ESTIMATION OF DOSE REQUIREMENTS FOR EXTREME LEVELS OF EFFICACY

Mark West

Guy Hallman

Estimation of dose requirements for extreme levels of efficacy

Mark West and Guy Hallman

The objective of this paper is to explore the extent of how dose-response models may be used to estimate extreme levels of efficacy for controlling insect pests and possibly other uses. Probit-9 mortality (99.9968% mortality) is a standard for treatment effectiveness in tephritid fruit fly research, and has been adopted by the United States Department of Agriculture for fruit flies and other pests. Data taken from the phytosanitary treatment (PT) literature are analyzed. These data are used to fit dose-response models with logit, probit and complimentary log-log links. The effectiveness of these models for predicting extreme levels of efficacy is compared using large (~100,000+ individuals) confirmatory trials that are also reported in the PT literature. We examine the role of model goodness-of-fit as a requirement for obtaining reliable dose requirements.

Mark West, Ph.D., Northern Plains Area Statistician, USDA Agricultural Research Service, 2150 Centre Ave., Fort Collins, Co 80526-8119. USA. Email: Mark.West@ars.usda.gov

Guy Hallman, Ph.D., Northern Plains Area Research Entomologist, USDA Agricultural Research Service, 1515 College Ave., Manhattan, Ks 66502. USA. Email: Guy.Hallman@ars.usda.gov

Introduction

Phytosanitary regulation is important to prevent the introduction of agricultural pests such as tephritidfruit flies. Government agencies, trade organizations and international consortiums establish guidelines for controlling the spread of such pests to limit their economic impact. For example, the USDA Animal and Plant Health Inspection Service (APHIS) often requires a standard of probit-9 efficacy for the treatment of Tephritidae fruit flies, the most important group of plant pests for which treatments are devised. Probit-x efficacy of a treatment is such that the treatment will inflict $P \cdot 100\%$ mortality on the insect population and where the value x is based on the relationship$\Phi(x-5)=P$. Φdenotes the standard normal cumulative distribution function. Therefore a treatment with probit-9 efficacy will kill $\Phi(9-5)\cdot 100\% = 99.9968\%$ of the pest it is formulated to control. Probit-9 efficacy dates back to Baker (1939) whose rationale of using this standard was to "assure no survival of [fruit fly] in the product treated." Products requiring regulation of fruit fly include mango, papaya, avocado, apple, zucchini and carambola. Treatments devised to control fruit fly have included fumigation, cold treatment, heat treatment and radiation. Because the treatments are quantitative and the efficacy standard is to kill a certain percentage of pests, establishing the level of treatment amounts to finding a dose requirement. Therefore, much of the literature on phytosanitary treatments of fruit fly involves dose-response studies that typically include small to moderate scale (size of experiment including amount of fruit and number of insects) experimental trials set up to find the required dose. These small scale studies are then followed by a separate and usually very large scale confirmatory experiment to validate the dose determined from the experimental trials. The methodology used to find the dose is typically accomplished by a probit analysis. Establishing the actual level of treatment needed for such an extreme efficacy requirement as probit-9 poses is very challenging. If the prediction is too low, the confirmatory test will fail and will have to be restarted at a higher level. If the prediction is too high the proposed treatment will result in a waste of treatment resources and time, and may reduce the quality of the commodity due to overtreatment. In this paper we report data from the literature where the dose requirements determined from the experimental trials almost but not quite achieved probit-9 mortality when tested in the confirmatory experiment. We did this in order to address the following questions:

1. Is model goodness-of-fit critically important for estimating the required dose?
2. How do other models compare to the probit for estimating the required dose?
3. Can we identify experimental conditions or methodology that would cause the predicted dose to fail?

The paper will be organized into following sections: 1) Methods 2) Literature Reviewed 3) Simulations and 4) Summary.

Methods

Small Scale Experiments

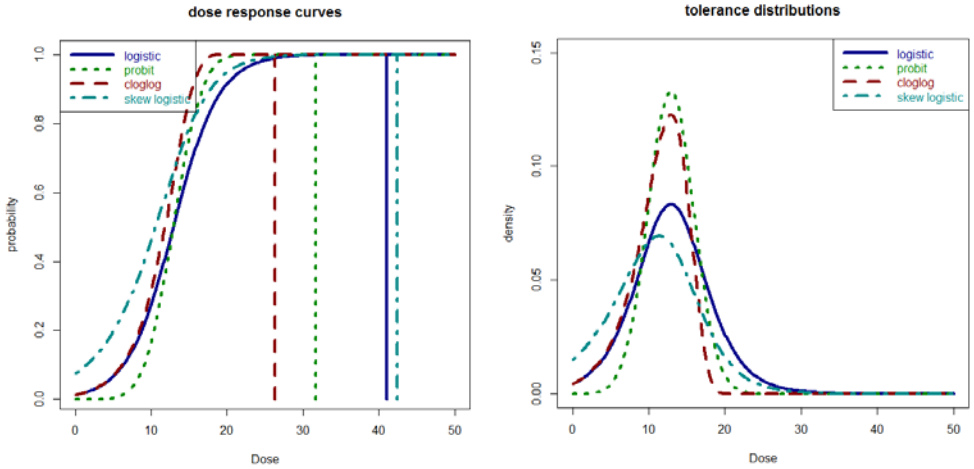
Phytosanitary Treatment (PT) studies use small-scale experiments to estimate dose requirements for attaining mortality at high levels. A moderate number of insects, say hundreds, are randomly divided into various treatment groups (including controls) and subjected to varying levels of treatment doses. This experimental process is often repeateda number of times so that the experiments are 'replicated in time'. The number of insects in each lot subjected to a treatment dose cannot always be measured directly but are estimated from the untreated lots though it is unlikely their numbers will be the same.Lots may consist of equal numbers of fruit but their weights will vary as will the insect numbers undergoing treatments. Therefore numbers of insects actually killed and survived are often only approximate.

Dose Estimation

Based on literature reviewed in this study it is common practice that the dose-response curve is modeled with a generalized linear model (glm) using the PROBIT procedure of SAS. The approximate number of insects killed at each dose is assumed to be Binomially distributed and a linkto the Binomial proportion$P$ is assumed to be linearly related to dosages of the treatment. The concept of the methodology for dose-response curve fitting relates a probability distribution for tolerances to dosages of the treatment(a tolerance distribution) to a nonlinear regression. More specifically, prediction for a proportion P of insects that fail to survive a dose D is based on the cumulative distribution function (CDF) of the assumed tolerance distribution. Thus the dose-response curve is based on the CDF of the tolerance distribution and the link to P is based on the inverse of its CDF. Users of SAS PROC PROBIT have the choice of Normal, Gompertz or Logistic tolerance distributions.The dose requirement $D$for $P \cdot 100\%$mortality can be determinedfrom the expression$F^{-1}(P) = a + b \cdot Dw$here F denotes the CDF for the tolerance distribution and F^{-1} its inverse. Regression coefficients a and $b$of the dose-response curve are relatedto the mean and scale parameters$\mu$and σ, respectively, of the tolerance distribution with $a = -\frac{\mu}{\sigma}$ and $b = \frac{1}{\sigma}$. The tolerance distributions used in this paper are listed in Table 1 with the corresponding dose-response curve. Included is the skew logistic model. This modelcorresponds to a tolerance distribution has an additional shape parameter and offers a more flexible model. We include it here for comparison to those offered by PROC PROBIT. The inverse function of the CDF gives the link function for the regression specified by the glm.

Model	Tolerance Distribution	Response-Curve
Probit	Normal	$P_D = \Phi(a + b \cdot D)$
Complimentary Log-Log	Gompertz	$P_D = 1 - exp\big(-exp(a + b \cdot D)\big)$
Logistic	Logistic	$P_D = \Big(1 + exp\big(-(a + b \cdot D)\big)\Big)^{-1}$
Skew Logistic	Skew Logistic	$P_D = \Big(1 + exp\big(-(a + b \cdot D)\big)\Big)^{-c}$

Table 1: Tolerance distributions and corresponding dose-response curve.



The choice of which dose-responsemodel to fit to the data is not cut and dry. The Probit model is a popular choice but the tolerance distribution for the Probitmodel is symmetric. If the researcher suspects that tolerance distribution may be skewedthen an asymmetric model such as the Gompertz (Complementary Log-Log) or the Skew Logistic model would be a more appropriate choice. We suggest trying both symmetric and asymmetric models and computing a Goodness-of-Fit statistic such a Pearson's Chi-Square and choose the best fitting model.Once the model is chosen and fitteda point estimate for dose requirement $D$for the desired efficacy (probit-9) is easily estimated using inverse regression. This is accomplished by setting the left hand side of the dose-response equation to the desired proportion and solving for D. Obtaining an interval estimate for D requires more involved computation. A common approach for obtaining an interval estimate for D and used with PROC PROBIT is Fieller's method. To construct a $(1-\alpha)\cdot 100\%$ confidence interval for $D$with this method, the expression $\frac{F^{-1}(P)-a-b\cdot D}{\sqrt{Var(a)+D^2\cdot Var(b)+2\cdot D\cdot Cov(a,b)}}$ is set equal to upper and lower $\frac{\alpha}{2}$th quantiles of the standard normal or student's t distribution and numerical methods are used to find the endpoints of the interval. When the data exhibit lack of fit the covariance of the coefficients a and $b$is usually scaled with an estimate of an overdispersion parameter and quantiles of the student's t distribution are used for obtaining the endpoints. There is no guarantee a solution to the expression exists for yielding either endpoint.

Large Scale Experiments

After thedose requirement from the small-scale studyis estimated,a large number of insects are usually collected in several batchesin order to test the estimated dose-requirementwith a Binomial test. The test involves treatingtheinsects with the dose requirement estimated from the small-scale study and the common decision rule is to conclude the dose effective if the dose kills all insects. The number of insects tested is selected to ensure at least a probit-9 efficacyat the 0.05 level of significance for no survivors.When no survivors are observed a one-sided $(1-\alpha)\cdot 100\%$ lower confidence bound on the true proportion P of insects killed can be obtained using the expression $\alpha^{1/n}$. Therefore the sample size requirement can be obtained from the inequalities $\alpha^{1/n} \geq P \rightarrow n \geq log(\alpha)/log(P)$.Sample sizes for y survivors remaining in general can be determined by solving

$$\sum_{i=n-y+1}^{n} \binom{n}{i} \cdot P^i \cdot (1-P)^{n-i} \leq \alpha$$

for n.Table 2 gives a summary of sample size requirements based on a decision rule for 0, 1 or 2 survivors observed using a Binomial test.

		number of survivors observed		
$P \cdot 100\%$	probit	0	1	2
99.7250	7.00	131	207	275
99.0000	7.33	299	473	628
99.9000	8.09	2,995	4,742	6,294
99.9900	8.18	29,956	47,437	62,956
99.9968	9.00	93,616	148,244	196,742

Table 2: Sample sizes required for testing $P \cdot 100\%$ efficacy at the 0.05 level of significance with the Binomial test

Literature Reviewed

Our literature review included select studies aimed at finding doses of treatments to control fruit flies for an assortment of commodities at the probit-9 level of efficacy (Table 3.) These were conducted using small-scale experiments to find the dose followed by a larger scale experiment to test the dose. Most of these studies involved hundreds or thousands of insects tested at each dose to fit a dose-response curve in the small-scale experiment. However, Armstrong et al. (1993 and 1995) used extremely large numbers per dose (10,000 to over 100,000) in their dose-finding experiments. Thus the term small-scale experiment may be used loosely in our discussions. In most cases a probit model was fit to data collected from the small-scale experiments to estimate the dose for probit-9 mortality. In all cases the number of insects tested at each dose was approximated based on the controls. We reported the range of mortality observed across the doses tested for each study (Table 3) to suggest the potential or lack of it for fitting a dose-response curve. In almost all of the studies reviewed dosages were replicated but the exact nature as to how they were replicated was not clear. We assume that the replication reported was based on repeating the experiment in time. In almost all cases counts were aggregated over the replications for dose estimation.

Study	Dose Tested	Commodity	# Tested	Range of mortality for doses tested	# Survivors	Dose Confirmed
Sharp et al. (1988)	65 min. in hot water	mango	147,993	75-100%	4	N
Gould (1990)	22 days in 5C air	starfruit	69,800	65-100%	14	N
Jessup, A. J. (1994)	14 days in 1C air	avacado	100,255	54-100%	1	N
Toba et al. (1991)	12 weeks in 0C air	apple	33,231	27-100%	3	N
Armstrong et al. (1995)	9 days in 1.1C air	starfruit	140,080	97-100%	1	N
Gould (1996)	12 days in 1.1C air	starfruit	107,221	32-100%	1	N
Corcoran et al. (1993)	30 min. in 45C air	zuchinni	178,219	32-100%	1	Y
Armstrong et al. (1993)	7 days in 1.1C air	starfruit	167,303	99-100%	1	Y
Hayes et al. (1984)	20 min. in hot water	papaya	82,089	82-100%	1	N
Jessup (1994)	14 days in 1C air	avacado	100,255	12-100%	1	N
Hallman et al. (1992)	40 g/m3	starfruit	104,303	31-100%	1	N

Table3: Literature on phytosanitary treatments of Tephritidae studied in this paper. All used Probit modeling to obtain estimate for dose tested.

Schortemeyer et al. (2011) point out the shortcomings of dose-finding studies such as reported above. These include:

1) No evidence of pilot studies before the reported study.
2) No discussion of how the number of organisms or the number of treatments, as well as the placement of doses, were selected.
3) No discussion of mortality in controls and how this affects modeling.
4) No discussion in the type of distribution selected for modeling, and how well the data obtained fit the model.
5) Confidence (fiducial) intervals are often reported, but their implications are seldom discussed in dose recommendations.
6) No discussion of how far a model can be meaningfully extrapolated beyond the range in the analyzed data set.

We found these shortcomings to be applicable to the studies in our literature review and we add the following:

7) No discussion of possible random effects of replication and their impacts on estimation.
8) Pooling over replicates to fit a dose-response model is a common practice but justification for doing so is not discussed.

Schortemeyer argues that the probit-9 efficacy standard for many pests such as those found in wood packaging materialfor such studies are often unrealistic and unachievableas the number of insects needed to test probit-9 mortality is prohibitive. Even so, the shortcoming listed above need to be understood. Our interests at the onset of this paper revolved around number 4) listed above. After literature review numbers 7) and 8) were raised as they related to 4).

To investigate 4) we fit each of the models listed in Table 1 to the data reported by each study. For each study reviewed, we report:the dose tested in the large-scale experiment, percent mortality achieved from the large-scale experiment, the best fitting model when applied to the small-scale experiment data, the estimated dose requirement, and Pearson's Goodness of Fit statistic and associated degrees of freedom (Table 4.)

Study	Dose tested	% mortality	Best fit	Dose estimate	Goodness of Fit (DF)
Sharp et al. (1988)	65 min	99.9973	skew logit	44 min	0.0 (2)
Gould (1990)	22 days	99.9921	skew logit	21 days	4.1(2)
Jessup, A. J. (1994)	14 days	99.9990	skew logit	17 days	3.5 (3)
Toba et al. (1991)	12 weeks	99.9940	Probit	11 weeks	3.0(4)
Armstrong et al. (1995)	9 days	99.9993	Logistic	11 days	19.1(3)
Gould (1996)	12 days	99.9991	Cloglog	13 days	46.1 (5)
Corcoran et al. (1993)	30 min	99.9994	Logistic	32 min	15.2 (3)
Armstrong et al. (1993)	7 days	99.9994	Logistic	6 days	7.0(4)
Hayes et al. (1984)	20 min	99.9988	skew logit	16 min	8.2 (1)
Jessup (1994)	14 days	99.9999	skew logit	17 days	3.5(3)
Hallman et al. (1992)	40 g/m^3	99.9990	Cloglog	38 g/m^3	8.3 (4)

Table 4: Best fitting models and Pearson's Chi-Square Goodness of Fit statistic reported with degrees of freedom (DF).

Simulations

Goodness of Fit

To investigate the importance of goodness of fit on model selection we generated 1,000 Monte Carlo trials to simulate Binomial samples Y_{D_i} from each of the Probit, Complimentary Log-Log and Logistic models for a sequence of doses D_i and for select sample size N_{D_i} at each dose D_i.Programming for simulation was done using the R language. The general form for all three models is given below with F^{-1} representing the appropriate inverse function for the CDF of the tolerance distribution.

$$Y_{D_i} \sim Binomial\left(N_{D_i}, \pi_{D_i}\right)$$
$$F^{-1}\left(\pi_{D_i}\right) = \alpha + \beta \cdot D_i$$

Parameters for each model were set so that the doserequirement for probit-9 efficacy was 25. The sequence of doses D_i selected were 4 to 14 in steps of 2. Table 5 lists the parameter values and inverse functions for each of the models. Data generated from each of these dose-response curves were fitted using the glm function of R. 95% confidence intervals for the probit-9 dose requirement were constructed using Fieller's method as previously described and Pearson's Chi-square goodness of fit statistic was computed. Coverage of the confidence intervals was computed from the percentage of intervals that contained the dose requirement of 25. The

percentage of Monte Carlo data sets each model fit best the data by having the lowest value of Pearson's χ^2 was also computed. Tables 6a-6c summarize results for these simulations.

Model	$F^{-1}(\pi_{D_i})$	Parameter values	
		α	β
Probit	$\Phi^{-1}(\pi_{D_i})$	$-8/3$	$4/15$
Complimentary Log-Log	$log\left(-log(1-\pi_{D_i})\right)$	$-2 \cdot cloglog(\Phi(4))/3$	$cloglog(\Phi(4))/15$
Logistic	$log(\pi_{D_i}/(1-\pi_{D_i}))$	$-2 \cdot logit(\Phi(4))/3$	$logit(\Phi(4))/15$

Table 5: Model parameters values for Monte Carlo simulation. $cloglog$ is used to denote the complimentary log-log link function and $logit$ the link function for the Logistic. They are given explicitly in the second column.

Number per dose	Data generated from Logistic					
	Coverage for model fit by			Proportion data sets best fit by		
N_{D_i}	logistic	probit	cloglog	logistic	Probit	cloglog
100	95.5%	7.2%	0.0%	59.3%	33.3%	7.4%
500	96.0%	0.2%	0.0%	81.9%	18.1%	0.0%
1000	95.6%	0.0%	0.0%	89.2%	10.8%	0.0%
5000	95.2%	0.0%	0.0%	100.0%	0.0%	0.0%
Mean dose requirement	25.0	20.2	16.4			

Number per dose	Data generated from Probit					
	Coverage for model fit by			Proportion data sets best fit by		
N_{D_i}	logistic	probit	cloglog	logistic	Probit	cloglog
100	1.5%	94.3%	6.1%	30.4%	56.9%	12.7%
500	0.0%	94.7%	0.0%	31.6%	67.3%	1.1%
1000	0.0%	95.3%	0.0%	27.9%	72.1%	0.0%
5000	0.0%	95.5%	0.0%	7.8%	92.2%	0.0%
Mean dose requirement	33.0	25.0	18.9			

Number per dose	Data generated from Gompertz					
	Coverage for model fit by			Proportion data sets best fit by		
N_{D_i}	logistic	probit	cloglog	logistic	Probit	cloglog
100	0.0%	4.2%	95.4%	23.1%	14.1%	62.8%
500	0.0%	0.0%	96.0%	6.8%	16.2%	77.0%
1000	0.0%	0.0%	96.6%	2.2%	11.3%	86.5%
5000	0.0%	0.0%	95.9%	0.0%	1.3%	98.7%
Mean dose requirement	51.3	35.1	25.1			

Table 6 a-c: Simulations were generated by the distribution specified in the first row header. Fiducial intervals were constructed after data were fitted to each of the Logistiic, Probit and Complimentary Log-Log dose-response curves and the coverage of these intervals reported for each sample size. Proportion of data sets best fit by reports the proportion of Monte Carlo trials for which the specified dose-response curve was the best fitting model as determined by Pearson's Goodness of Fit. The mean dose requirement taken over the 1,000 Monte Carlo trials for each model is reported for the largest sample size.

Tables 6a-6c demonstrate model selection is critical for estimation as illustrated by the mean dose requirement from each model. However choosing the model with the smallest Pearson's Chi-square statistic is not sufficient. Very large sample sizes are needed for Pearson's Chi-square statistic to be useful for distinguishing among models. These data suggest the number of samples per dose needs to be over 1,000for the correct model to have a good chance of being selected. Incorrect model selection almost guarantees the model fit will produce unreliable estimates. For example, when data are generated from a logistic model, estimates based on either probit or complimentary log-log models tend to be too low and the coverage based on these are inadequate for any sample size. These data also demonstrate the feasibility of extrapolating the dose requirement from test doses far below it as the coverages for the correct model when fitted are at or slightly above the nominal coverage of 95%. We reran the simulation for where the sequence of doses D_i were set to be 14 to 18 in steps of 2 for sample sizes of 500 to 5,000. We did thisto explore model selection when the dose levels correspond to a narrow range of mortalitiesbut simulations only included the Logistic model being the parent distribution with doses having an expected range of mortality between 0.8 and 0.98. We didn't report coverage from these intervals because Fieller's method failed to provide both endpoints for more than 10% of the trials and coverage was not really needed to demonstrate the need for careful dose placement. Proportion of data sets best fit by Pearson's Goodness of Fit are reported for each model along with the mean estimated dose requirement.

Number per dose	Proportion data sets best fit by		
N_{D_i}	logistic	probit	Cloglog
500	47.4%	8.6%	44.0%
1000	55.3%	11.9%	32.8%
5000	62.6%	21.3%	16.1%
Mean dose requirement	25.0	22.6	21.4

Table 6d: Results of data simulated of from Logistic model with expected probit-9 dose requirement of 25 and for data generated at doses 14(18)2 .

Table 6d. demonstrates that when doses are tested in the high but narrow range of expected mortalities model selection becomes more difficult with estimates of the dose requirement remaining appreciably different although less so than for doses with a greater range of expected mortalities.

Random Effects

To investigate the impacts of random effects and pooling across replications on estimation we performed another similar simulation study. We considered a random-effects logistic dose-response model for a randomized complete block design. The model was as follows:

$$
\begin{aligned}
&Y_{D_{ij}} | a_j, b_j \sim Binomial\left(N_{D_{ij}}, \pi_{D_{ij}}\right) \\
&logit\left(\pi_{D_{ij}}\right) = \alpha + a_j + \left(\beta + b_j\right) \cdot D_i \\
&a_j \sim N(0, \sigma_a^2) \\
&b_j \sim N(0, \sigma_b^2) \\
&j = 1, \cdots, r
\end{aligned}
$$

$Y_{D_{ij}}$ denotes a random binomial count for dose D_i conditional on the j[th] block effects and $Y_{D_{ij}}$ conditionally follows a binomial distribution with parameters $N_{D_{ij}}$ and $\pi_{D_{ij}}$. The $logit$ is the logit function $log(x/(1-x))$. The random effects a_j and b_j for the j[th] block follow normal distributions with 0 means and variances σ_a^2 and σ_b^2 respectively. Programming for simulation was done using the R language.

To study effects of pooling on obtaining an interval estimate on the dose requirement we generated 1,000 Monte Carlo trials to simulate a small scale dose-finding study with dose-response curve parameters $\alpha = -4$ and $\beta = 1/4$ using $r = 4$ replications with doses $D_i = 2(20)(2)$ and $N_{D_{ij}} = 500$ for all i doses and j blocks. We set both random effects parameters σ_a^2 and σ_b^2 to 0 to simulate data with no random effects of replicates and computed 95% confidence intervals on the dose requirement using Fieller's method as previously described.

With no random effects to condition on, the model is an unconditional and is known as a marginal model. The probit-9 dose requirement corresponding to these parameters is 57.4. For each Monte Carlo trial we constructed confidence intervals before and after pooling over the replicates. The results are reported in Table 7.

	Comparison of Pooling vs. Not Pooling using 1,000 Monte Carlo trials when no random effects of replicates are present			
	Mean	Mean Lower Limit	Mean Upper Limit	Coverage
Not pooled	57.39	56.04	58.84	0.951
Pooled	57.39	56.00	58.87	0.949

Table 7: Comparison of 95% fiducial intervals for estimating the probit-9 dose-requirement before and after pooling. Coverage reported is the fraction of intervals containing the true requirement of 57.4.

Simulations summarized in Table 5 demonstrate that if no random effects of replicates are present then interval estimates are essentially the same with coverage adequately close to the target of 0.95whether the data are pooled over replicates or not . This result can be explained by a property of binomial random variables which is that the sum of binomial random variables sampled from the same distribution with parameter$P$will itself be binomially distributed with parameter P. What is unexpected is the test doses of 2 to 20 were far from the probit-9 dose requirement of 57.4 which supports the argument that extrapolating an accurate estimate of the dose requirement is conceptually possible. We note that parameter values for this simulation were chosen arbitrarily as were the test doses and there was a bit of luck involved in choosing these parameter values. The simulation often failed computationally to give intervals when test doses of 40 to 60 (the upper quantiles of the tolerance distribution) were used. Thissubstantiatesthe argument that careful planning of dose placement is needed even when the dose requirement is known to be in a certain range. Although this topic is not explored here insight to planning dose response studies to ensure might be found in Freeman (1970) and Hu el al. (2010).

To study effects of pooling over replicates withrandom effects on estimation we generated 1,000 Monte Carlo trials as before but set the parameter σ_a^2 to 1 to emulate random effects of the replicates. This model can be described as a random intercepts logistic regression model with intercepts varying among replicates and implies the dose requirement varies among replicates.Modeling logistic regression models with random effects can be accomplished using generalized linear mixed models (GLMM) or generalized estimating equations (GEE).The choice of which of these to use depends on the question to be answered. GLMM are conditional models and provide estimates suited to answer question “What is the dose requirement for the typical replicate?” whereas GEE are marginal models suited to answer the question “What is the dose requirement for the entire population of replicates?”The latter is a GEE approach. Using either approach is defensible but care is needed when interpreting results. We used a GEE approach akin to that which would be gotten to those familiar with PROC PROBIT in SAS for

fitting dose-response models.The consequences of random effects on estimation of marginal parameters usingGEE models are described in McCulloch and Searle (2001). In fact the marginal distribution of the $Y_{D_{ij}}$ will not follow a Logistic distribution but can be well approximated by one with parameters $\alpha^* = \alpha/\sqrt{1+\lambda\cdot\sigma_a^2/r}and\beta^* = \beta/\sqrt{1+\lambda\cdot\sigma_a^2/r}$ where $\lambda = 256/(75\cdot\pi)$. The impact that random effects have on the dose requirement is that it gets larger with the variance of the random effects as the parameter values α^* and β^*becomesattenuated with increasing values of σ_a^2. We computed the marginal expected dose requirement (the dose requirement needed for all replicates to comply to the probit-9 standard) to be 62.68 based on approximation formulas given in McCulloch and Searle (page 107) and used this value to estimate the coverage of the Fieller intervals based on this approximation.Results are presented in Table 8.

	Comparison of Pooling vs. Not Pooling using 1,000 Monte Carlo trials when random effects of replicates are present			
	Mean	Mean Lower Limit	Mean Upper Limit	Coverage
Not pooled	62.93	53.54	79.93	0.925
Pooled	62.93	61.07	64.96	0.337

Table 8: Comparison of 95% fiducial intervals for estimating the probit-9 dose-requirement before and after pooling. Coverage reported is the fraction of intervals containing the true requirement of 57.4.

Whether or not data are pooled the dose requirement estimated from the fitted dose-response curve will be the same. However the coverage of the confidence intervals from pooled data will be far too low because variation among replicates is averaged out. These intervals are intended to provide estimates of the dose requirement that achieve a requiredlevel of mortality P and are based on a mean level of mortality.Therefore these estimates will not ensure with a specified level of confidence that the level of mortality achieved will be P or more for that dose. However, large-scale confirmatory experimentsare used to testjust that. Thus the methodology of the small-scale study to estimate the dose followed by a large-scale Binomial test is statistically flawed.We followed up on this by modifying the simulation in Table 6a to include a simulation for a large-scale follow-up sample to test the dose estimate for the Logistic model. Only 7% of the Monte Carlo trials resulted in no survivors (the requirement for 95% confidence the dose will have at least probit-9 mortality) when the dose estimate $D$was tested after simulating a sample from $Binomial\left(N = 93{,}616,\ P = \left(1+exp(-a+b\cdot D)\right)^{-1}\right)$. A different methodology is needed that can place a bound on the dose requirementobtained from a small-scale study that will have a high assurance level it can be validated.

Summary

It is clear from the phytosanitaryliterature that predicting doses for extrememortalityis a challenge. Our literature review included select studies for controlling fruit flies on various commodities that nearly met the probit-9 efficacy standard(Table3) based on doses estimated from a Probit model.

A primary objective of this paper was to address the question as to whether model goodness of fit is critically important for estimating a dose requirement. Goodness of fit for fitted models was not reportedin the studies we reviewed. Wecomputed Pearson's χ^2 for the pooled data that was reported but only two of these (Toba et al. 1991) and (Jessup 1994) fit a Probit model which was the standard for these studies. Another objective was to compare other models to the probit. For most studies reviewed (8 of 11) alternative models from among the Logistic, Skew Logistic or Complimentary Log-Log were able to fit the data with no significant lack of fit but we note again that this was after pooling and that only pooled data were available(Table 4.) Nevertheless estimates of the dose requirements for probit-9 efficacy were in most cases close to the doses tested and the level of mortality was close to the probit-9 level intended.

We also simulated data from specific dose-response models and fit these data to mispecified models in order to explore the sensitivity of Pearson's Chi-square to select the appropriate model. Our simulations reveal that rather large samples per dose (5,000 or more) are required in order to select the model that generated the data (Table 6a-c).Mispecifiedmodels produced poor estimates of the true dose requirement in our simulations as indicated by the average estimate and coverage of confidence intervals constructed using Fieller's method. These simulations suggest that Goodness of Fit is important for dose estimation but this presumes real data can be expected to be generated in a similar process to what we used in simulation. If the underlying population sampled for a real-life dose-response study consists of mixtures of insects from different age classes, genetics, or species then we cannot expect models used in our simulations to be useful for estimating dose requirements.

A final objective was to identify experimental conditions or methodology that would cause estimation of the dose required for the desired control of insects to fail. Our literature review revealed several common practices that are problematic to estimation including insect counts at test doses being estimated from controls, doses tested over an inadequate range of mortalities, disregarding possible random effects associated with experimental replicates and aggregating insect counts across these. The most problematic of practices is attempting to confirm a point estimate of the dose requirement obtained from the small-scale experiments by testing it against a large-scale sample for verification that it exceeds the level of desired mortality (probit-9) with a specified level of confidence, usually 95%. A point estimate obtained from a small-scale experiment cannot be expected to be confirmed from a large-scale

experiment since the point estimate is for a dose associated with a mean level of mortality. If the underlying tolerance distribution associated with the dose response is symmetric then we could expect that the dose estimate would pass the confirmatory test at the very most half the time. Simulations suggest that these estimates will pass the confirmatory tests with low probability. Methods are needed to place lower limits on the dose requirement so that a specified level of control can be assured, usually 95% as this is the established standard of phytosanitary control for commodities susceptible to infestation of fruit flies.

References

Sharp, J. L., M. T. Ouye, R. Thalman, W. Hart, S. Ingle & Victor Chew. 1988. Submersion of 'Francis' mango in hot water as a quarantine treatment for the West Indian fruit fly and the Caribbean fruit fly (Diptera: Tephritidae). J. Econ. Entomol. 81: 1431-1436.

Gould, W. P. & J. L. Sharp. 1990. Cold-storage quarantine treatment for carambolas infested with the Caribbean fruit fly (Diptera: Tephritidae). J. Econ. Entomol. 83: 458-460.

Jessup, A. J. 1994. Quarantine disinfestation of 'Hass' avocados against Bactrocera tryoni (Diptera: Tephritidae) with a hot fungicide dip followed by cold storage. J. Econ. Entomol. 87: 127-130.

Toba, H. H. & H. R. Moffitt. 1991. Controlled atmosphere-cold storage as a quarantine treatment for nondiapausing codling moth (Lepidoptera: Tortricidae) larvae in apples. J. Econ. Entomol. 84: 1316-1319.

Armstrong, J. W., S. T. Silva & V. M. Shishido. 1995. Quarantine cold treatment for Hawaiian carambola fruit infested with Mediterranean fruit fly, melon fly, or oriental fruit fly (Diptera: Tephritidae) eggs and larvae. J. Econ. Entomol. 88: 638-687.

Gould, W. P. 1996. Cold treatment, the Caribbean fruit fly, and carambolas, pp. 489-493. In B. A. McPheron and G. J. Steck [eds.], Fruit fly pests: a world assessment of their biology and management. St. Lucie Press, Delray Beach, Fla.

Corcoran, R. J., N. W. Heather & T. A. Heard. 1993. Vapor heat treatment for zucchini infested with Bactrocera cucumis (Diptera: Tephritidae). J. Econ. Entomol. 86: 66-69.

Armstrong, J. W., S. T. Silva & V. M. Shishido. 1995. Quarantine cold treatment for Hawaiian carambola fruit infested with Mediterranean fruit fly, melon fly, or oriental fruit fly (Diptera: Tephritidae) eggs and larvae. J. Econ. Entomol. 88: 638-687.

Hayes, C. F., H. T. G. Chingon, F. A. Nitta & W. J. Wang. 1984. Temperature control as an alternative to ethylene dibromide fumigation for the control of fruit flies (Diptera: Tephritidae) in papaya. J. Econ. Entomol. 77: 683-686.

Jessup, A. J. 1994. Quarantine disinfestation of 'Hass' avocados against Bactrocera tryoni (Diptera: Tephritidae) with a hot fungicide dip followed by cold storage. J. Econ. Entomol. 87: 127-130.

Hallman, G. J. & J. R. King. 1992. Methyl bromide fumigation quarantine treatment for carambolas infested with Caribbean fruit fly (Diptera: Tephritidae). J. Econ. Entomol. 85: 1231-1234.

Schortemeyer, M., et al. (2011). "Appropriateness of probit-9 in the development of quarantine treatments for timber and timber commodities." Journal of economic entomology 104(3): 717-731.