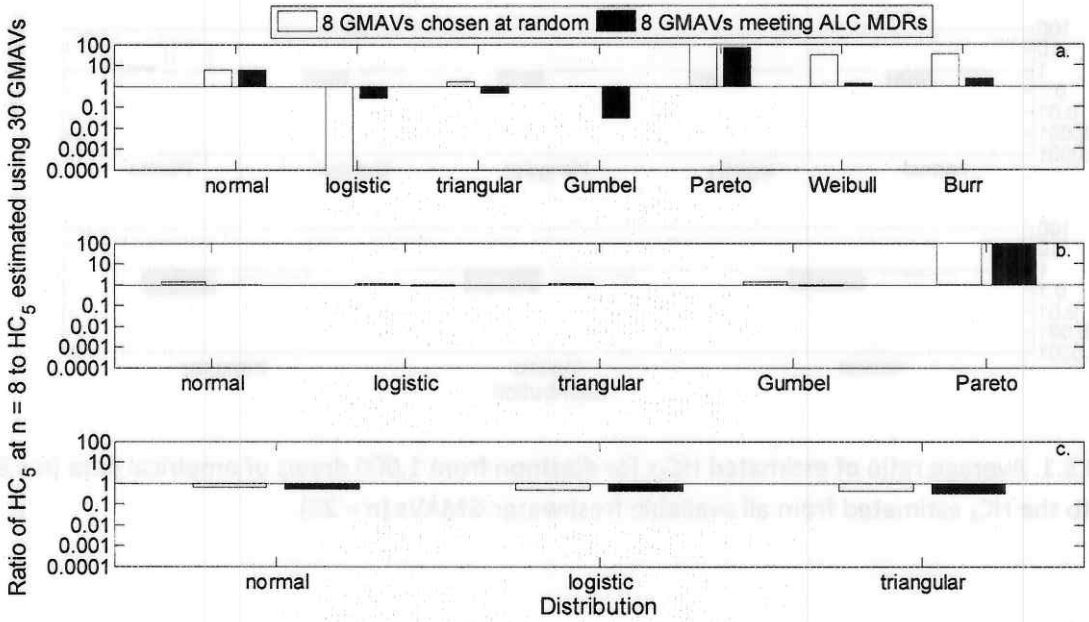


Figure S8.3.3. Average ratio of estimated HC_5s for **fenitrothion** from 1,000 draws of empirical data (n = 8 GMAVs) to the HC_5 estimated from all available freshwater GMAVs (n = 30).

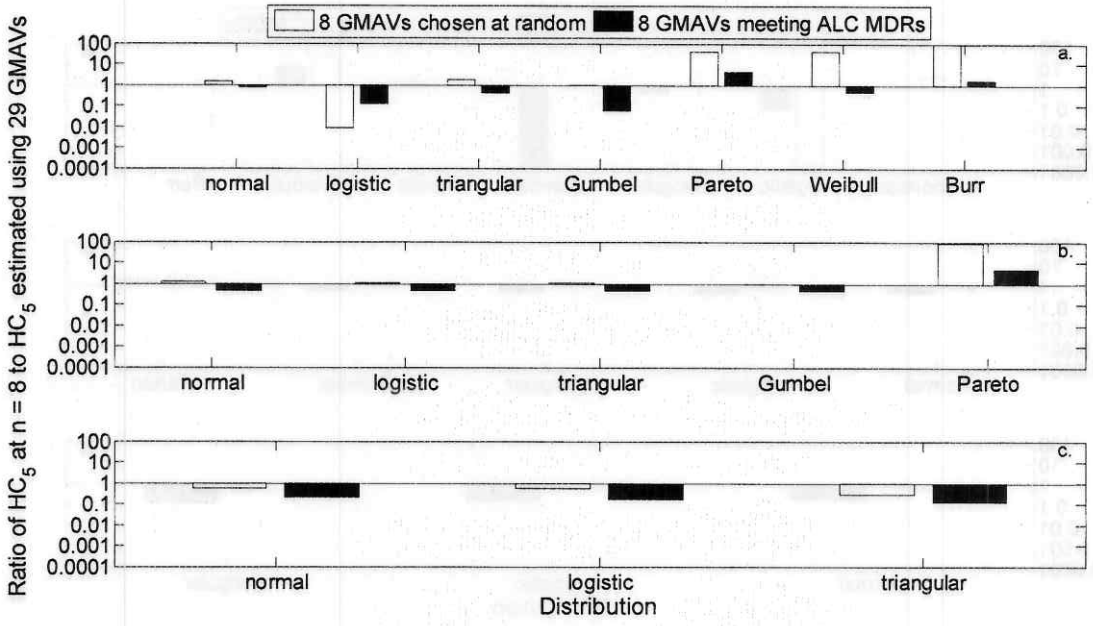


Figure S8.3.4. Average ratio of estimated HC_5s for **methyl parathion** from 1,000 draws of empirical data (n = 8 GMAVs) to the HC_5 estimated from all available freshwater GMAVs (n = 29).

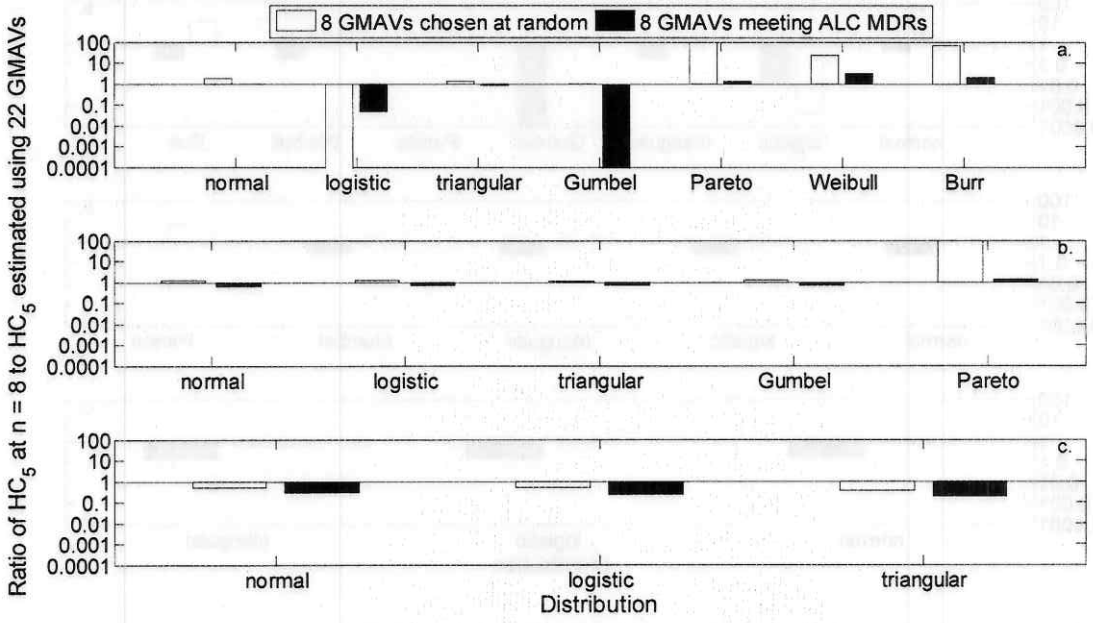


Figure S8.3.5. Average ratio of estimated HC_5s for **dichlorvos** from 1,000 draws of empirical data (n = 8 GMAVs) to the HC_5 estimated from all available freshwater GMAVs (n = 22).

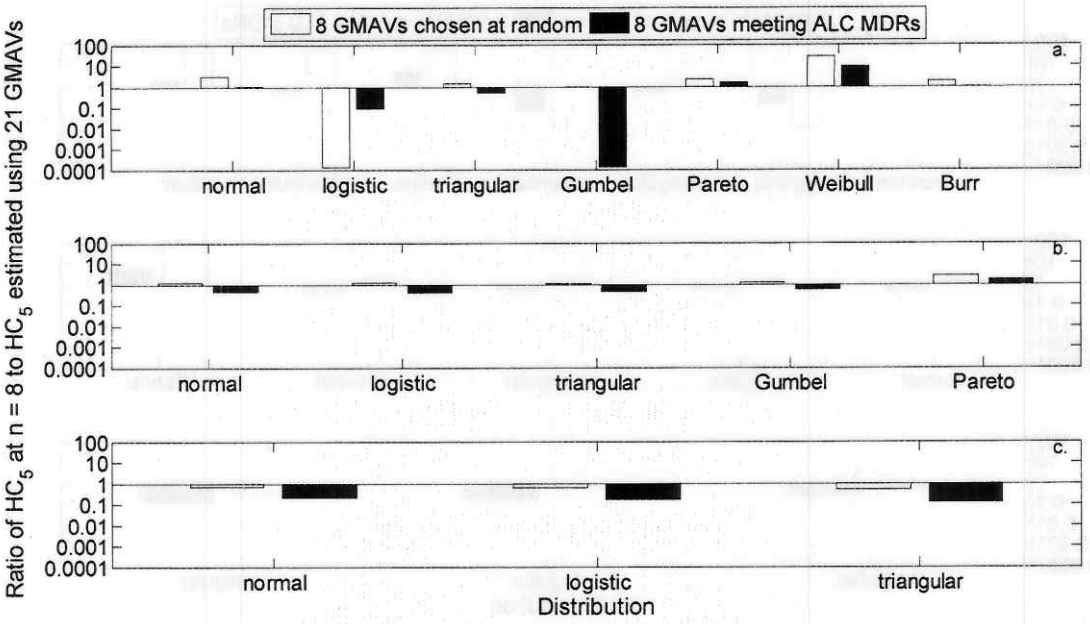


Figure S8.3.6. Average ratio of estimated HC_5s for **chlorpyrifos** from 1,000 draws of empirical data (n = 8 GMAVs) to the HC_5 estimated from all available saltwater GMAVs (n = 21).

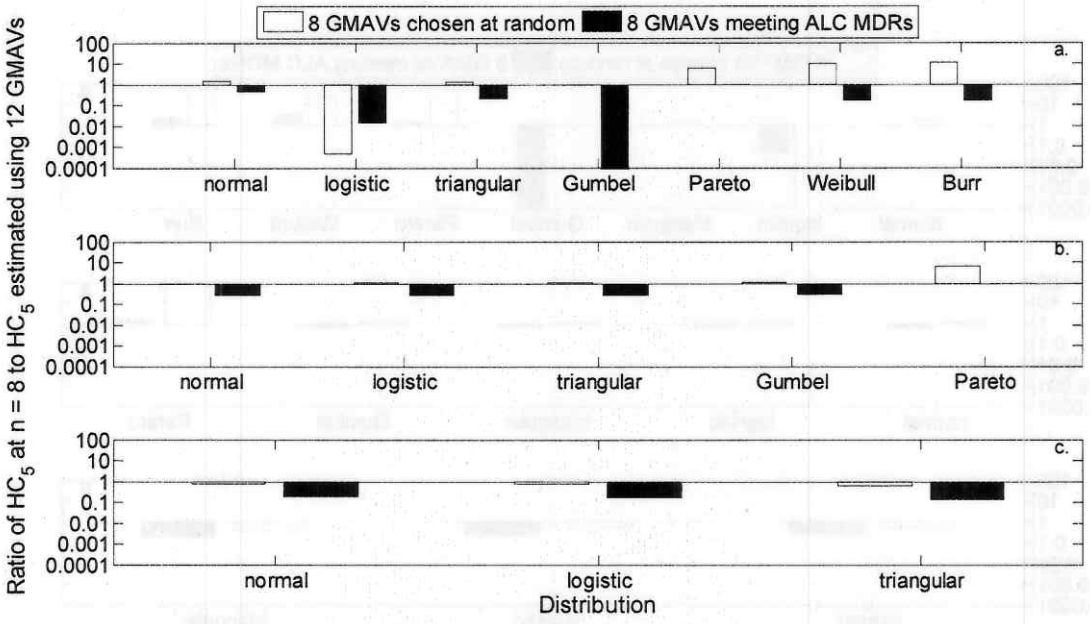


Figure S8.3.7. Average ratio of estimated HC_5s for **fenthion** from 1,000 draws of empirical data (n = 8 GMAVs) to the HC_5 estimated from all available saltwater GMAVs (n = 21).

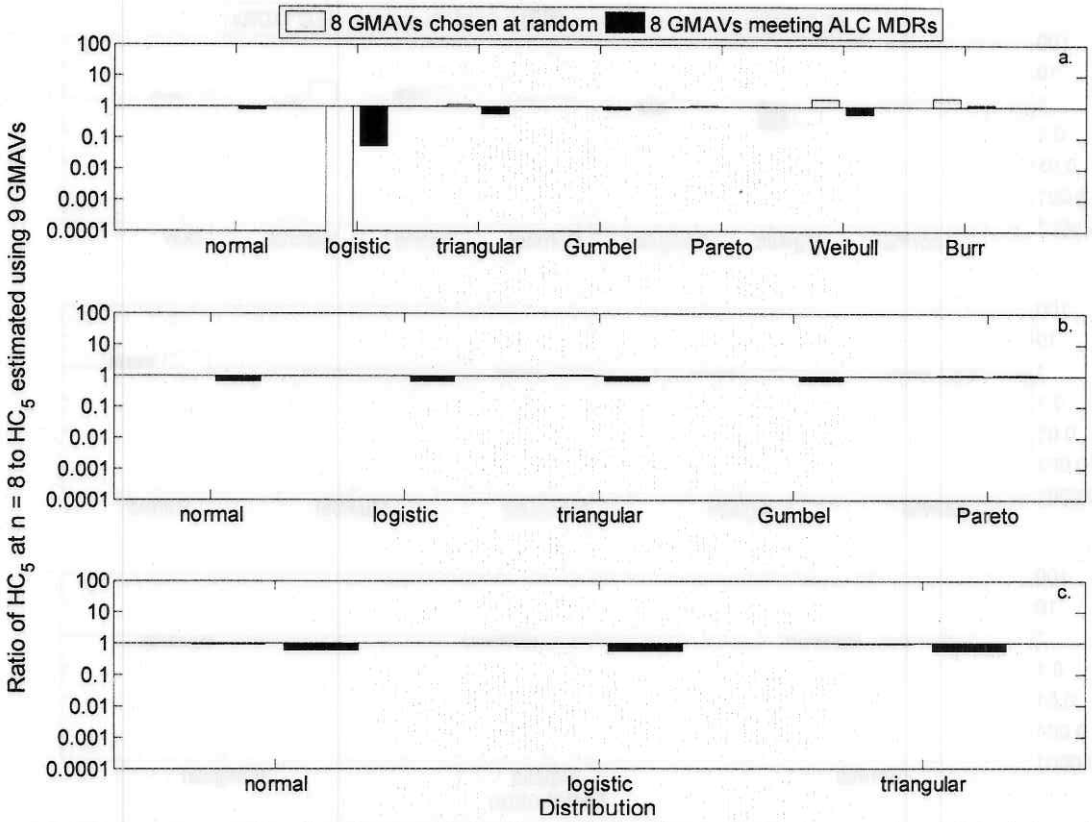


Figure S8.3.8. Average ratio of estimated HC_5s for **carbaryl** from 1,000 draws of empirical data (n = 8 GMAVs) to the HC_5 estimated from all available saltwater GMAVs (n = 9).

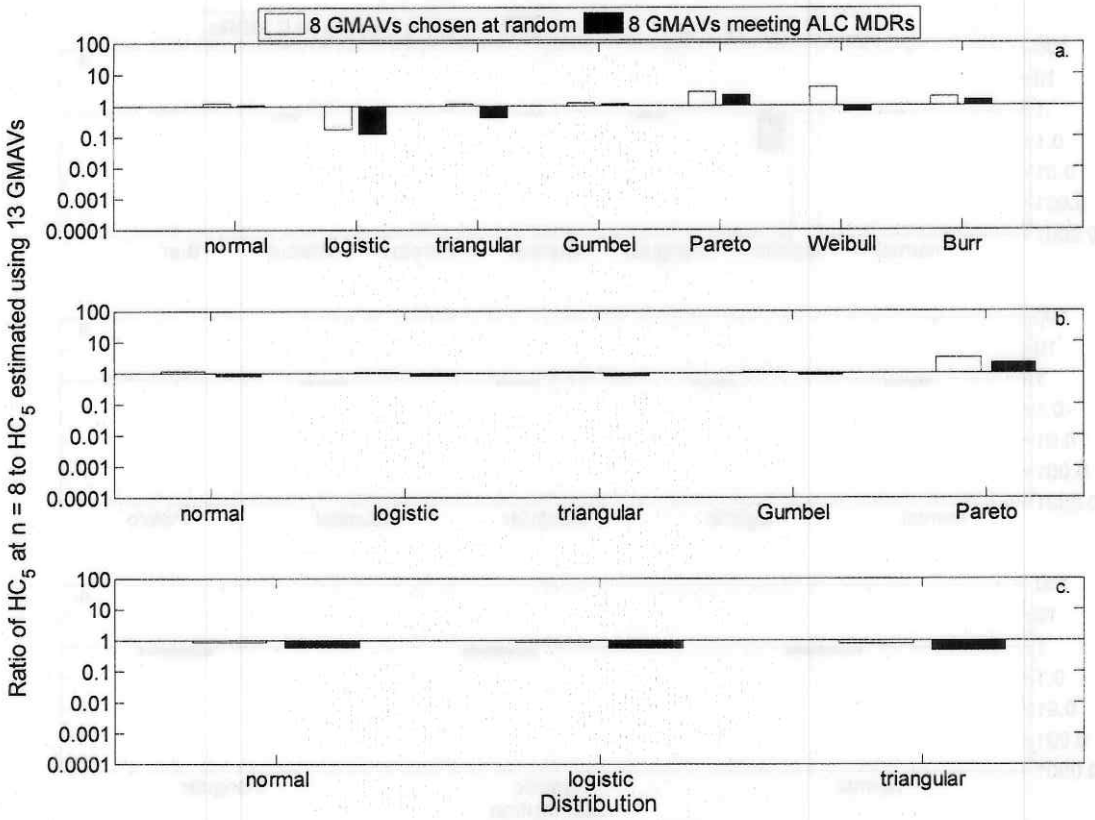


Figure S8.3.9. Average ratio of estimated HC_5s for **propoxur** from 1,000 draws of empirical data (n = 8 GMAVs) to the HC_5 estimated from all available freshwater GMAVs (n = 13).

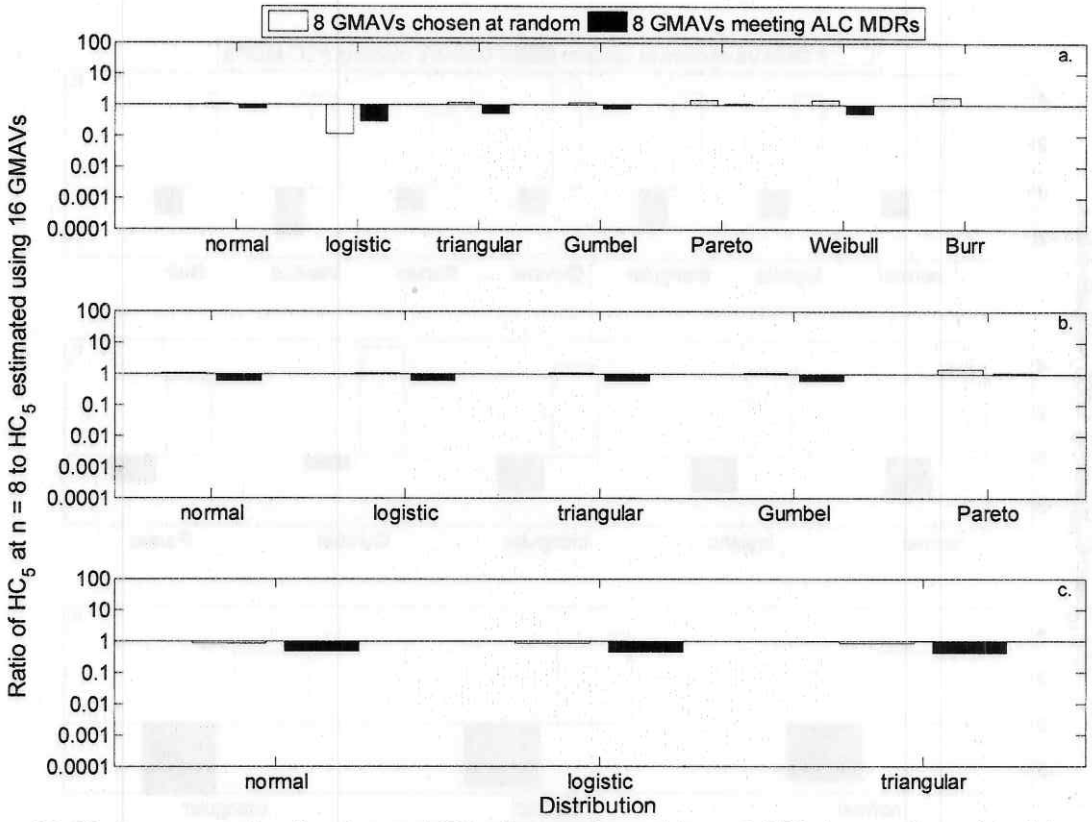


Figure S3.10. Average ratio of estimated HC_5s for **methomyl** from 1,000 draws of empirical data (n = 8 GMAVs) to the HC_5 estimated from all available freshwater GMAVs (n = 16).

8.4. Variance plots of analyses at n = 8 (ALC MDRs imposed versus not imposed)

Notes: Figure in this series for **malathion** freshwater tests is Figures 7 in the main text. Panels represent a. maximum likelihood estimates, b. moment estimates, c. graphical estimates. Note $\log_{10}$ scale of Y-axes.

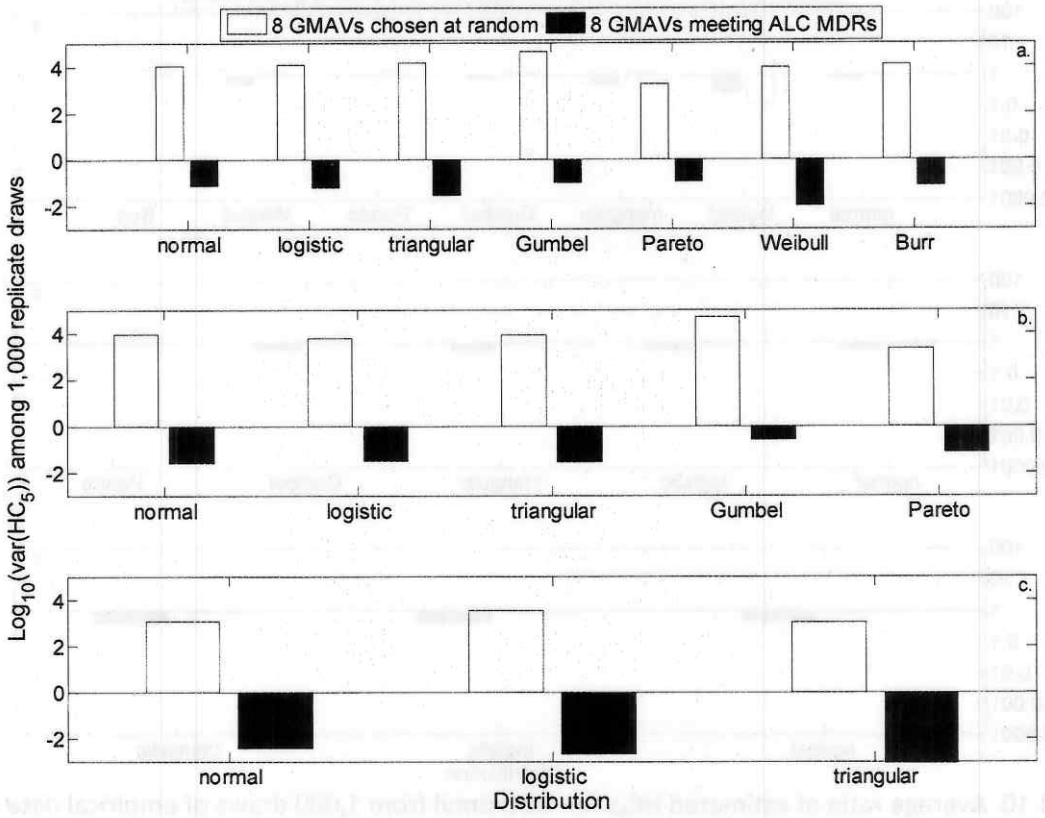


Figure S8.4.1. Comparison of variance among HC_5 estimates from 1,000 draws of 8 randomly chosen freshwater GMAVs for **diazinon** compared to when the 8 ALC MDRs are imposed.

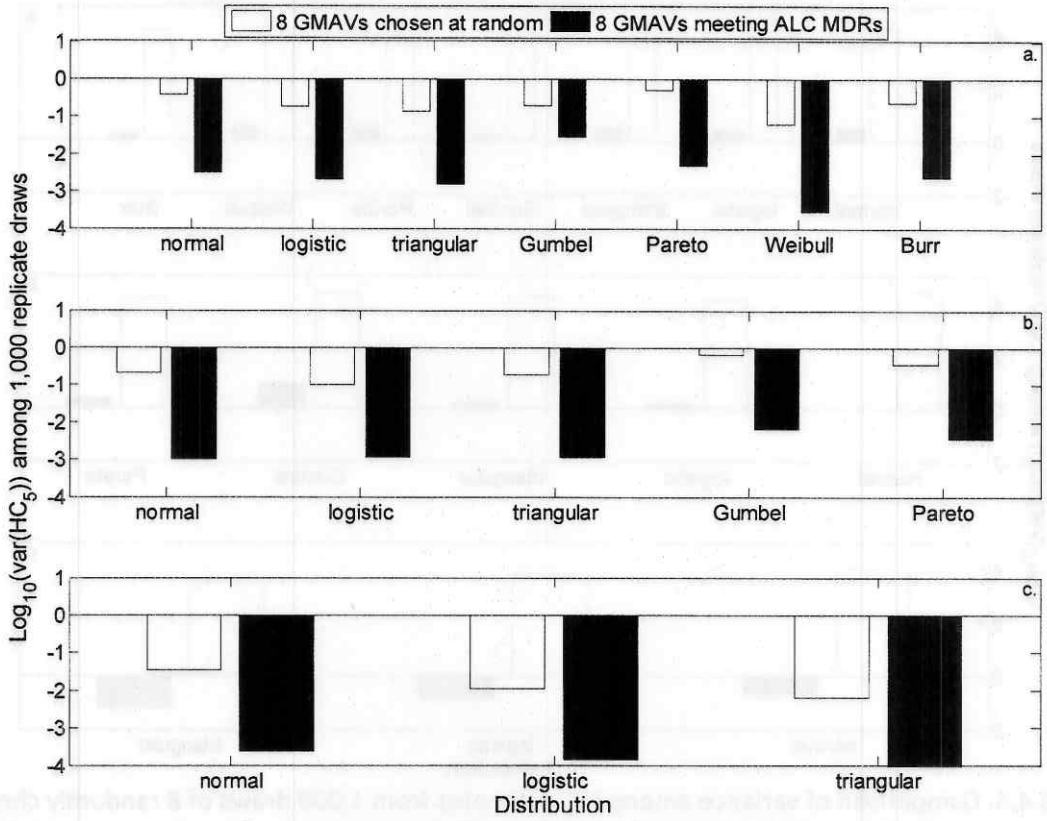


Figure S8.4.2. Comparison of variance among HC_5 estimates from 1,000 draws of 8 randomly chosen freshwater GMAVs for **chlorpyrifos** compared to when the 8 ALC MDRs are imposed.

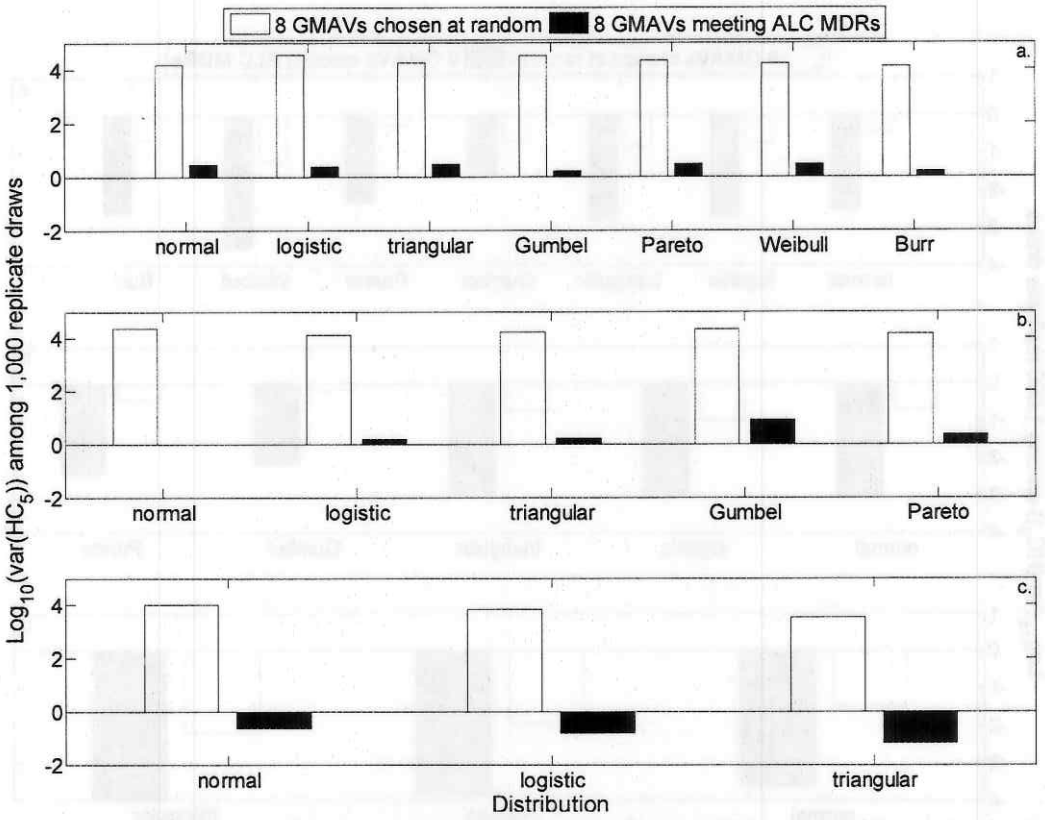


Figure S8.4.3. Comparison of variance among HC_5 estimates from 1,000 draws of 8 randomly chosen freshwater GMAVs for **fenitrothion** compared to when the 8 ALC MDRs are imposed.

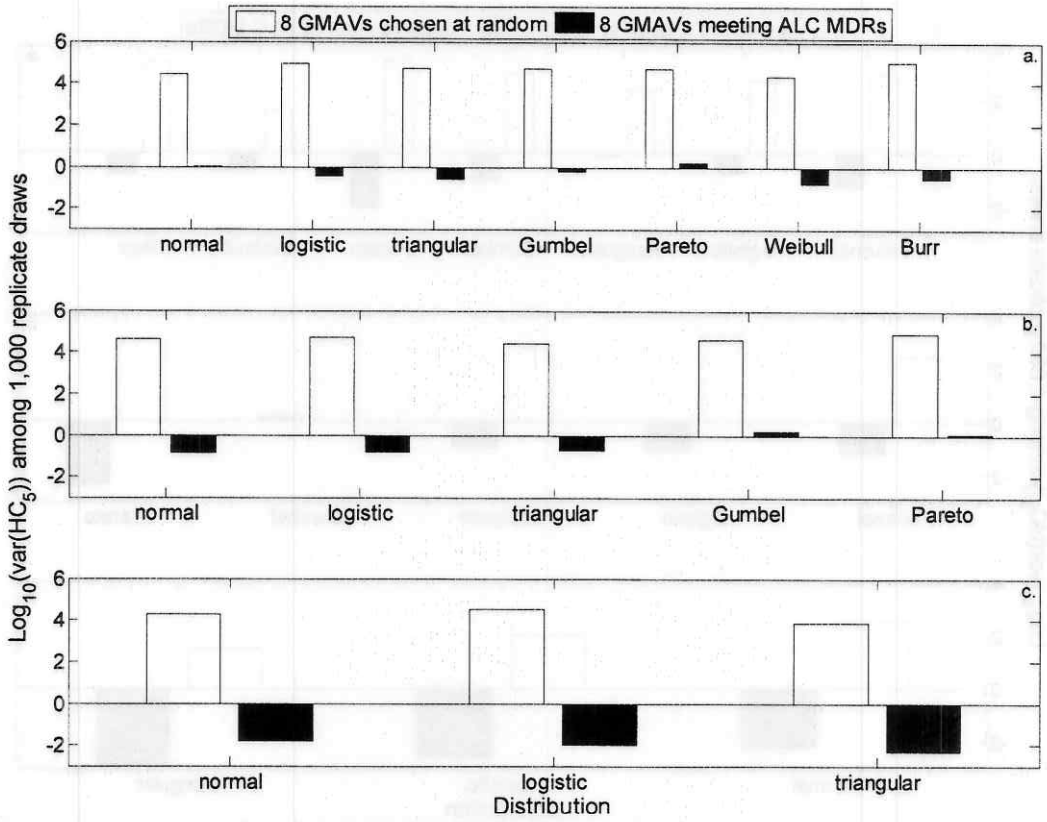


Figure S8.4.4. Comparison of variance among HC_5 estimates from 1,000 draws of 8 randomly chosen freshwater GMAVs for **methyl parathion** compared to when the 8 ALC MDRs are imposed.

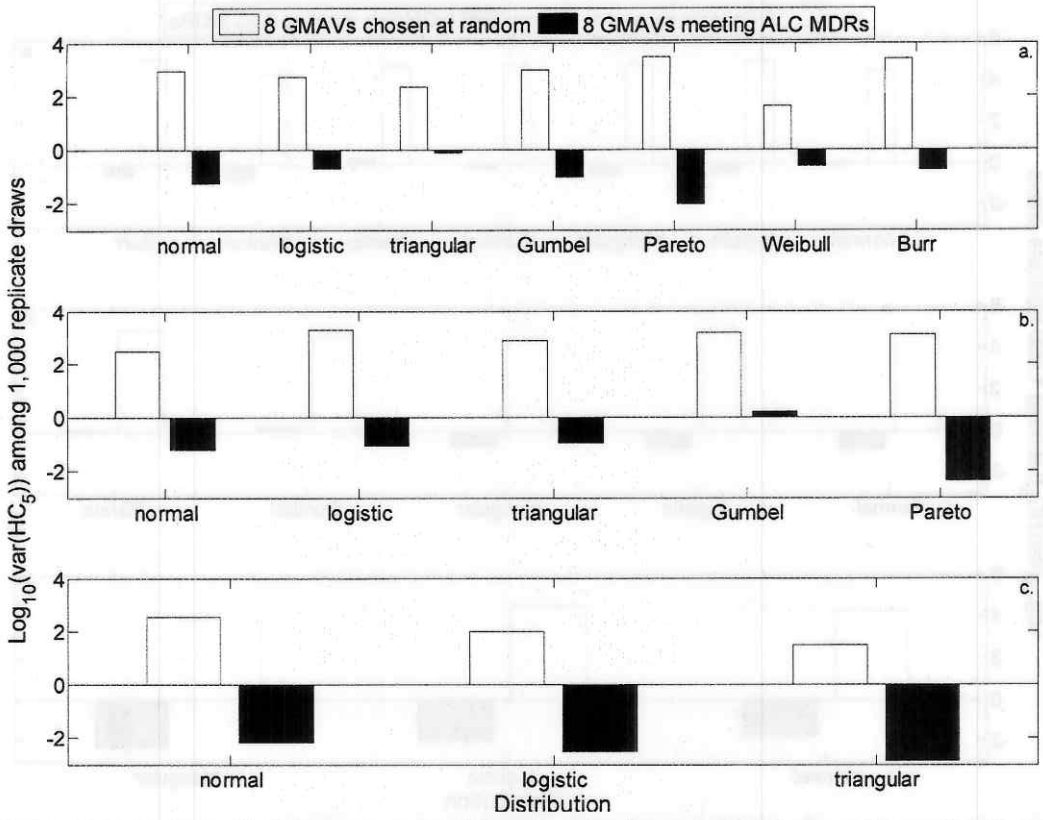


Figure S8.4.5. Comparison of variance among HC_5 estimates from 1,000 draws of 8 randomly chosen freshwater GMAVs for **dichlorvos** compared to when the 8 ALC MDRs are imposed.

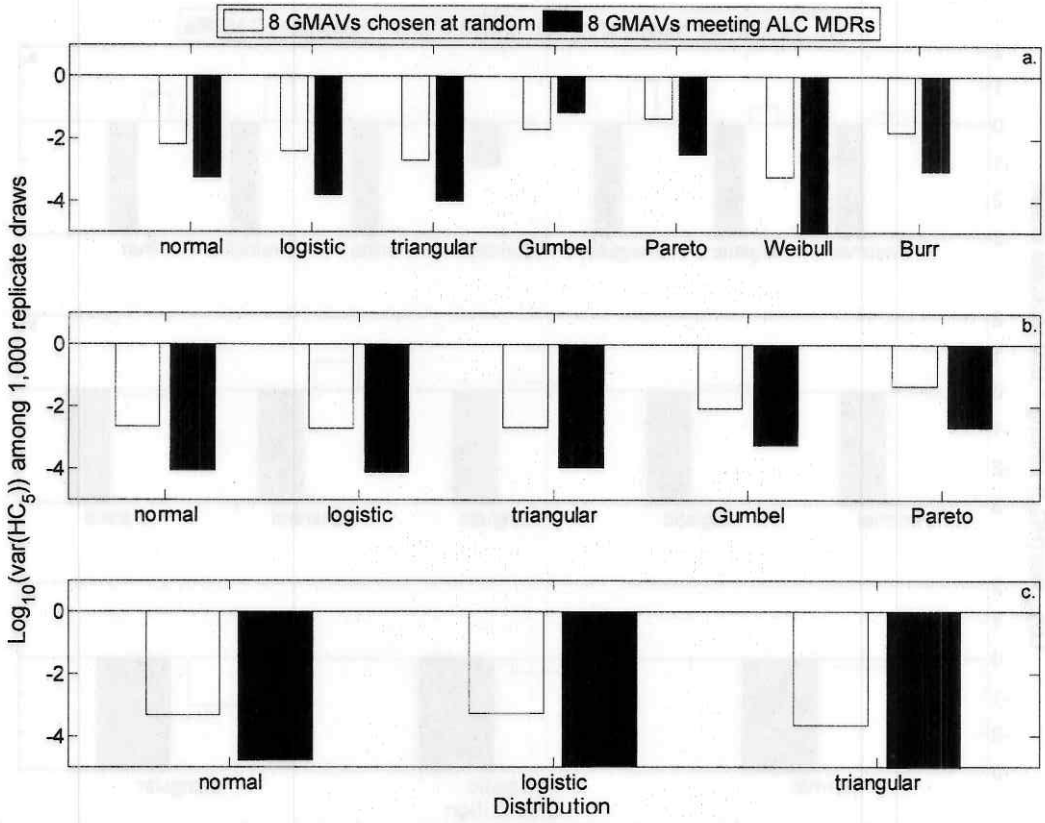


Figure S8.4.6. Comparison of variance among HC_5 estimates from 1,000 draws of 8 randomly chosen saltwater GMAVs for **chlorpyrifos** compared to when the 8 ALC MDRs are imposed.

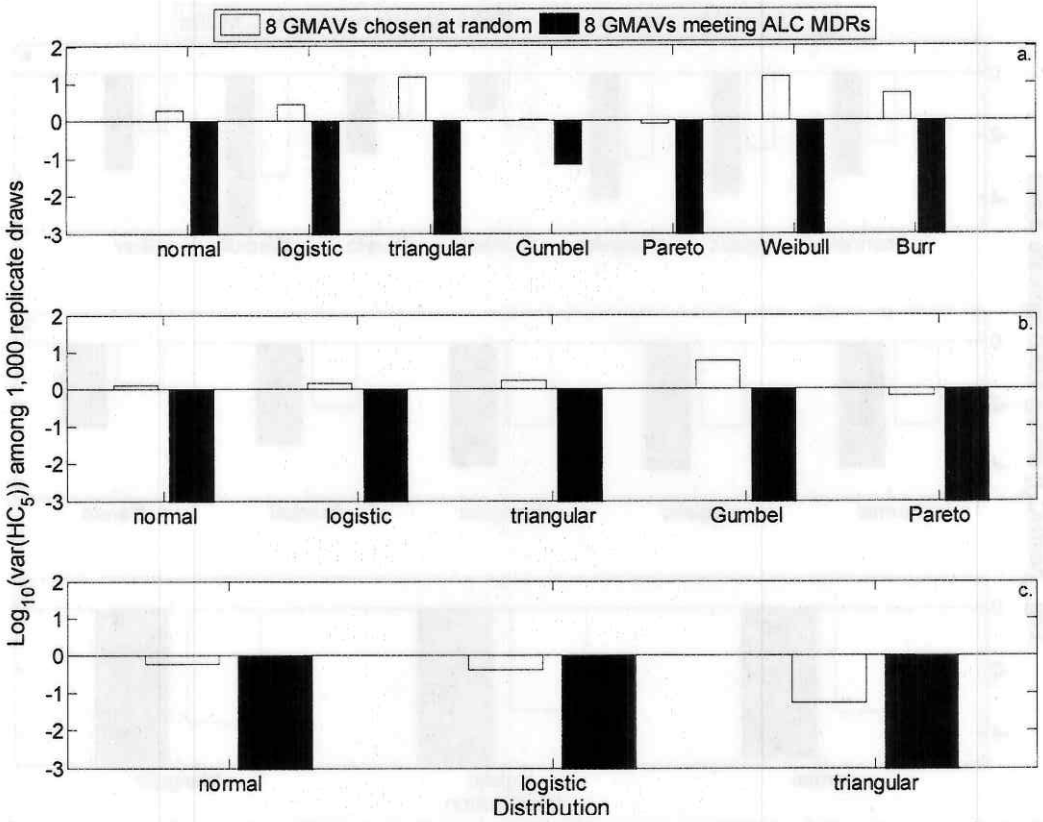


Figure S8.4.7. Comparison of variance among HC_5 estimates from 1,000 draws of 8 randomly chosen saltwater GMAVs for **fenthion** compared to when the 8 ALC MDRs are imposed.

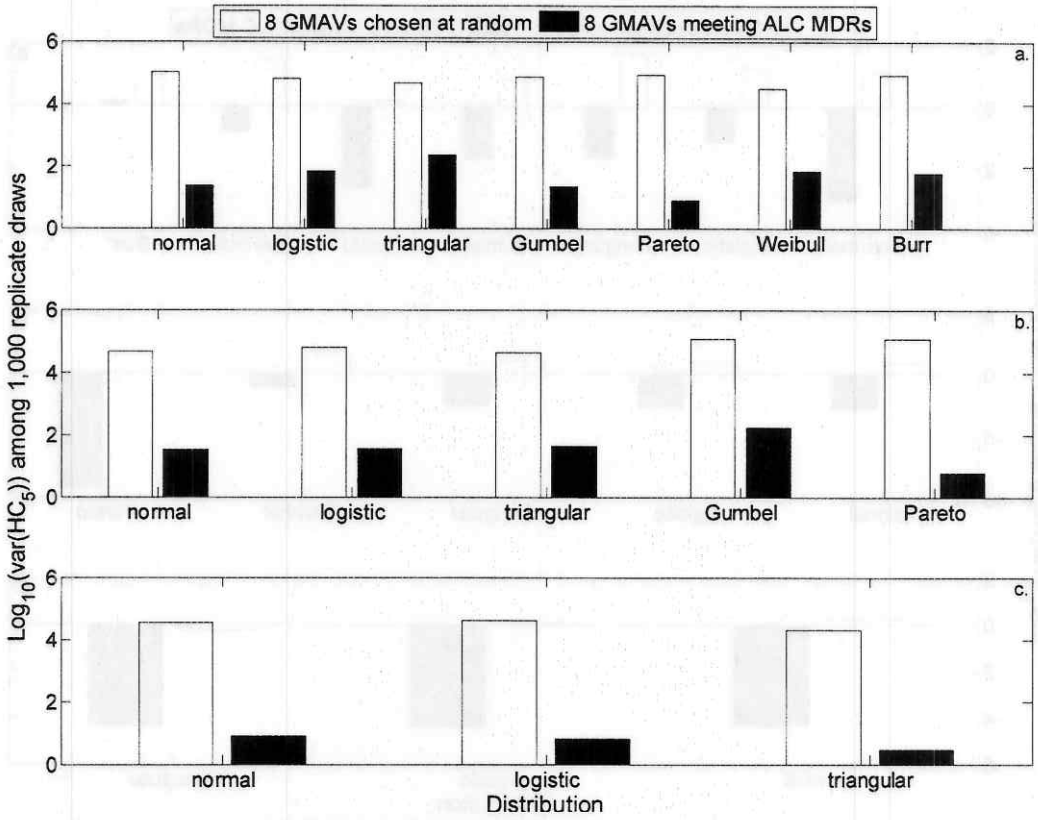


Figure S8.4.8. Comparison of variance among HC_5 estimates from 1,000 draws of 8 randomly chosen freshwater GMAVs for **carbaryl** compared to when the 8 ALC MDRs are imposed.

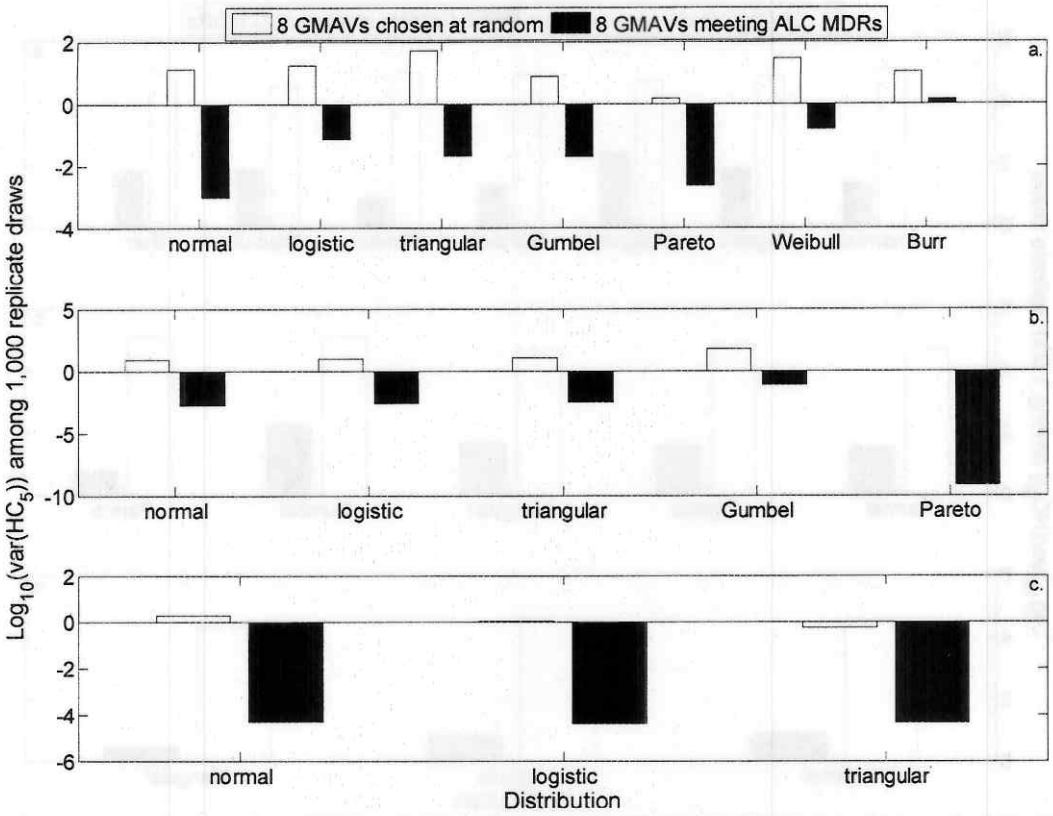


Figure S8.4.9. Comparison of variance among HC_5 estimates from 1,000 draws of 8 randomly chosen saltwater GMAVs for **carbaryl** compared to when the 8 ALC MDRs are imposed.

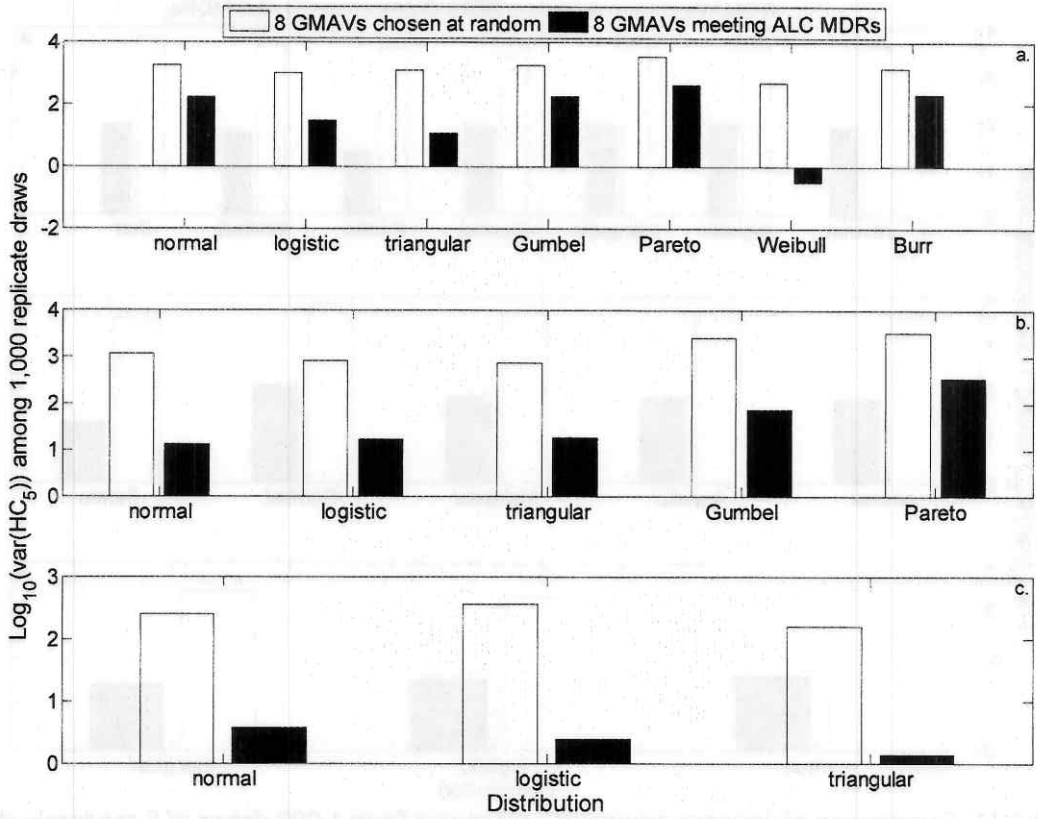


Figure S8.4.10. Comparison of variance among HC_5 estimates from 1,000 draws of 8 randomly chosen freshwater GMAVs for **propoxur** compared to when the 8 ALC MDRs are imposed.

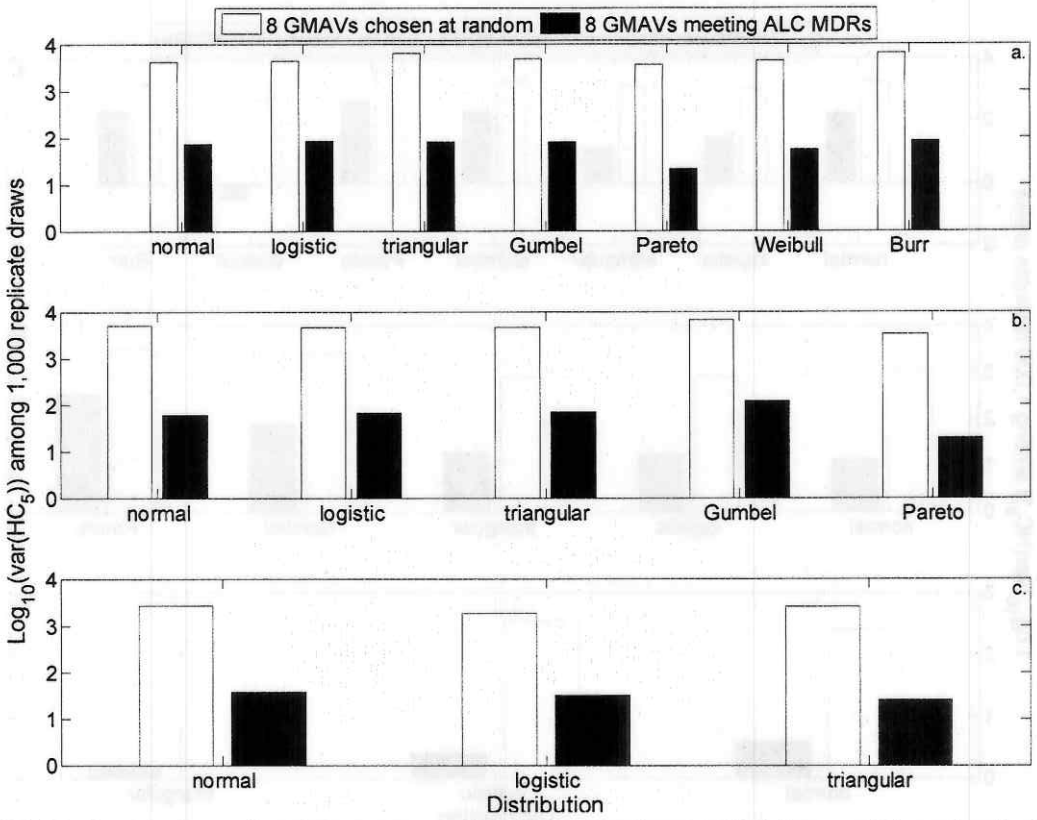


Figure S8.4.11. Comparison of variance among HC_5 estimates from 1,000 draws of 8 randomly chosen saltwater GMAVs for **methomyl** compared to when the 8 ALC MDRs are imposed.

8.5. Comparison plots of analyses at n = 3 (random subset versus 3 typical OPP species)

Notes: Figures in this series for **malathion** and **propoxur** freshwater tests are Figures 5 and 6 in the main text. Panels represent a. maximum likelihood estimates, b. moment estimates, c. graphical estimates. Note log_{10} scale of Y-axes.

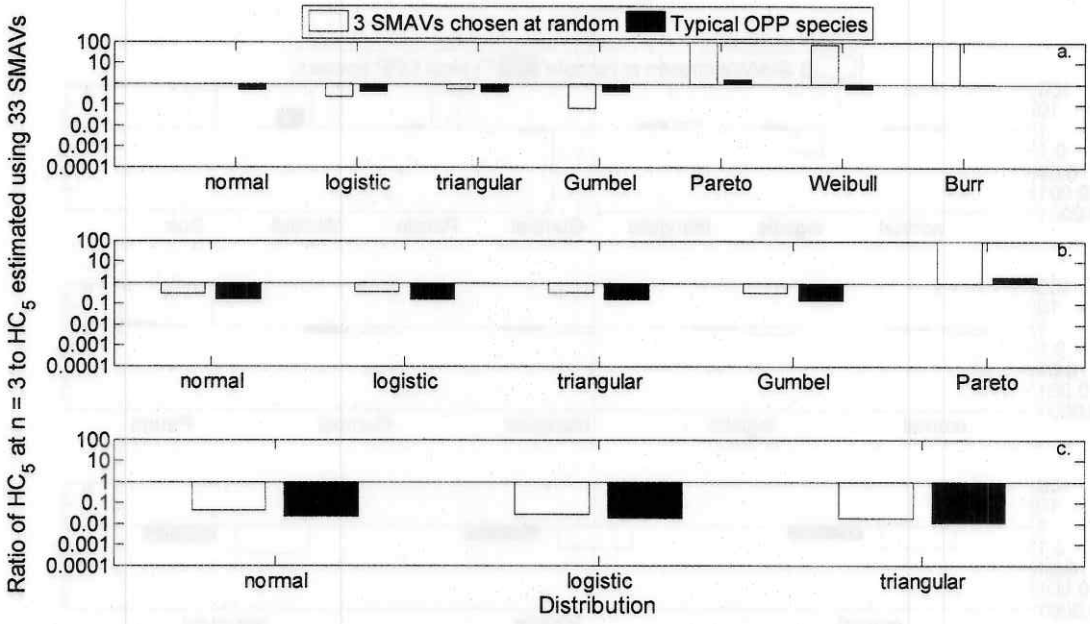


Figure S8.5.1. Average ratio of estimated HC_5s for **diazinon** from 1,000 draws of empirical data (n = 3 SMAVs) to the HC_5 estimated from all available freshwater SMAVs (n = 33).

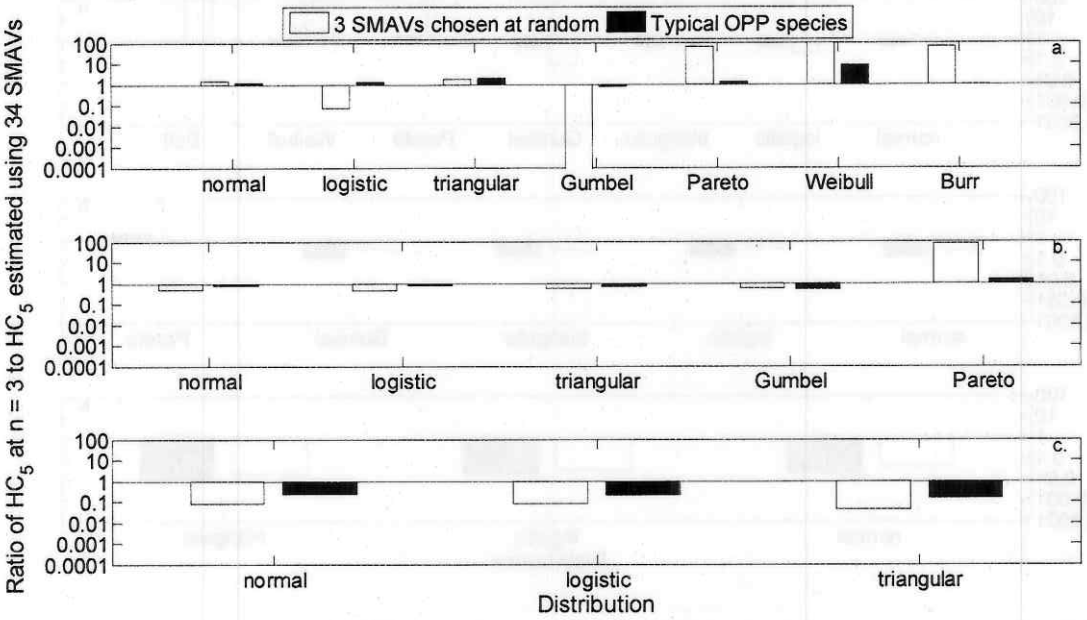


Figure S8.5.2. Average ratio of estimated HC_5s for **chlorpyrifos** from 1,000 draws of empirical data (n = 3 SMAVs) to the HC_5 estimated from all available freshwater SMAVs (n = 34).

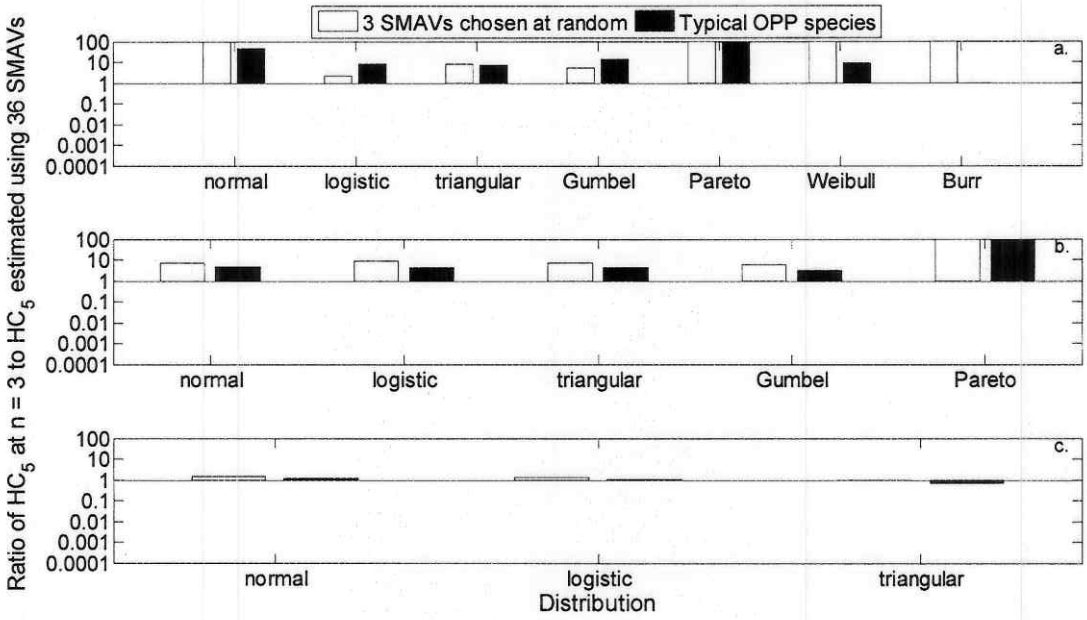


Figure S8.5.3. Average ratio of estimated HC_5s for **fenitrothion** from 1,000 draws of empirical data (n = 3 SMAVs) to the HC_5 estimated from all available freshwater SMAVs (n = 36).

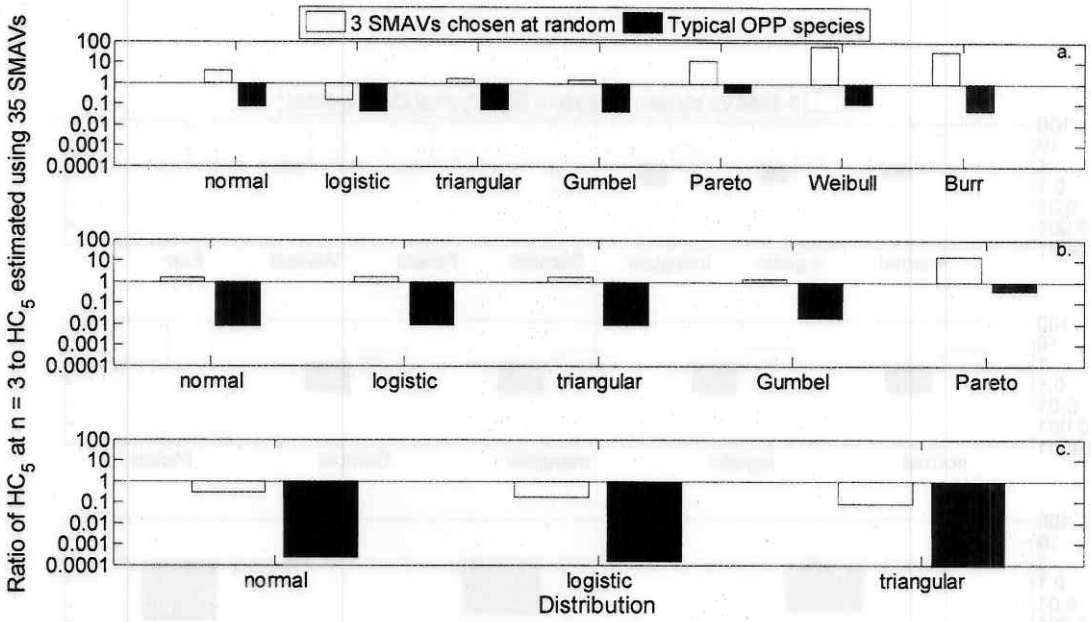


Figure S8.5.4. Average ratio of estimated HC_5s for **methyl parathion** from 1,000 draws of empirical data (n = 3 SMAVs) to the HC_5 estimated from all available freshwater SMAVs (n = 35).

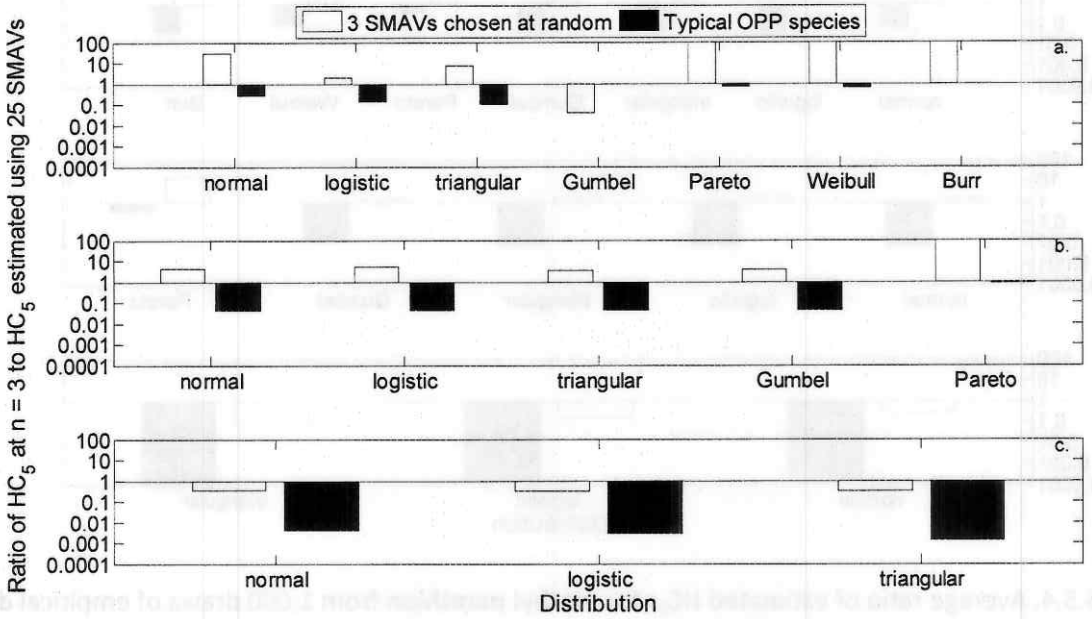


Figure S8.5.5. Average ratio of estimated HC_5s for **dichlorvos** from 1,000 draws of empirical data (n = 3 SMAVs) to the HC_5 estimated from all available freshwater SMAVs (n = 25).

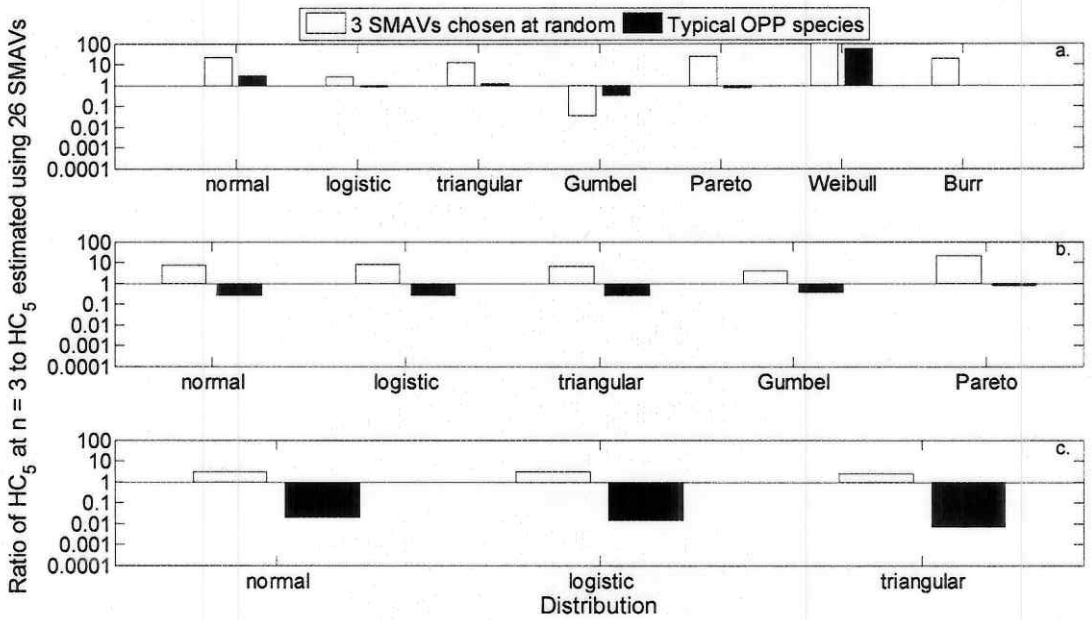


Figure S8.5.6. Average ratio of estimated HC_5s for **chlorpyrifos** from 1,000 draws of empirical data (n = 3 SMAVs) to the HC_5 estimated from all available saltwater SMAVs (n = 26).

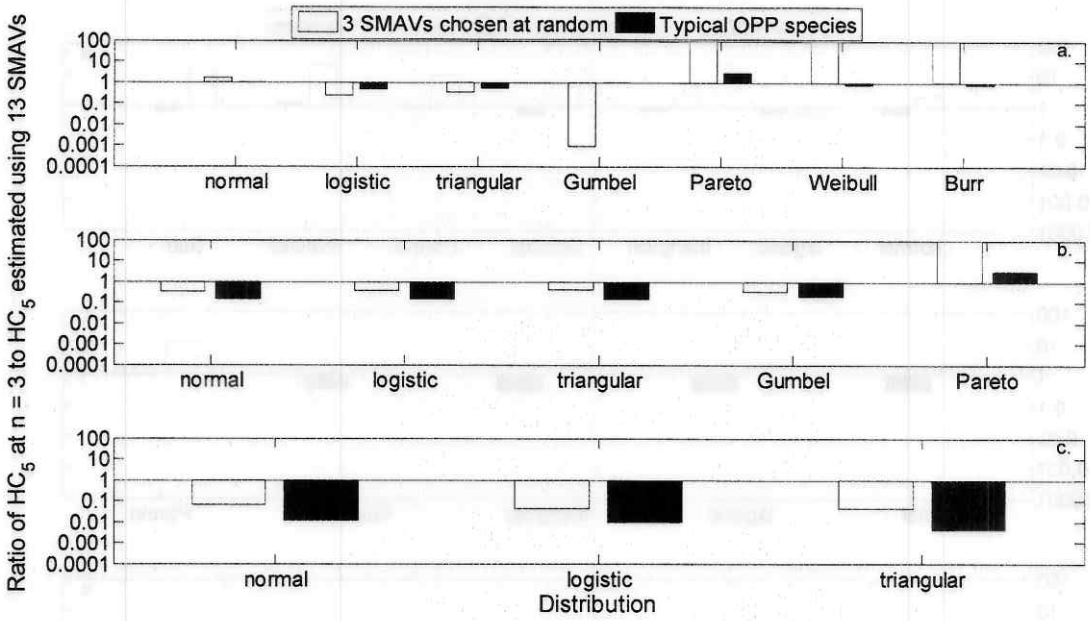


Figure S8.5.7. Average ratio of estimated HC_5s for **fenthion** from 1,000 draws of empirical data (n = 3 SMAVs) to the HC_5 estimated from all available saltwater SMAVs (n = 13).

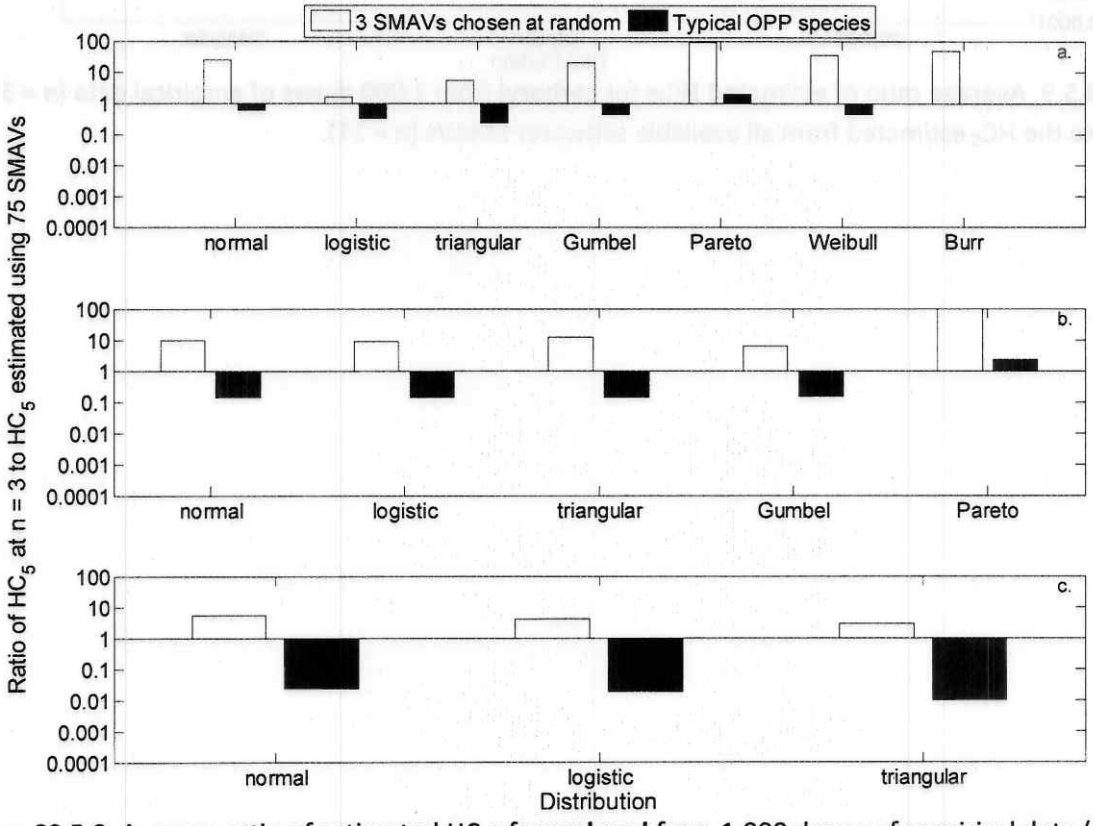


Figure S8.5.8. Average ratio of estimated HC_5s for **carbaryl** from 1,000 draws of empirical data (n = 3 SMAVs) to the HC_5 estimated from all available freshwater SMAVs (n = 75).

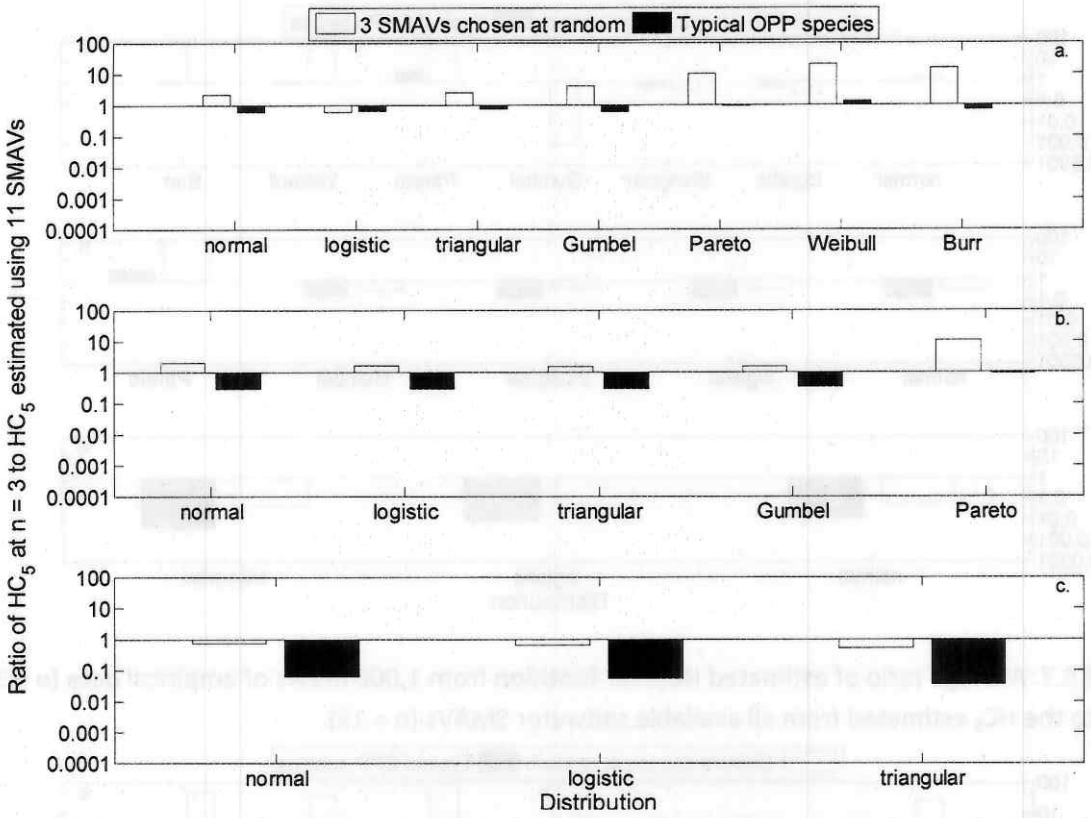


Figure S8.5.9. Average ratio of estimated HC_5s for **carbaryl** from 1,000 draws of empirical data (n = 3 SMAVs) to the HC_5 estimated from all available saltwater SMAVs (n = 11).

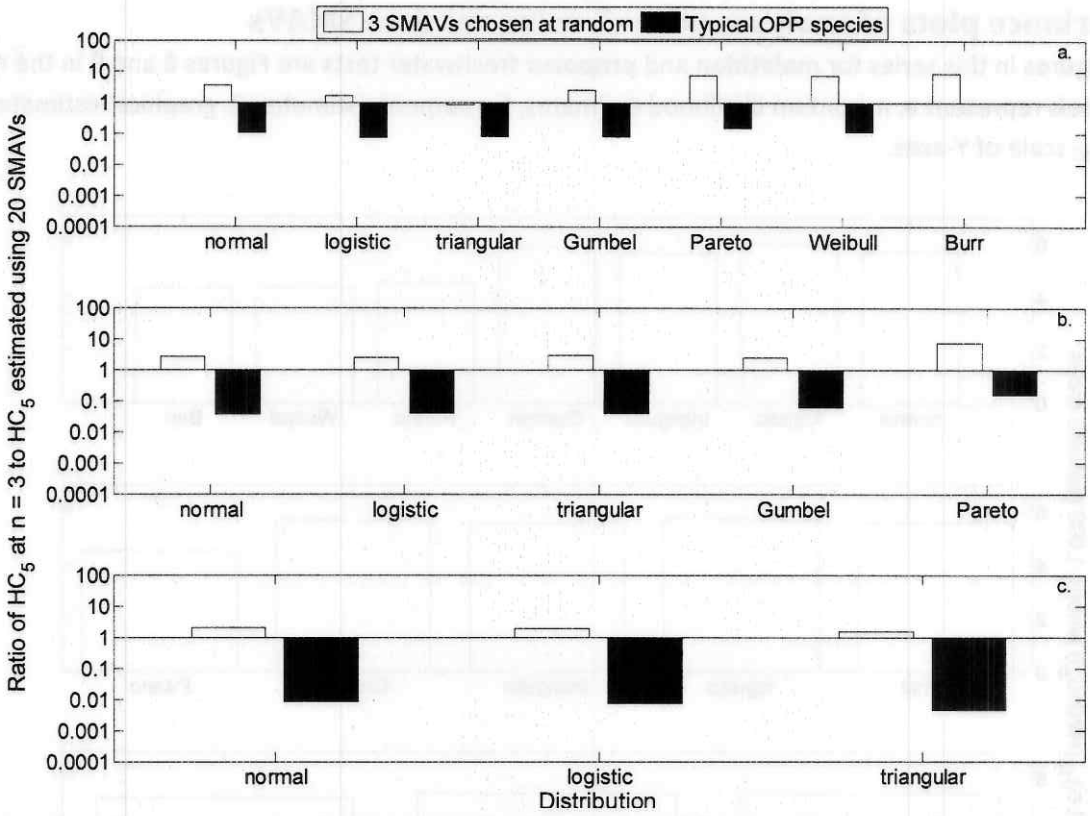


Figure S8.5.10. Average ratio of estimated HC_5s for **methomyl** from 1,000 draws of empirical data (n = 3 SMAVs) to the HC_5 estimated from all available freshwater SMAVs (n = 20).

8.6. Variance plots of analyses at n = 3 using random SMAVs

Notes: Figures in this series for **malathion** and **propoxur** freshwater tests are Figures 8 and 9 in the main text. Panels represent a. maximum likelihood estimates, b. moment estimates, c. graphical estimates. Note $\log_{10}$ scale of Y-axes.

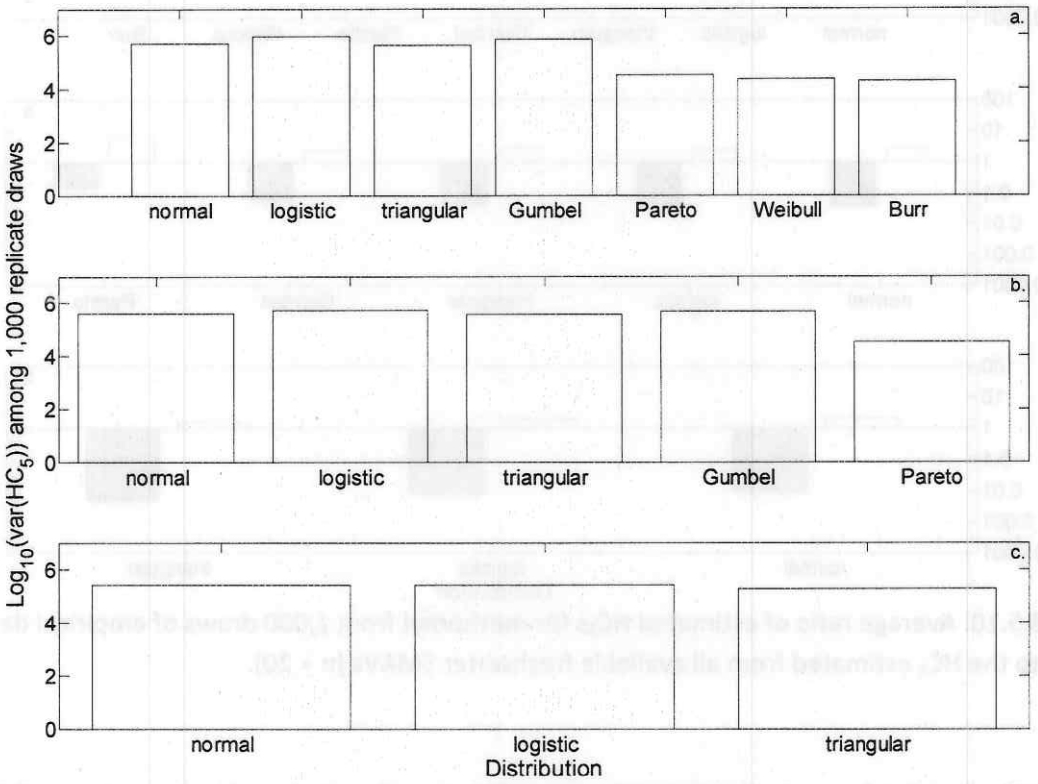


Figure S8.6.1. Variance in HC_5 estimates from 1,000 draws of 3 randomly chosen freshwater SMAVs for **diazinon**.

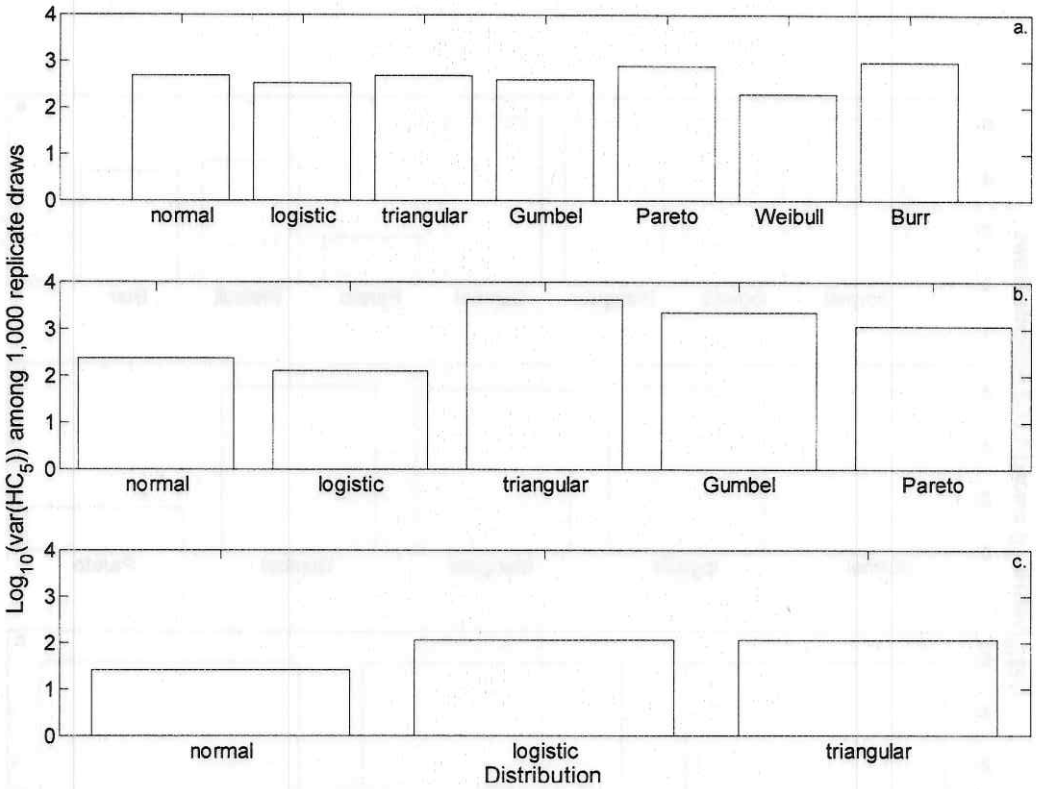


Figure S8.6.2. Variance in HC_5 estimates from 1,000 draws of 3 randomly chosen freshwater SMAVs for **chlorpyrifos**.

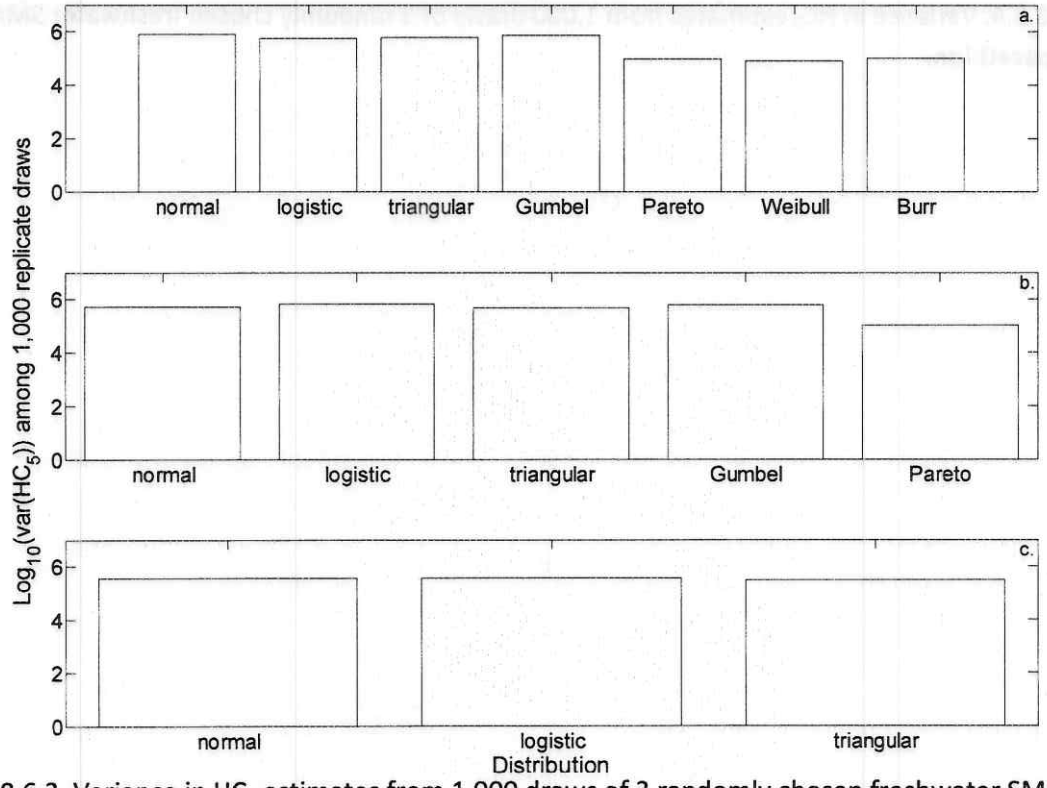


Figure S8.6.3. Variance in HC_5 estimates from 1,000 draws of 3 randomly chosen freshwater SMAVs for **fenitrothion**.

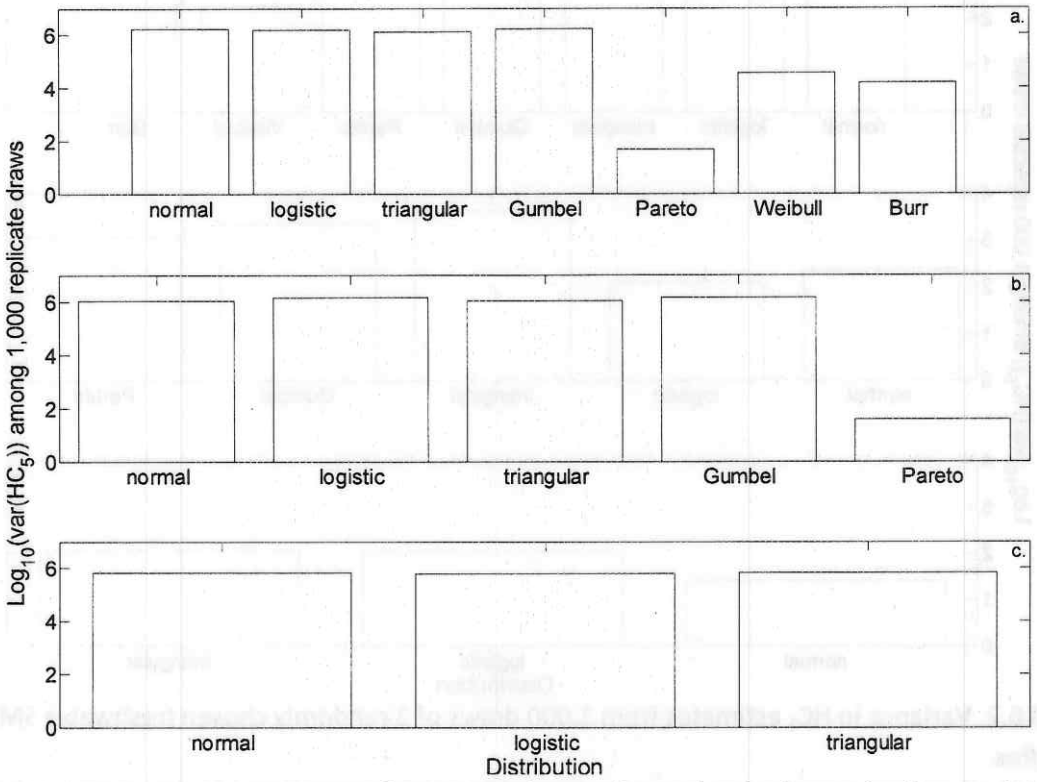


Figure S8.6.4. Variance in HC_5 estimates from 1,000 draws of 3 randomly chosen freshwater SMAVs for **methyl parathion**.

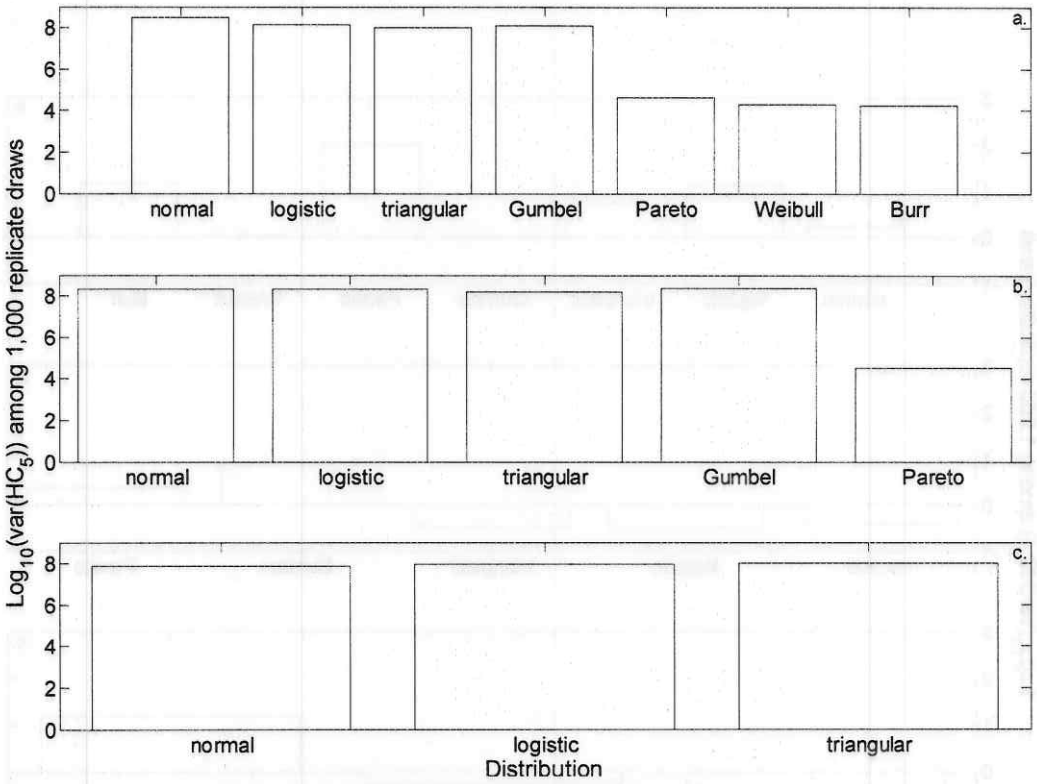


Figure S8.6.5. Variance in HC_5 estimates from 1,000 draws of 3 randomly chosen freshwater SMAVs for **dichlorvos**.

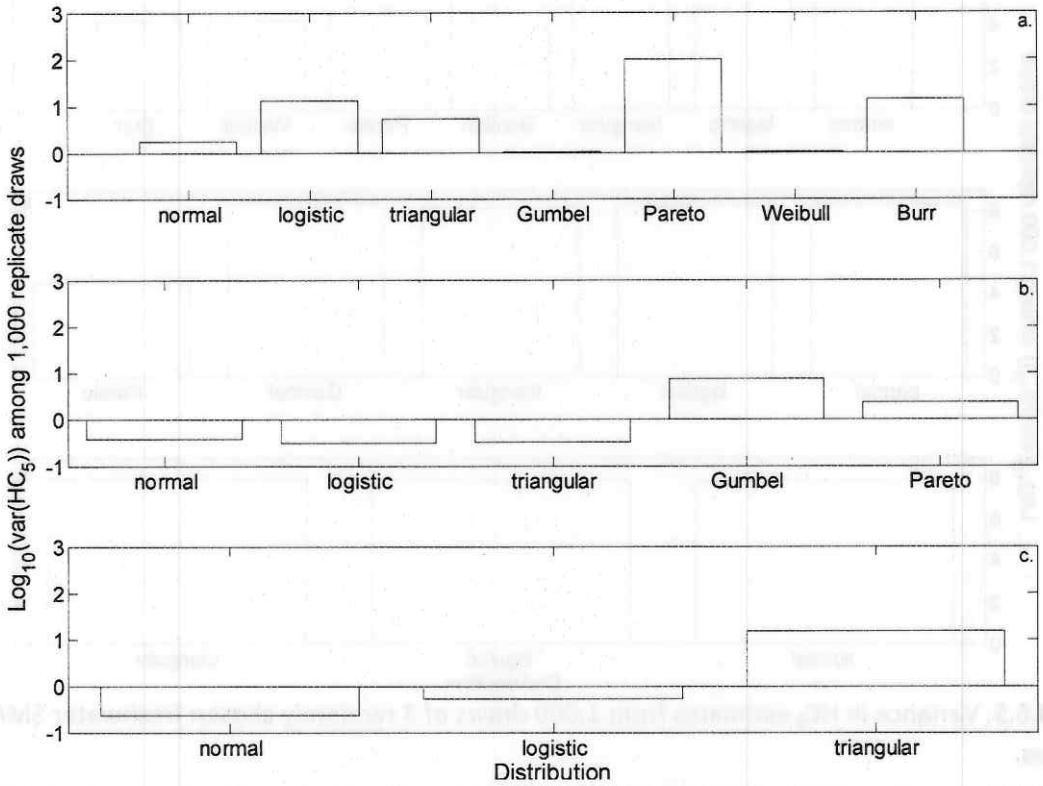


Figure S8.6.6. Variance in HC_5 estimates from 1,000 draws of 3 randomly chosen saltwater SMAVs for **chlorpyrifos**.

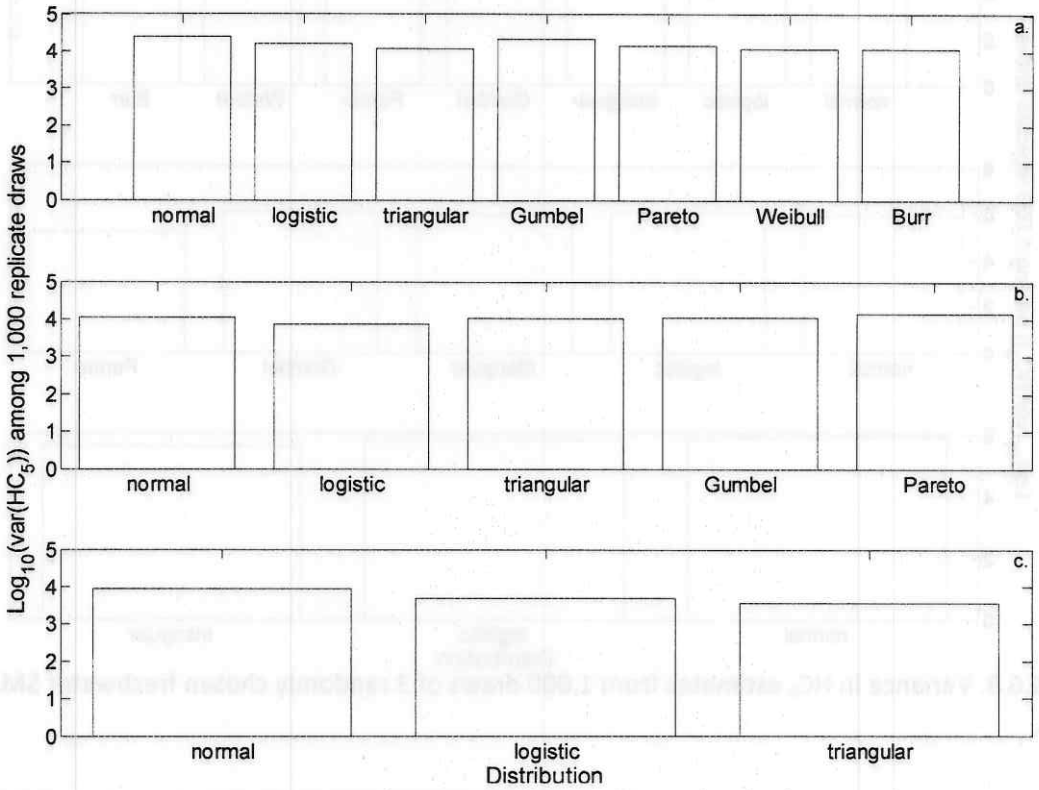


Figure S8.6.7. Variance in HC_5 estimates from 1,000 draws of 3 randomly chosen saltwater SMAVs for **fenthion**.

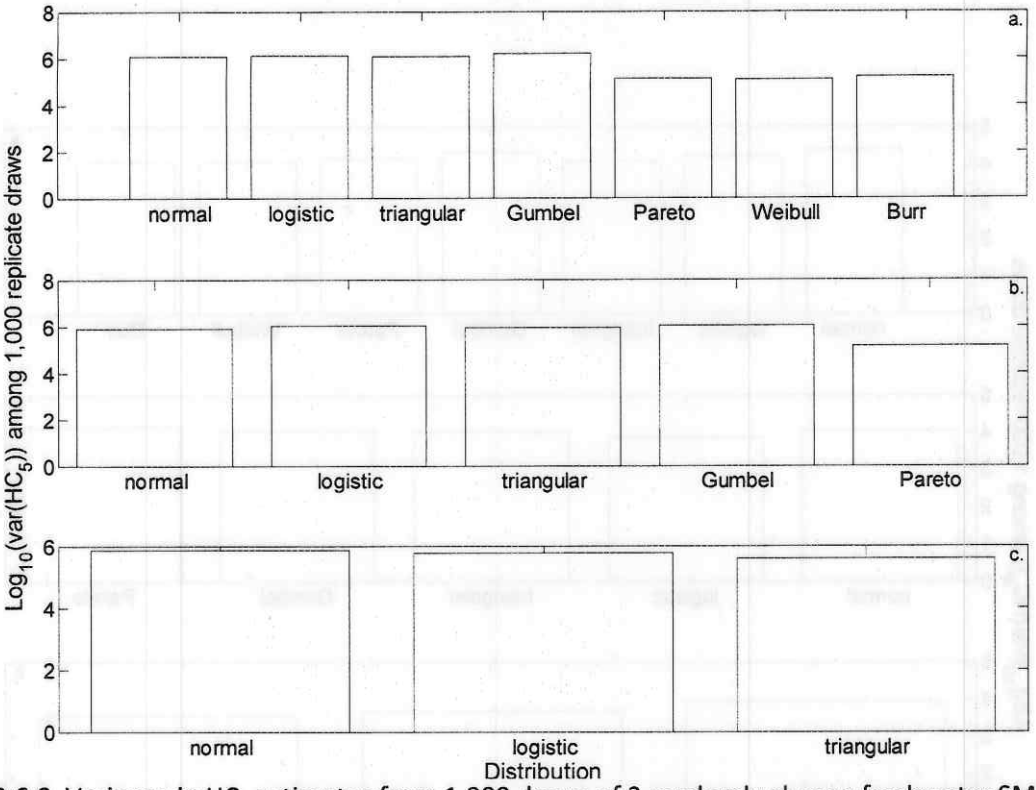


Figure S8.6.8. Variance in HC_5 estimates from 1,000 draws of 3 randomly chosen freshwater SMAVs for **carbaryl**.

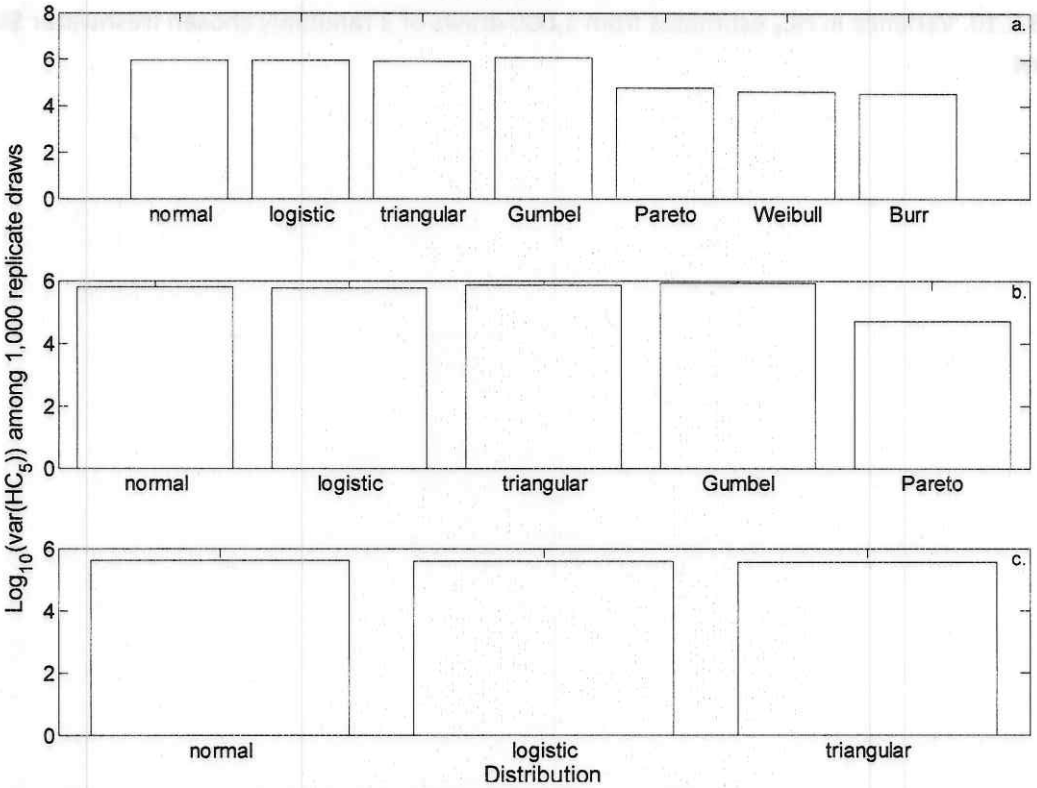


Figure S8.6.9. Variance in HC_5 estimates from 1,000 draws of 3 randomly chosen saltwater SMAVs for **carbaryl**.

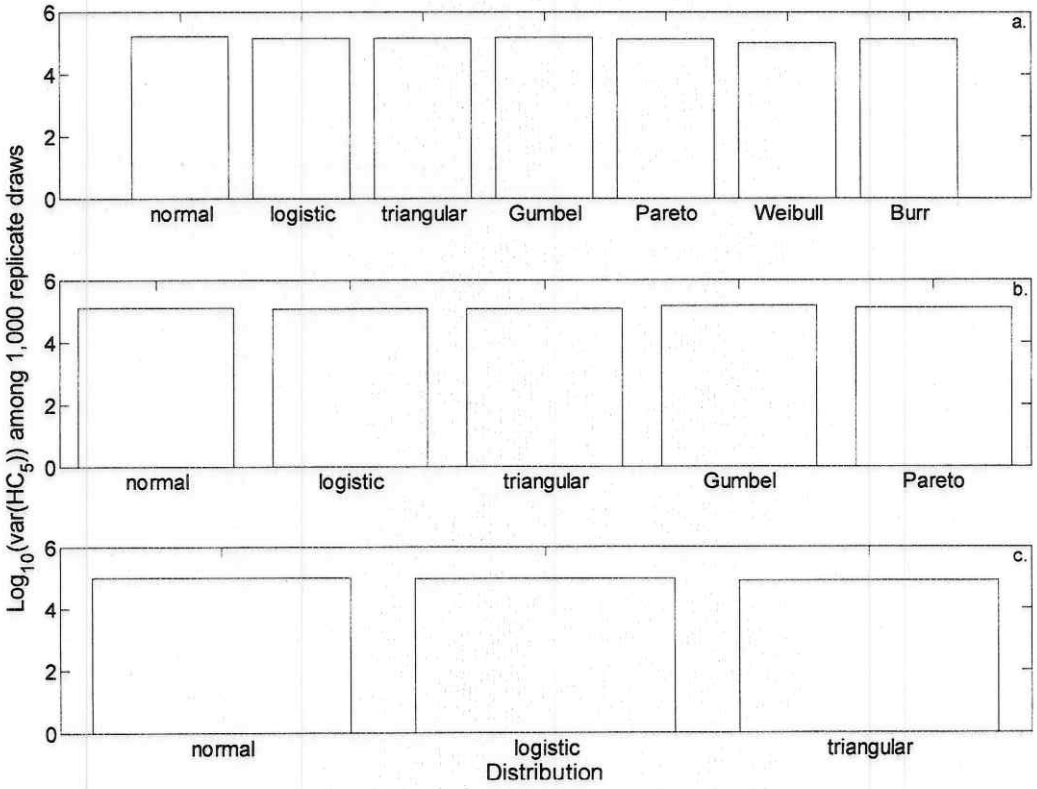


Figure S8.6.10. Variance in HC_5 estimates from 1,000 draws of 3 randomly chosen freshwater SMAVs for **methomyl**.

8.7. Goodness-of-fit results for $n = 20$ and $n = 50$.

Table S8.7.1. Power of parametric bootstrap goodness-of-fit tests for maximum likelihood at $n = 20$.

	normal	logistic	triangular	Gumbel	Pareto	Weibull	Burr
normal	**0.020**	0.067	0.080	0.087	0.733	0.167	0.033
logistic	0.080	**0.073**	0.193	0.173	0.773	0.133	0.093
triangular	0.080	0.107	**0.047**	0.127	0.727	0.120	0.080
Gumbel	0.167	0.113	0.427	**0.040**	0.433	0.373	0.047
Pareto	0.720	0.587	0.913	0.113	**0.040**	0.213	0.473
Weibull	0.273	0.187	0.467	0.247	0.940	**0.040**	0.353
Burr	0.067	0.067	0.247	0.093	0.733	0.240	**0.033**

Table S8.7.2. Power of parametric bootstrap goodness-of-fit tests for moment estimates at $n = 20$.

	normal	logistic	triangular	Gumbel	Pareto
normal	**0.040**	0.047	0.040	0.187	1.000
logistic	0.093	**0.067**	0.133	0.267	1.000
triangular	0.073	0.087	**0.053**	0.240	1.000
Gumbel	0.240	0.193	0.300	**0.047**	1.000
Pareto	0.747	0.693	0.813	0.233	**1.000**
Weibull	0.307	0.247	0.360	0.733	1.000
Burr	0.093	0.067	0.107	0.133	1.000

Table S8.7.3. Power of parametric bootstrap goodness-of-fit tests for graphical estimates at $n = 20$.

	normal	logistic	triangular
normal	**0.033**	0.007	0.053
logistic	0.127	**0.053**	0.180
triangular	0.053	0.027	**0.067**
Gumbel	0.260	0.160	0.313
Pareto	0.793	0.660	0.820
Weibull	0.327	0.207	0.347
Burr	0.113	0.027	0.180

Table S8.7.4. Power of parametric bootstrap goodness-of-fit tests for maximum likelihood at $n = 50$.

	normal	logistic	triangular	gumbel	Pareto	Weibull	Burr
normal	**0.080**	0.067	0.247	0.293	1.000	0.513	0.073
logistic	0.073	**0.053**	0.520	0.320	1.000	0.553	0.067
triangular	0.073	0.113	**0.087**	0.260	1.000	0.407	0.107
Gumbel	0.460	0.227	0.840	**0.040**	0.960	0.647	0.073
Pareto	1.000	0.973	1.000	0.107	**0.060**	0.253	0.860
Weibull	0.500	0.307	0.880	0.400	1.000	**0.047**	0.607
Burr	0.113	0.067	0.507	0.340	1.000	0.533	**0.080**

Table S8.7.5. Power of parametric bootstrap goodness-of-fit tests for moment estimates at n = 50.

	normal	logistic	triangular	Gumbel	Pareto
normal	**0.093**	0.087	0.120	0.460	1.000
logistic	0.100	**0.067**	0.200	0.447	1.000
triangular	0.067	0.140	**0.087**	0.473	1.000
Gumbel	0.500	0.427	0.580	**0.060**	1.000
Pareto	1.000	1.000	1.000	0.553	**1.000**
Weibull	0.547	0.440	0.633	0.973	1.000
Burr	0.133	0.073	0.227	0.387	1.000

Table S8.7.6. Power of parametric bootstrap goodness-of-fit tests for graphical estimates at n = 50.

	normal	logistic	triangular
normal	**0.100**	0.027	0.153
logistic	0.173	**0.060**	0.247
triangular	0.040	0.047	**0.067**
Gumbel	0.507	0.380	0.607
Pareto	0.993	0.993	1.000
Weibull	0.607	0.400	0.680
Burr	0.193	0.073	0.300

8.8. Model selection results for SSDs fit to GMAVs when N = 8 and ALC MDRs are met

Notes: The tables in this series pertaining to **malathion** is Tables 44 in the main text. Average is the average of model-averaged estimates using Akaike weights, calculated following methods described by Burnham and Anderson (2002).

Table S8.8.1. Mean estimated **diazinon** HC_5s when n = 8 freshwater GMAVs meeting the ALC MDRs

Fitting Method	normal	logistic	triangular	Gumbel	Pareto	Weibull	Burr	average
Maximum Likelihood	0.33	0.19	0.57	0.51	0.85	0.10	0.41	0.39
Moment Estimators	0.23	0.24	0.21	0.78	0.69			
Graphical Methods	0.05	0.03	0.07					

Table S8.8.2. Mean estimated **chlorpyrifos** HC_5s when n = 8 freshwater GMAVs meeting the ALC MDRs

Fitting Method	normal	logistic	triangular	Gumbel	Pareto	Weibull	Burr	average
Maximum Likelihood	0.05	0.04	0.09	0.10	0.14	0.01	0.07	0.07
Moment Estimators	0.04	0.04	0.04	0.11	0.12			
Graphical Methods	0.01	0.01	0.02					

Table S8.8.3. Mean estimated **fenitrothion** HC_5s when n = 8 freshwater GMAVs meeting the ALC MDRs

Fitting Method	normal	logistic	triangular	Gumbel	Pareto	Weibull	Burr	average
Maximum Likelihood	1.64	1.45	2.29	1.77	2.43	1.41	1.30	1.14
Moment Estimators	1.16	1.24	1.03	3.25	2.06			
Graphical Methods	0.34	0.23	0.45					

Table S8.8.4. Mean estimated **methyl parathion** HC_5s when n = 8 freshwater GMAVs meeting the ALC MDRs

Fitting Method	normal	logistic	triangular	Gumbel	Pareto	Weibull	Burr	average
Maximum Likelihood	0.76	0.53	1.53	1.22	2.25	0.39	0.91	0.89
Moment Estimators	0.53	0.59	0.50	1.92	1.89			
Graphical Methods	0.12	0.08	0.17					

Table S8.8.5. Mean estimated **dichlorvos** HC_5s when n = 8 freshwater GMAVs meeting the ALC MDRs

Fitting Method	normal	logistic	triangular	Gumbel	Pareto	Weibull	Burr	average
Maximum Likelihood	0.44	0.64	0.33	0.33	0.21	0.23	0.37	0.20
Moment Estimators	0.29	0.33	0.24	1.27	0.14			
Graphical Methods	0.05	0.03	0.07					

Table S8.8.6. Mean estimated **chlorpyrifos** HC_5s when n = 8 saltwater GMAVs meeting the ALC MDRs

Fitting Method	normal	logistic	triangular	Gumbel	Pareto	Weibull	Burr	average
Maximum Likelihood	0.013	0.009	0.020	0.056	0.090	0.001	0.046	0.040
Moment Estimators	0.008	0.010	0.008	0.030	0.076			
Graphical Methods	0.002	0.001	0.003					

Table S8.8.7. Mean estimated **fenthion** HC_5s when n = 8 saltwater GMAVs meeting the ALC MDRs

Fitting Method	normal	logistic	triangular	Gumbel	Pareto	Weibull	Burr	average
Maximum Likelihood	0.039	0.023	0.061	0.070	0.065	0.015	0.008	0.041
Moment Estimators	0.025	0.028	0.023	0.101	0.050			
Graphical Methods	0.005	0.003	0.007					

Table S8.8.8. Mean estimated **carbaryl** HC_5s when n = 8 freshwater GMAVs meeting the ALC MDRs

Fitting Method	normal	logistic	triangular	Gumbel	Pareto	Weibull	Burr	average
Maximum Likelihood	6.97	8.88	7.17	5.50	6.16	5.44	6.31	2.96
Moment Estimators	4.84	5.43	4.43	12.19	5.06			
Graphical Methods	1.70	1.10	1.93					

Table S8.8.9. Mean estimated **carbaryl** HC_5s when n = 8 saltwater GMAVs meeting the ALC MDRs

Fitting Method	normal	logistic	triangular	Gumbel	Pareto	Weibull	Burr	average
Maximum Likelihood	3.24	2.38	6.78	4.36	7.74	1.79	3.50	4.61
Moment Estimators	2.30	2.47	2.09	7.00	6.50			
Graphical Methods	0.64	0.41	0.83					

Table S8.8.10. Mean estimated **propoxur** HC_5s when n = 8 freshwater GMAVs meeting the ALC MDRs

Fitting Method	normal	logistic	triangular	Gumbel	Pareto	Weibull	Burr	average
Maximum Likelihood	18.0	13.1	31.6	34.5	38.6	2.0	33.3	27.7
Moment Estimators	13.5	14.4	12.4	34.1	33.9			
Graphical Methods	4.7	3.3	5.9					

Table S8.8.11. Mean estimated **methomyl** HC_5s when n = 8 freshwater GMAVs meeting the ALC MDRs

Fitting Method	normal	logistic	triangular	Gumbel	Pareto	Weibull	Burr	average
Maximum Likelihood	45.6	37.1	59.1	56.0	71.1	27.1	55.2	53.5
Moment Estimators	39.3	40.7	37.4	63.4	67.9			
Graphical Methods	22.8	18.7	25.6					

8.9. Model selection results for SSDs fit to GMAVs when N = 8 regardless of ALC MDRs

Notes: The tables in this series pertaining to **malathion** and **fenitrothion** are Tables 43 and 45 in the main text. Average is the average of model-averaged estimates using Akaike weights, calculated following methods described by Burnham and Anderson (2002).

Table S8.9.1. Mean estimated **diazinon** HC5s when n = 8 freshwater GMAVs regardless of MDRs

Fitting Method	normal	logistic	triangular	Gumbel	Pareto	Weibull	Burr	average
Maximum Likelihood	14.3	19.1	12.4	17.7	6.5	17.1	14.1	0.6
Moment Estimators	11.1	11.7	13.0	30.4	7.6			
Graphical Methods	4.5	4.0	5.0					

Table S8.9.2. Mean estimated **chlorpyrifos** HC5s when n = 8 freshwater GMAVs regardless of MDRs

Fitting Method	normal	logistic	triangular	Gumbel	Pareto	Weibull	Burr	average
Maximum Likelihood	0.16	0.12	0.21	0.21	0.29	0.06	0.19	0.08
Moment Estimators	0.12	0.11	0.12	0.30	0.24			
Graphical Methods	0.04	0.02	0.05					

Table S8.9.3. Mean estimated **methyl parathion** HC_5s when n = 8 freshwater GMAVs regardless of MDRs

Fitting Method	normal	logistic	triangular	Gumbel	Pareto	Weibull	Burr	average
Maximum Likelihood	43.6	49.2	17.8	24.1	21.1	33.5	50.4	0.9
Moment Estimators	29.0	17.5	28.1	28.5	32.0			
Graphical Methods	22.6	8.5	15.4					

Table S8.9.4. Mean estimated **dichlorvos** HC_5s when n = 8 freshwater GMAVs regardless of MDRs

Fitting Method	normal	logistic	triangular	Gumbel	Pareto	Weibull	Burr	average
Maximum Likelihood	7.7	4.9	9.3	9.8	15.1	1.9	12.5	0.6
Moment Estimators	6.6	5.6	4.9	12.1	9.5			
Graphical Methods	2.5	1.5	2.7					

Table S8.9.5. Mean estimated **chlorpyrifos** HC_5s when n = 8 saltwater GMAVs regardless of MDRs

Fitting Method	normal	logistic	triangular	Gumbel	Pareto	Weibull	Burr	average
Maximum Likelihood	0.044	0.031	0.059	0.093	0.130	0.008	0.091	0.055
Moment Estimators	0.031	0.031	0.029	0.077	0.125			
Graphical Methods	0.011	0.008	0.013					

Table S6.8.6. Mean estimated **fenthion** HC_5s when n = 8 saltwater GMAVs regardless of MDRs

Fitting Method	normal	logistic	triangular	Gumbel	Pareto	Weibull	Burr	average
Maximum Likelihood	0.49	0.89	0.63	0.38	0.39	0.89	0.57	0.05
Moment Estimators	0.37	0.40	0.31	0.93	0.34			
Graphical Methods	0.14	0.06	0.19					

Table S8.9.7. Mean estimated **carbaryl** HC_5s when n = 8 freshwater GMAVs regardless of MDRs

Fitting Method	normal	logistic	triangular	Gumbel	Pareto	Weibull	Burr	average
Maximum Likelihood	83.9	95.5	103.8	82.6	82.4	75.7	95.0	15.0
Moment Estimators	84.0	74.5	72.4	142.5	95.9			
Graphical Methods	53.4	38.9	57.3					

Table S8.9.8. Mean estimated **carbaryl** HC_5s when n = 8 saltwater GMAVs regardless of MDRs

Fitting Method	normal	logistic	triangular	Gumbel	Pareto	Weibull	Burr	average
Maximum Likelihood	6.00	7.50	9.16	6.02	8.21	5.66	5.46	4.86
Moment Estimators	4.37	4.68	3.91	12.05	6.84			
Graphical Methods	1.31	0.89	1.72					

Table S8.9.9. Mean estimated **propoxur** HC_5s when n = 8 freshwater GMAVs regardless of MDRs

Fitting Method	normal	logistic	triangular	Gumbel	Pareto	Weibull	Burr	average
Maximum Likelihood	31.4	28.3	43.2	44.0	50.7	11.5	40.4	22.6
Moment Estimators	25.7	27.0	26.1	54.3	47.1			
Graphical Methods	12.2	8.8	13.1					

Table S8.9.10. Mean estimated **methomyl** HC_5s when n = 8 freshwater GMAVs regardless of MDRs

Fitting Method	normal	logistic	triangular	Gumbel	Pareto	Weibull	Burr	average
Maximum Likelihood	91.3	98.4	98.1	96.4	92.6	77.3	93.0	68.3
Moment Estimators	85.7	87.2	82.5	125.6	86.5			
Graphical Methods	51.7	49.0	57.0					

8.9. Preliminary analysis of the use of MOA-specific parameters for the logistic distribution

Kristina Garber
US Environmental Protection Agency
Office of Pesticide Programs
Washington, DC

de Zwart (2002) proposed that knowledge of mode of action (MOA) could be used to improve the ability to fit SSDs at smaller sample sizes. He hypothesized that within an MOA, the average test result among species would vary, but that the variance among responses would remain relatively constant. Thus, given a set of pre-existing data concerning chemicals with a given MOA, the expected variation in response to a new chemical with the same MOA could be specified a priori, and test results for the new chemical could be used only to estimate the mean response. To illustrate this idea, de Zwart used the logistic distribution and hypothesized that the location parameter (α) would vary among chemicals, but that the scale parameter (β) would remain relatively constant.

As a preliminary assessment of de Zwart's method, I estimated β for 23 chemicals (Table ?)and estimated mean values of the estimated β by mode of action. These were compared to the mean β values reported by de Zwart (2002). In total, it was possible to compare β values from 5 modes of action, including nonpolar narcosis, polar narcosis, AChE inhibition by carbamate insecticides, AChE inhibition by organophosphate insecticides and neurotoxicity caused by organochlorine insecticides. Of these modes of action, the average β calculated for neurotoxicity for this analysis is closest to the average β reported by de Zwart (2002), although the sample size used by de Zwart for this mode of action was very small (*i.e.*, N=2). The average β calculated in this analysis for the nonpolar and polar narcosis modes of action falls within the standard deviation of the average β reported by de Zwart. For both AChE inhibition modes of action, the average β calculated in this analysis falls outside of the standard deviation of the average β reported by de Zwart. This indicates that the average β reported by de Zwart may be applicable to narcosis modes of action and neurotoxicity of organochlorine insecticides, but may be limited for AChE inhibitors.

Table 1. FAVs derived according to 85 guideline methodology and HC_5 derived using de Zwart (2002) methodology.

Chemical	MOA	FAV	C5	# acute GMAVs	β
Carbaryl	AChE inhibition (carbamate)	4.4	96	55	0.50
Methomyl	AChE inhibition (carbamate)	50	61	16	0.50
Propoxur	AChE inhibition (carbamate)	9.6	116	13	0.50
Chlorpyrifos	AChE inhibition (organophosphate)	0.074	0.47	30	0.71
Diazinon	AChE inhibition (organophosphate)	0.54	13	28	0.71
Dichlorvos	AChE inhibition (organophosphate)	0.083	15	22	0.71
Fenitrothion	AChE inhibition (organophosphate)	0.11	11	30	0.71
Malathion	AChE inhibition (organophosphate)	2.1	29	57	0.71
Methyl parathion	AChE inhibition (organophosphate)	7.3	17	29	0.71
DDT	neurotoxicity	0.85	1.9	49	0.50
Dieldrin	neurotoxicity	0.54	2.9	18	0.50
Endosulfan	neurotoxicity	0.21	0.67	32	0.50
Endrin	neurotoxicity	0.21	0.26	37	0.50
Lindane	neurotoxicity	1.9	6.3	23	0.50
Benzenamine*	Narcosis (polar)	45	24000	39	0.31
Nonylphenol*	Narcosis (polar)	41	50	21	0.31
Ammonia	Narcosis (nonpolar)	4800	7100	20	0.39
Benzene*	Narcosis (nonpolar)	6300	12000	9	0.39
Chloroform*	Narcosis (nonpolar)	26000	24000	8	0.39
Phenol	Narcosis (nonpolar)	8600	8300	42	0.39
Toluene*	Narcosis (nonpolar)	10000	9900	11	0.39
Pentachlorophenol	Uncoupling oxidative phosphorylation	20	24	40	0.38
1,3 Dichloropropene	Reactivity	630	???	10	???

*Not a pesticide

Table 2. Comparison of mean β values reported by de Zwart 2002 and mean β values of example chemicals.

Mode of Action	De Zwart 2002			Current analysis		
	Average β	Standard deviation of β *	N**	Average β	Standard deviation of β	N**
Nonpolar narcosis	0.39	0.17	34	0.22	0.07	5
Polar narcosis	0.31	0.11	13	0.36	0.26	2
AChE inhibition (carbamates)	0.71	0.099	11	0.52	0.19	3
AChE inhibition (organophosphates)	0.50	0.26	27	0.87	0.10	6
Neurotoxicity (organochlorines)	0.50	0.18	2	0.51	0.10	5

*Calculated as: $stdev\ of\ \beta = SEM * \sqrt{N}$

** N = number of chemicals used to determine average β

In comparing the results of this analysis to those of de Zwart, (2002) it is important to note three major differences in the two methods. First, this analysis had more strict definitions of acute toxicity for aquatic animals (i.e., 48-h duration for crustaceans, 96-h for fish and other invertebrates). Second, this analysis developed SSDs based on acute toxicity data for genera, while de Zwart (2002) developed SSDs for species. Third, this analysis developed SSDs using only aquatic animals, while de Zwart (2002) included effects data to derive SSDs for aquatic animals, plants and bacteria. This limitation arises because the modes of action of the different chemicals may be different for animals, plants and

bacteria. The latter difference in methods represents a significant limitation of the application of mean β values to estimation of C5 values intended to approximate FAVs, which are based on effects data for aquatic animals only. This may explain why the AChE inhibitors, which primarily act upon animals, but not plants, have average β values for only animals which differ from the average β reported by de Zwart, which include plants.

Because Pennington's (2003) work depends upon that of de Zwart (2002), it should be noted that the uncertainties and limitations of the β values recommended by de Zwart (2002)would also extend to the extrapolation factors reported by Pennington (2003). This is especially relevant to the use of the extrapolation factors for AChE inhibitors, which are based on β values that may be affected by inclusion of non-animal toxicity data in de Zwart's (2002) analysis.

HC_5 values were generated for each example chemical using de Zwart's (2002)approach (logistic distribution with MOA specific β value) along with the full data set available for that chemical. For 20 of the 23 chemicals considered, the de Zwart (2002)method generated HC_5 values that overpredict the FAVs, with factors ranging 1.2x to 530x. Three of the 23 chemicals considered had HC_5 values that underpredicted the FAV, but were still on the same order of magnitude as the FAV. In these cases, all three chemicals (chloroform, phenol and toluene) are described by the nonpolar narcosis MOA.