Final Report: A time-course model to allow the prediction of the optimal period after box closure to commence resampling in King Scallop boxes in Scottish offshore waters.

Prepared for:

Food Standards Agency Scotland
Aberdeen, Scotland UK

Prepared by:

The ENVIRON Health Sciences Institute
ENVIRON International Corporation
Arlington, Virginia USA
Ruston, Louisiana, USA
London, England UK
Edinburgh, Scotland UK

November, 2004

EXECUTIVE SUMMARY

ENVIRON International has created and updated a database of information pertaining to domoic acid (DA) concentrations in king scallops in Scottish offshore waters. In addition to geospatial, bathymetric, and some phytoplankton data, the database contains DA concentration measurements taken between July, 1998 and March, 2004.

That database has been used to investigate the time-course of DA concentrations in boxes that have been "closed" due to concentrations exceeding a risk-based threshold of 20 μg DA per g of tissue. Both gonad and whole tissues were examined; only samples collected up to the minimum of 335 days after closure or the date of reopening were included. The probabilistic analysis that is reported here used the statistical modeling approach of logistic regression to predict, based on the observations in the database, the likelihood that samples taken from closed boxes at various times after closure will have DA concentrations less than 20 μg/g (referred to as a "low concentration").

One of the major efforts of the analysis was to determine what factors might affect the time-course of interest. It was determined that the following factors played an important and significant role, in the sense of modifying the probability of obtaining a low-concentration sample from a closed box:

1. the area (as taken from the FSAS designation, being E, M, NM, etc.) in which the closed box is located;
2. the month in which closure occurred; and
3. the DA concentration causing closure.

These factors were determined to modify the probability of obtaining a low-concentration sample at any time after closure. The magnitude of their effects, and the specific effect on the time-course probability predictions was determined by finding the best-fitting logistic regression models that included some function of time after closure. Models that provided a satisfactory fit to the data were found; the predictions (in both graphical and tabular form) of those best fitting models have been made available in an accompanying Excel spreadsheet named LOGISTIC_PREDICTIONS.xls.

Implementation of the modeling results can be based on the LOGISTIC_PREDICTIONS.xls spreadsheet. If one wants to base the timing of resampling of a closed box on the likelihood of reopening it, then the predicted probabilities in that spreadsheet can be used. To successfully reopen a box, one needs 4 consecutive samples (2 per day on 2 separate days at least seven days apart) all with low DA concentration. If, for example, one wanted the probability of successfully sampling for reopening to be at least 50% (P = 0.5), then that means that there would need to be a high probability (in this example, about $0.84 = (0.5)^{1/4}$) that individual samples will have low concentrations. The spreadsheet tabular read-outs of the individual-sample probabilities can be used to approximate when that will be the case.

The time after closure when such a high probability will occur depends on the factors of area, month of closure, and level causing closure, as indicated above. In general, however, it appears that at short times after closure, e.g., less than 60 days and in some instances well beyond 60 days after closure, the probability of getting low-concentration samples may not be sufficiently high to offer much likelihood of successful reopening. Individual instances (based on area, month of closure, and level causing closure) should be evaluated on a case-by-case basis using the spreadsheet predictive tools, but when the likelihood of successful reopening is estimated to be relatively low, then allocation of resources to other sampling efforts (e.g., monitoring of open boxes that would still be yielding scallops for consumption) may be more cost-effective and health-protective.

The logistic regression analysis was successful in identifying models that appeared to be consistent with the data included in the database. There are uncertainties associated with those models, as there are with any modeling effort. We have identified several key follow-on tasks that would help characterize or reduce those uncertainties, such as refinement of the definition of variables in the models, consideration of additional covariables, calculation of bounds on predicted probabilities, and elaboration of the modeling approach. Of greatest importance, however, is the continued augmentation of the database used for this analysis. Additional data are already available (from April to October 2004); those data offer the rare opportunity to test the model predictions against data that were not used to fit the models. Such a model validation effort would identify if and how the models need to be improved. In general, an on-going effort to maintain the database and integrate the new data with the old (ideally, in terms of using those data to update the model predictions) has been identified as a key effort for improving the ability of FSAS to allocate sampling resources.

INTRODUCTION

In a previous work assignment, ENVIRON International created a database of information pertaining to the measurement of domoic acid (DA) in king scallops in Scottish offshore waters. The database contained information on the areas (designated by one- or two-letter identifiers, e.g., E, M, NM, IS, etc.) where samples were collected, the specific "boxes" within those areas in which data were collected, date of collection, and DA concentrations. Certain bathymetric data were included for most boxes (minimum depth, maximum depth, and average depth).

That database has been expanded as the first task in this work assignment. Data collected from October of 2003 to March of 2004 were added. In addition, sample size information (number of scallops pooled to create a sample that was then analyzed for DA concentration) was added for some of the more recent observations. However, the major effort under this work order has been the analysis of the data, with respect to the timing of sampling in closed boxes and the corresponding probability that a box would be re-opened.

In this analysis, the main issue under consideration was to determine if it is possible to suggest optimal times after closure of a box at which resampling may be done. It was desired to estimate resampling dates that would improve "efficiency" by reducing the number of times such resampling returned the verdict that the box should remain closed. From the standpoint of Food Standards Agency Scotland (FSAS), resampling a closed box too soon, before natural forces or typical patterns have had a chance to reduce DA levels in scallop tissues to levels considered safe, means extra effort and money spent to no avail. That is, such sampling does not change the status of a box from closed to open, so in some sense the resampling might just as well not have been undertaken. On the other hand, to maximize the utilization of the scallop fishery, closure of boxes should not be extended beyond the time period necessary to achieve the reduction in DA concentrations necessary to return scallops to a safe condition. Hence it is important to attempt to optimize expenditure of resources by conducting resampling when there is a good chance of affecting some change in box status.

To date, decisions on the closure of boxes have been based on DA concentrations in gonad tissues. Hence, the analyses reported below have examined the patterns in gonad DA concentrations. But FSAS has expressed the desire to investigate the use of whole-tissue samples for closure determinations. While there are fewer measurements of whole tissues present in the database, we have applied the same analysis approaches to deal with the whole tissue data as we have applied to the gonad samples. In both cases, a DA concentration of 20 μg/g has been taken as the "threshold" for determining closure.[1] In the following, mention of "low concentrations" or "low-concentration samples" refers

[1] Technically, the 20 μg/g threshold for whole tissues is now associated with the issuing of a shucking advisory. A DA concentration from a box in excess of 250 μg/g in whole tissues would result in closure of the box. But, for the purposes of this analysis, the 20 μg/g threshold is what defines "closure" for either gonads or whole tissues.

to that threshold, for both tissue types; a low concentration is one that is less than 20 μg/g.

The issue of resampling as a function of time is complicated by the fact that there appears to be a great deal of variability with respect to the timing of DA concentration change. On top of that, there is variability across samples on any given date; i.e., the concentrations in samples collected in the same box on the same day will vary, and it is possible that one sample will indicate DA levels in the safe range while another sample will indicate that DA concentrations are too high. It is not clear even that the important sources of or contributors to variability over time have been identified. From the perspective of this analysis, that has entailed that a substantial part of the effort has been to find those parameters that best describe or contribute to observed patterns of DA concentration and that the variability not accounted for by such parameters is treated as random variation. The manner in which the parameter investigation and random variation were treated is described in the Methods section.

The remainder of this document includes Methods, Results, Discussion, and Conclusion sections. The Methods section defines the analysis methodology, the assumptions made for that methodology to apply, and the specific approaches applied in the case of DA in scallops. The Results section provides the major findings with respect to resampling of closed boxes. These findings are presented in terms of the probability of a sample yielding a concentration estimate below the cut-off point for closure (20 μg/g). The Discussion section suggests the ways in which the probability estimates presented in the Results section can be used to aid in the determination of the time for resampling closed boxes. Graphical aids are provided as are supplemental spreadsheets that can be used by FSAS to calculate the desired probabilities in specific instances. Recommendations for follow-on work are also provided. Finally, the Conclusion section summarizes all of the findings and offers observations about the timing of resampling in closed boxes.

METHODS

Two sets of analyses were completed, distinguished by tissue type. One set of analyses was based on gonad tissue, with the 20 μg/g cutoff. The other, independent set of analyses looked at DA concentrations in whole tissue. While box closure has not heretofore been based on whole tissue DA concentrations, FSAS has expressed an interest in determining the effect of doing so. For the second set of analyses, only whole tissue concentrations were examined, with gonad tissue concentrations ignored. Box "closure" was determined by DA concentrations in whole tissue, with 20 μg/g also serving as the cutoff for that determination here as well.

The focus of this analysis was on sampling in closed boxes. Therefore, the data points that were used in this analysis were those that came from closed boxes. For our purposes, we desired to know the date of closure, and we needed to know when a box was reopened. Thus, for these analyses, the following definitions applied:

Date of closure: the day on which a sample (of the appropriate tissue type) from a particular box was determined to have a DA concentration greater than or equal to 20 μg/g, if the preceding two sample dates that were at least seven days apart had all concentrations for those dates less than 20 μg/g.

Date of reopening: the day on which two consecutive dates of sampling (at least seven days apart) from a particular box returned all concentrations for those dates less than 20 μg/g.

Observations that were between the date of closure and the date of reopening, inclusive, for a given box were included in the analyses. The following points are consequences of those definitions:

1. A box that had a DA concentration greater than or equal to 20 μg/g as one of the first two sampling dates in the database did not contribute observations to the analysis until after that box had two consecutive dates of sampling (at least seven days apart) that had all concentrations below 20 μg/g. In such instances, one could not determine when closure occurred (it was sometime before data for that box were available) for those observations at the "beginning" of the record for that box, so those observations were not included in the analysis.
2. Any concentration measurement was included in at most one closure for one box. That is, each observation could be associated with at most one closure period, so there was no overlap in counting, and the probabilities estimated by the methods described below could be determined without fear of "double counting."
3. A variable defined as the days since closure, referred to as "DAYS" below, could be defined. The minimum value for that variable was zero; concentrations measured on the same day as the date of closure, but not including the concentration that determined closure, were included in the analysis. Operationally, on the date of closure, the maximum concentration (by definition, greater than or equal to 20 μg/g) was ignored, but all other concentrations for that day were considered for the analysis.
4. A maximum value for DAYS did not, in theory, exist. It was possible that a box would never satisfy the criteria for the definition of the date of reopening (which otherwise would define the maximum value of DAYS for a given closure incident in a given box) so that the date of reopening would never occur. In the database, we observed concentrations that would have been included in our analysis that had DAYS up to and including 1321 (i.e., more than 3.5 years for a box to fail to be reopened by the above definition) for gonad concentrations. For whole tissues, similar large values of DAYS included values equivalent to more than 4.5 years.
5. A box could contribute more than one set of observations if that box was closed, then reopened, and then closed again.

In practice, the analyses included only those observations for which DAYS was less than or equal to 335. This decision was made for the following reasons. First, it was recognized during preliminary examination of some of the data that there was the strong potential for yearly cycles in DA concentration. Thus, if a box had remained closed for nearly a year, then it was likely that another cycle of changes in DA concentration similar to the one just completed (during the first year of closure) would begin anew. Such

cyclic patterns might very easily mask any patterns over shorter time periods. That is, any model that modeled the probability of getting concentrations below 20 μg/g as a function of days since closure might not fit the patterns over the shorter term as it tried to fit all of the data over the longer term. Second, it seems likely that FSAS would be more interested in the best estimates for the probability of getting a sample below 20 μg/g in the time period before one year after closure, because there would be "pressure" to reopen closed boxes as soon as safely possible. The specific value of 335 days was selected so that no observation would enter the analysis that was in the same month (but one year later) as the closure date for the box from which that observation was taken (365 days in a year minus 30 days per month, roughly). There were only 16 gonad samples and 28 whole tissue samples collected between 330 and 365 days, inclusive, after the date of closure, so the impact of changing the cutoff date for inclusion of samples in the analysis would be minimal.

Because the main interest was in whether or not a sample would fall above or below 20 μg/g DA, as input to a decision about reopening a box, the dependent variable of interest that was the object of all our analyses was defined as follows:

Y = 1 if the DA concentration was less than 20 μg/g
= 0 if the DA concentration was greater than or equal to 20 μg/g.

This definition of the outcome of interest was also prompted by the fact that some of the concentration measurements were recorded as being "> 250" which makes treatment of concentrations as a continuous variable problematic. Dichotomization of the concentration results was not hampered by the cutoff of some samples at 250 μg/g and it still captured the event of interest, i.e., whether a low concentration was or was not obtained, and the estimation of the probability of such low concentrations being observed.

Given that definition of Y as the dichotomous variable of interest, we were interested in estimating the following probabilities:

$$P(Y = 1 \mid DAYS = d; Z = z),$$

that is, the probability of getting a sample with a DA concentration below 20 μg/g on day d after closure, where Z represents the covariates (variables other than the number of days since closure, DAYS) that might help determine the probability of interest, and z indicates the value of those other parameters. Note that Z and z might be vectors (i.e., Z might consist of more than one covariate related to the probability of interest).

From the database of sample values that was compiled and presented to FSAS in the first task of this work assignment (and which was extended with more recent data as

the first step in the current task), each of the observations[2] included in the analysis had the following variables available (or derived) for it.

Variable	Description
Y	The dichotomous outcome variable of interest; Y = 0 or Y = 1.
BOX	The box from which the sample was collected.
DAYS	The number of days since the date of closure of the box from which the sample was obtained (0 $\leq$ DAYS $\leq$ 335).
AREA	The major areas in which the box was located (e.g., M, S, O, NM, etc.).
COAST	Coast from which the sample was obtained; COAST = east for areas E, M, O and S; COAST = west for all other areas.
MON	The month in which the box from which the sample was obtained was closed.
LEV	The concentration of DA that caused closure of the box from which the sample was obtained; LEV $\geq$ 20 μg/g.
MINDEPTH	Minimum depth (meters) of the box from which the sample was obtained.
AVGDEPTH	Average depth (meters) of the box from which the sample was obtained.
MAXDEPTH	Maximum depth (meters) of the box from which the sample was obtained.

The parameter LEV was treated as a continuous variable. The concentration that initiated closure of a box was never one recorded as ">250" so it was possible to unambiguously assign a concentration level associated with closure of a box. The values of LEV ranged from 20 to 140 μg/g for gonads and 20.7 to 112 μg/g for whole tissue.

A common approach used for the analysis of dichotomous outcome variables (e.g., Y in our case) is logistic regression. Logistic regression looks very similar to the well-know linear regression approach, but it accounts for the fact that the outcome is not continuous (Y is either 0 or 1) and that the probability of interest (that Y = 1) must fall between 0 and 1. The logistic regression analyses used here for all of the following is given by

$$\text{Logit}[P(Y = 1 \mid \text{DAYS} = d; Z = z)] = \beta_0 + \beta_1 * f(d) + \beta_Z * z$$

where logit[x] = x/(1-x), the β coefficients are to be estimated, f(d) is some function of d (e.g., d^2), and β_Z and z may be vectors. The right side of that equation looks like the familiar linear regression. But, if we let p denote the probability of interest on the left hand side of that equation, solving for p yields

[2] Some observations were lacking the bathymetric measurements that determined the values of the parameters MINDEPTH, AVGDEPTH, and MAXDEPTH.

$$p = \exp\{\beta_0 + \beta_1 * f(d) + \beta_Z * z\} / [\, 1 + \exp\{\beta_0 + \beta_1 * f(d) + \beta_Z * z\}],$$

which can be seen to fall between 0 and 1, as desired. The statistical software package SAS (PROC Logistic) was used to solve for the β coefficients and to give the diagnostics of fit used to pick the best fitting model from among those considered. As will be seen below, those models included ones with different functional forms for f(d) (to determine the function of days since closure that best describe the change in probability of a low concentration) and they considered different sets of possible covariables which make up the vector Z.

Maximum likelihood methods were used for fitting the models to the data. Models were compared to one another via the associated values of the maximized log-likelihood, L. When one model is a special case of a second model (e.g., when the second model includes all the parameters of the first model plus some others) it is known that the maximized likelihood of the second model will be at least as large as that of the first model. Formal comparison of the two models is possible by comparing

$$-2(L_1 - L_2),$$

where L_i is the log-likelihood of model i, to a chi-square distribution with v degrees of freedom (where v is the number of extra parameters in the second model, compared to the first). Such statistical comparisons were used as part of the model selection effort.

RESULTS

Gonads

There were a total of 1787 measurements of DA in gonad tissues of scallops collected from closed boxes eligible for the analyses. As stated above, that number includes only those samples collected between 0 and 335 days (inclusive) from the date of closure.

For most of the analyses reported here, only 1779 observations were used for the logistic regression modeling of gonad concentrations. Eight samples from areas E and IS were excluded from most considerations. Those areas were excluded because of the relatively few samples collected from closed boxes there. In area E, box E03 was closed once and all samples collected during the closure periods after the day of closure (2 collected 48 days after closure and 1 collected 163 days after closure) were below 20 μg/g. The sample collected on the day of closure (not the one that caused closure) had a concentration of 29 μg/g. Similarly, box IS14 was closed twice, and all four samples collected during the closure periods (collected on days 7 and 21 after one closure and on days 111 and 119 after the other closure) were below 20 μg/g. Ignoring these 8 samples

out of 1787 would not have a major impact on the results.[3] In the areas remaining in the analyses (H, J, M, NM, O, and SM) there was more than just one closed box that contributed observations to the analyses. Thus, for those areas, it is likely that estimates of the probability of a low concentration are more robust than would be the case if only one box contributed to the observations. The numbers of data points in those areas, by month of closure, are summarized in the following table:

	Area						
Month of Closure	**H**	**J**	**M**	**NM**	**O**	**SM**	**All**
Jan		5	3	15	5	14	42
Feb	5			8		5	18
Mar		25			3	11	39
Apr	30			27		8	65
May	10	5		22	25	18	80
Jun		43		31	45	21	140
Jul		132	22	40	110	172	476
Aug	34	63	42	112	117	110	478
Sep	24	45	12	9		33	123
Oct	40	42	10	31	24	55	202
Nov		15		16	44	20	95
Dec	5			13	3		21
Total	148	375	89	324	376	467	1779

A total of 1132 of the 1779 observations included in the analysis of gonad tissues had DA concentration less than 20 μg/g. The following description summarizes the process of determining the best set of covariables for estimating the probability that such a low concentration would arise.

The modeling of interest always included DAYS, because the primary interest of the analyses was to examine time-trends in DA concentration following closure. Thus, the base model would include some function of the DAYS variable, but it was not clear *a priori* if the best description of the observations would be attained using DAYS itself or some alternative function of DAYS. The alternatives examined were power functions of DAYS, $DAYS^x$, where the values of x examined were ½, 1, 2, 3 and 4. These five choices span curve shapes that are steeper for lower values of DAYS (i.e., concave down, for $DAYS^{½}$), that are linear (for $DAYS^1$), and that have concave upward shapes of various degrees of curvature (for $DAYS^2$, $DAYS^3$, and $DAYS^4$). The likelihoods associated with the fits of these basic models are given in the following table:

Model: logit(p) = $\beta_0 + \beta_1$*$DAYS^x$ x =	**Maximized Log-Likelihood**

[3] The formal reason for the exclusion of these observations was that logistic regression analyses are adversely impacted by variables that completely (or almost completely) separate the Y=0 observations from the Y=1 observations. Because all but one of the observations in these two areas had Y=1, the SAS PROC Logistic routine flagged AREA as a parameter that might lead to such separation. In order to maintain the same set of observations for comparison of all models fit to the data, the observations in boxes E03 and IS14 were excluded from all models, even those that did not include AREA as a covariable.

½	-1161.23
1	-1159.1
2	-1156.48
3	-1156.14
4	-1157.04

Recalling that a larger log-likelihood indicates a better fit, it can be seen that the model with $DAYS^3$ provides the best fit, with the model with $DAYS^2$ nearly as good. Models including both those functions of DAYS were examined further. In addition, even though $DAYS^1$ was not among the best, we included models with that function of DAYS just because it may appear to be a more natural choice. Note however, that the modeling pursued here is not intended to uncover some underlying physical or biological reality so that choice of the function of DAYS to include is merely intended to maximize the prediction of the probability of getting low concentrations of DA in scallops collected from closed boxes. So, even though we have included $DAYS^1$ in the modeling exercise, that function of DAYS should not be automatically preferred *a priori* over any other function of DAYS.

A preliminary investigation of the impact of including the bathymetry variables (MINDEPTH, AVGDEPTH, and MAXDEPTH) suggested that MINDEPTH was at least as good as the other two variables in terms of matching the observed proportions of low-concentration samples when combined in a logistic model with $DAYS^1$. So, subsequent analyses reported here examined only MINDEPTH and not AVGDEPTH or MAXDEPTH, for all three functions of DAYS.

The procedure for determining what other variables to include in the modeling was a forward selection procedure, where, one at a time, other variables were added to the model and it was determined if such an addition significantly improved the fit of the model to the data (determined by the change in the maximized likelihood). When several possible additional variables significantly improved the fit, the one that was "most significant" (had the lowest p-value associated with the difference in the likelihoods) was added and then other possible additions were considered subsequently. Adding of variables stopped when no addition had a p-value associated with the change in likelihoods less than 0.10.

The results of the model selection described above for the models starting with the three functions of DAYS are summarized below:

	Base Model (Function of DAYS)					
	$DAYS^1$ (-1159.1)		**$DAYS^2$ (-1156.48)**		**$DAYS^3$ (-1156.14)**	
Step	Var. Added	p-value	Var. Added	p-value	Var. Added	p-value
1	MON (-1102.86)	5.8E-19	MON (-1101.64)	2.1E-18	MON (-1103.21)	1.2E-17
2	AREA (-1085.65)	1.9E-06	AREA (-1084.72)	2.5E-06	AREA (-1086.61)	3.4E-06
3	LEV (-1081.77)	0.0054	LEV (-1081.02)	0.0065	LEV (-1083.06)	0.0077
4	$DAYS^1$xLEV (-1066.89)	4.8E-08	$DAYS^2$xLEV (-1065.53)	2.6E-08	$DAYS^3$xLEV (-1066.29)	7.0E-09

5	$DAYS^1$xMON (-1052.44)	0.0024	$DAYS^2$xAREA (-1057.42)	0.0062	$DAYS^3$xAREA (-1057.82)	0.0046
6	--		$DAYS^2$xMON (-1048.73)	0.097	--	

Notes: In parentheses are the log-likelihood values associated with the corresponding models.

The table above can be understood as follows. Take the $DAYS^1$ base model as an example. With $DAYS^1$ in the model, the other four variables considered for these analyses (AREA, LEV, MON, MINDEPTH) were examined to see if they improved the fit of the base model. MON was the most significant with respect to improving the fit to the data (observe the big change in log-likelihoods from -1159.1 to -1102.86, with an associated p-value of 5.8E-19, i.e., 5.8×10^{-19}, where a smaller p-value is associated with a bigger improvement in the fit of the model), so MON was added to the model. With MON in the MODEL the other variables were examined to see if, with $DAYS^1$ *and* MON in the model, they still significantly improved the fit.[4] The AREA variable did so, and so it was added in step 2. Similarly, the LEV variable improved the fit of the model that included $DAYS^1$, MON, *and* AREA, and so was added in step 3.

At that point, for the $DAYS^1$ base model and both of the other base models, all of the "main effects" (covariables under consideration, other than MINDEPTH) were included in the model, and so interaction terms were considered next. The interaction terms that were considered were all of the two-way interactions of the DAYS function and the other main effects included in the model. Interaction terms are designated by something of the form "RxS" (e.g., $DAYS^1$xLEV, the interaction of $DAYS^1$ and the LEV variable). Because three main effects in addition to the DAYS function were included in all the models, there were three possible interaction terms to consider.

In the case of the $DAYS^1$ base model, the first interaction to enter the model was $DAYS^1$xLEV, followed by $DAYS^1$xMON. The $DAYS^1$xAREA interaction did not enter the model, because with the main effects and the other interactions in the model, $DAYS^1$xAREA did not significantly improve the fit.

Note that the three base models differed in terms of addition of variables with respect to the interaction terms that were added. For $DAYS^2$, all interactions of other main effects with $DAYS^2$ were significant. For $DAYS^3$, the first interaction to enter the model was the one with LEV (as for the case of $DAYS^1$) but the second one to enter involved AREA (unlike the case for $DAYS^1$) and that was sufficient (i.e., $DAYS^3$xMON did not add significantly to the fit once $DAYS^3$xAREA was included, unlike the $DAYS^2$ model). The three base models ended up including different sets of variables and interactions. The impact of those different sets with respect to model fit and to predictions of probabilities of low DA concentrations at various times after closure are examined below.

[4] The variable MINDEPTH was uniformly the least important variable, for all the models that we examined. It never added significantly to the fit of the base model; p-values ranged from 0.14 to 0.17 across the base models examined, including those that used the COAST variable instead of AREA (see below). So, MINDEPTH was no longer considered after the first step for all of the base models shown here and below.

But first some mention of the need for interaction terms may be necessary. Addition of only the main effects changes the fit of the model to the data by shifting up or down the straight lines describing the logits as a function of DAYS (or some function thereof). That is, they merely change the intercept (or, in other words, make the intercept a function of the main effects) while keeping the slope constant. This can be seen from the equation for the logits given in the Methods section:

$$\begin{aligned} \text{Logit}(p) &= \beta_0 + \beta_1 * f(d) + \beta_Z * z \\ &= (\beta_0 + \beta_Z * z) + \beta_1 * f(d), \end{aligned}$$

after rearrangement of the terms (for d days after closure and covariable values z). Since $\beta_Z * z$ does not involve DAYS when Z (the vector of covariables) includes no interaction terms, this shows that the same slope (β_1) is assumed no matter what the values of the covariables may be (i.e., for all areas, months of closure, and levels that caused closure).

Conversely, suppose that the Z vector includes interaction terms; take DAYSxLEV as an example and assume that it is the only interaction term in the model. Then the logit equation can be rewritten

$$\begin{aligned} \text{Logit}(p) &= \beta_0 + \beta_1 * d + \beta_Z * z \\ &= \beta_0 + \beta_1 * d + \beta_{Z'} * z' + \beta_{DL} * d * l \\ &= (\beta_0 + \beta_{Z'} * z') + \beta_1 * d + \beta_{DL} * d * l \\ &= (\beta_0 + \beta_{Z'} * z') + (\beta_1 + \beta_{DL} * l) * d, \end{aligned}$$

where l is the value of LEV and Z' is the vector of covariables other than DAYSxLEV (in this example, including no interaction terms) which have values given by the vector z'. This rearrangement shows that the DAYSxLEV interaction introduces a variable-slope model, a model where the slope for DAYS ($\beta_1 + \beta_{DL} * l$) varies and depends on the value of LEV.

Note that at the end of the third step of the model selection process, all of the base models had added all the main effects. Interestingly, the $DAYS^1$ base model, which started out with the lowest log-likelihood, was at that point fitting the data slightly better than the $DAYS^3$ model. But, all of the base models added two or more interaction terms (starting with $DAYS^x$xLEV in all cases) before the selection process was completed. For all of those models, the probability of obtaining a sample with DA concentration less than 20 μg/g at any given time after closure depended on the area in which that sample is to be collected, the month closure occurred, and the level of DA that caused closure. Moreover, because of the interactions, the rate of change of that probability also depended on two or more of those factors. Specific model predictions are discussed next.

Figures 1-3 display the fits of the base models to the data from all areas, months of closure, and levels causing closure. Recall that the base models do not take into account any of these factors, so the model predictions can be presented on a single plot, for each function of DAYS. Note that the observations themselves have been

"summarized" in the following sense. The observations are actually either 0's or 1's; plotting 1779 points that are either 0 or 1 would produce an uninterpretable jumble of points. Instead, Figures 1-3 display the proportion of times a low-concentration sample was obtained, over several ranges of the DAYS variable. The ranges were selected arbitrarily, but they do function to show the overall trends in the proportion of samples with low concentrations. The data underlying those figures are summarized here:

Range of DAYS	Number of Samples	Proportion of Samples with Low Concentration
0	128	0.57
1-19	129	0.65
20-40	250	0.67
41-80	379	0.55
81-120	322	0.66
120-200	382	0.63
201-335	189	0.78

The number of samples per range is reasonably large for all ranges; the ranges more towards the middle have somewhat greater sample size. Since this is merely for the sake of illustration and has no impact on the model selection or fitting, we considered this breakdown to be adequate. The range that is composed of only DAYS=0 was included to show that there were many samples collected on the day that a box was closed, and that even then more than half of the samples collected had low DA concentrations.

While the plots and the above table suggest an overall trend of increasing proportions of low-concentration samples, it is easy to see that there is considerable "scatter," that the change in proportions of low-concentration samples is not monotonically increasing. Some of that may be due to the fact that those plots lump together all areas, months of closure, and levels causing closure, which have been shown in the model selection to be important covariables. On the other hand, the stochastic nature of the sampling and measurement of DA (leading to a random component in the determination of whether a sample has low concentration or not) means that even after adjusting for those covariables, the data will continue to show some degree of scatter.

The base models do a reasonable job of matching the data. They all show the increasing trend and they do tend to be "close to" the proportions shown. Of course, the models are constrained to be monotonic with DAYS, and so they can not reflect all of the nonmonotonicities in the observed proportions, nor would we necessarily want them to. The $DAYS^1$ base model has a more steady increase in predicted probability of a low-concentration sample; it matches the DAYS=0 observation very well. The $DAYS^2$ and $DAYS^3$ base models show a slower rate of increase for shorter times after closure (and so they tend to "split the difference" between the observed proportions for those shorter times) while they increase more steeply as time progresses. Those two models provide a better fit to the proportions at later times (nearly bisecting the line representing the proportion in the last DAYS range, which would represent the ideal relationship between the model predictions and the observed proportions plotted in this way). The $DAYS^3$ base model does also intersect the line representing the proportion in the penultimate DAYS range (although it does not come close to bisecting it), and so it is perhaps not

surprising that the $DAYS^3$ base model gave the greatest likelihood of these three models (indicating the best fit).

Because the model selection discussed above suggested that area, month of closure, and level causing closure were significant covariables, it becomes difficult to display all of the comparisons of model predictions and observations. That is the case because each area (there were 6 of them in the analysis) and month of closure (all 12 of them) would have a different plot (yielding 72 such plots). Moreover, since LEV is a continuous variable, it is not feasible to plot the comparisons of the model predictions to the observations broken down by values of LEV (although LEV could be categorized and plots, by area and month of closure also, could be developed for the LEV categories).

Instead, to illustrate the fits of the selected model for each function of DAYS, we have selected three combinations of area and month of closure (ones that have a reasonable number of sample points) to illustrate the fits for those selected models. The areas (H, O, and SM) were selected to be from differing geographical regions of the Scottish coast and the months of closure (April, July, and October) were selected to span a good part of the year (recognizing that most of the observations were obtained from closures that occurred in the summer months). The specific combinations examined were of observations from area H having an April closure date; from area O having a July closure date; and from area SM having an October closure date. The visual comparisons are displayed in Figures 4-12. In those figures the observed values themselves (the "0" or "1" observations that reflect samples having greater than or less than 20 μg/g DA, respectively) have been added to the plots. Proportions calculated from those observations (as described above) have been included as well.

Those figures display the variety of trends and model predictions that can be attained. For area H with closure starting in April, it appears that the likelihood of obtaining a sample with low concentration actually decreases as time after closure increases. While the best $DAYS^1$ model does reflect that downward trend (Figure 4), the other two models predict very slight increases in the probability over time (Figures 5 and 6). The trend in the proportion of samples with low concentrations appears to be increasing in area O when closure is in July and in area SM when closure is in October (Figures 7-9 and 10-12, respectively). All the models appear to describe the increasing pattern for area O. The best $DAYS^2$ model, however, predicts decreasing probability of a low-concentration sample as time after an October closure increases in area SM; the other two models show an increasing trend for that area and month of closure.

These nine examples are a small set of the matches between observations and model predictions that have gone into the model selection and parameter coefficient estimation process. Even here, the impact of the LEV covariable was only approximated – for the model predictions it was set equal to the average for the corresponding area and month of closure. Some discrepancy between observed trends and model prediction may be due to that simplification; in actuality, each observation and the corresponding model prediction would depend also on the specific level causing closure.

So it would be inappropriate to judge among the models on the basis of these plots alone. Indeed, the parameter estimates obtained for the selected models were derived by finding the values that maximized the likelihood of obtaining *all* the observations, across areas, months of closure, and levels causing closure. So, discrepancies shown here are the result of parameter estimation that does best overall, but may deviate from the observations in particular instances, as we know they must because of the randomness of the observations and the fact that there are undoubtedly other factors (potential covariables) besides those at our disposal that affect the likelihood of obtaining a sample with low DA concentration.

Three statistical measures of fit are presented here to compare the three models and help determine if any of them do an adequate job of fitting the data overall. The first two, the R^2 value and the concordance rate, are not associated with a formal test of fit. The third, the Hosmer-Lemeshow goodness-of-fit statistic, does actually compare the observed and expected number of low-concentration samples, and uses a chi-squared-based test to see if the expected number is acceptably close to the observed number. The values for those measures for the three selected models are shown in the following table:

	Statistical Measure		
	R^2	**Concordance (%)**	**H-L Statistic (p-value)**
DAYS[1] Selected Model	0.120	69.5	11.11 (0.20)
DAYS[2] Selected Model	0.124	69.6	10.52 (0.23)
DAYS[3] Selected Model	0.115	69.1	5.46 (0.71)

The R^2 value relates to the amount of variation in the observations that is explained by the model. All three selected models can account for approximately 12% of the total variation. That is a relatively small amount, but as discussed above, it is not surprising given the amount of scatter displayed in the example graphics (Figures 1-12) and the fact that there may be some important explanatory variables that we have not considered (e.g., temperature information is one possibility that one might want to consider).

Concordance measures the degree to which observations that have different predicted probabilities of a low-concentration sample are "properly sorted." That is, it looks at all pairs of observations with different response values (0 and 1 in this analysis) and determines for each pair if the observation that has the value 1 has a greater model-predicted probability of having a value of 1 than the observation with the value 0. If that is the case, that pair of observations is concordant; otherwise it is discordant (or tied if the members of the pair have the same model-predicted probability). All three models have similar concordance rates; between 69% and 70% of the pairs are concordant.

The Hosmer-Lemeshow goodness-of-fit statistic extends the common and familiar chi-squared tests of fit to models that have continuous covariables (as these models do with the function of DAYS and with LEV). It is based on a partition of the observations into approximately 10 groups based on their predicted probability of a low-concentration sample (roughly, the observations with the lowest ten percent of the predicted probabilities go in the first group, and on up to the observations with the highest ten percent of the probabilities). Then the observed numbers of low-concentration samples

can be compared to the model predicted expected number of such samples, and a chi-squared test (with 8 degrees of freedom) can be applied. For our models, the match between the $DAYS^3$ selected model and the observations is strikingly better than that for the other two models; a smaller p-value is indicative of a poorer fit. The excellent match between the observed and predicted numbers for the $DAYS^3$ model is summarized in the following table:

	Low-Concentration Samples		Higher-Concentration Samples	
	Observed	**Expected**	**Observed**	**Expected**
Group 1	64	68.17	114	109.83
Group 2	90	83.78	89	95.22
Group 3	95	92.42	84	86.58
Group 4	104	100.28	74	77.72
Group 5	103	109.13	75	68.87
Group 6	111	115.43	68	63.57
Group 7	128	122.62	52	57.38
Group 8	126	132.58	52	45.42
Group 9	150	147.67	28	30.33
Group 10	161	159.90	11	12.10

Based on the similar proportion of variability explained by the three models, on the similarity of the concordance measures, and on the clearly better fit by the $DAYS^3$ selected model, that model would appear to be the best choice for use in planning dates for gonad tissue collection in closed boxes.

As noted at the beginning of this section, 1779 observations were used in the above model estimation procedures, because observations in two areas (E and IS) almost completely separated the Y=0 observations from the Y=1 observations (all but one observation in those areas had a low DA concentration). In addition, the above models use area as a covariable, so application of the model results to areas other than those included in the database would be problematic. As an alternative to the modeling above, one that allows use of the 8 observations from areas E and IS and that avoids the problematic extrapolation to other areas, we completed logistic regression modeling using COAST in lieu of AREA as a covariable.

Based on the favorable results obtained when $DAYS^3$ was the chosen function of DAYS, the model selection procedure for the "COAST" model used $DAYS^3$ as well. When the selection procedure described above was applied, the main effects of MONTH, then COAST, then LEV entered the model. The only interaction that significantly improved the fit was the $DAYS^3$ x LEV interaction. The final model had an R^2 of 0.097, concordance of 67.6%, and p-value for the Hosmer-Lemeshow goodness-of-fit test of 0.41. While these statistics indicate an adequate fit, they are slightly worse than the corresponding estimates for the selected $DAYS^3$ model that uses area. As one would expect, lumping areas into coasts causes some loss of resolution in the model predictions, but the "COAST" model is still useable.

For comparison purposes, Figure 13 displays the model predicted time-course of the probability of a low concentration for area SM (a west-coast area) with October being

the month of closure. The predictions use the average level causing closure in that month and area (36.06 μg/g). The model prediction shown and the data can be compared to those in Figure 12, which has predictions for that same area and month of closure, but based on the area-specific model. Some differences in the model predictions in Figures 12 and 13 are apparent; from Figure 13 one would estimate a slightly lower initial probability of getting a low-concentration sample, and the rate of increase in the probability is also slightly less.

Whole Tissues

The analysis of whole tissue DA concentration data proceeded as described above for the analysis of the gonad concentrations. The database contained many fewer measurements of DA in whole tissue; a total of 345 observations were available from closed boxes, for 335 days or less from the date of closure. Of those samples, only 102 (less than one third) were found to have DA concentrations less than 20 μg/g. It appeared that if a box were closed because of whole tissue DA concentration, the overall likelihood of subsequently obtaining low-concentration samples that would justify re-opening was fairly small. That observation does not imply that there are necessarily no trends with time since closure; indeed, given a low overall probability of finding low-concentration samples from closed boxes, it might be even more important to determine the best time to collect additional samples.

In the case of the whole tissue analyses, it was not clear that the bathymetry variables were the least informative (as was the case for the analysis of the gonad samples). Thus, analyses were conducted and model selection procedures were pursued that included the bathymetric variables. Unfortunately, only 310 of the 345 whole tissue observations were from boxes that had such bathymetric data; 35 observations from boxes E01, E33, IS09, and IS14 were excluded from the analyses because of missing data.

The initial examination of the functions of DAYS that appeared to best fit the data on low-concentration whole tissue samples showed the following patterns:

Model: logit(p) = $\beta_0 + \beta_1$*DAYS^x x =	**Maximized Log-Likelihood**
½	-209.092
1	-208.038
2	-205.962
3	-204.519
4	-203.676
5	-203.257
6	-203.128
7	-203.203
8	-203.42

Note: The log-likelihoods shown were calculated using all 345 of the whole tissue samples. Conclusions about the best function of DAYS based on these values were considered satisfactory for determining the best function for the portion of the whole-tissue database having bathymetry data.

As that table shows, the best choice for a function of DAYS would be $DAYS^6$, the function that gave the largest log-likelihood. Higher powers of DAYS were not as good. In addition to $DAYS^6$, we also investigated $DAYS^2$ and $DAYS^3$, since those choices were also part of the analysis of the gonad concentrations.

The three bathymetric variables were examined with those three functions of DAYS to see if one was clearly better than the others. In terms of the likelihoods, AVGDEPTH, the average depth of a box, was uniformly better, over the three functions of DAYS, as shown by the log-likelihoods in the following table:

	Function of DAYS		
Bathymetric Variable	**$DAYS^2$**	**$DAYS^3$**	**$DAYS^6$**
MINDEPTH	-186.191	-185.159	-184.401
AVGDEPTH	-178.862	-178.279	-177.822
MAXDEPTH	-188.105	-187.3	-186.706

Because the variable AVGDEPTH gave larger log-likelihoods than the other two bathymetric variables, no matter the choice of DAYS function, AVGDEPTH was used in the following model selection procedure as the only bathymetric parameter.

As discussed above in connection with the analysis of the gonad concentrations, variables that completely or almost completely separate Y=1 observations from Y=0 observations invalidate the maximum likelihood estimation procedure used to fit logistic regression models. For the whole tissue data, this was an issue for the MON variable; some values of MON had all Y=0 observations and so MON as defined above could not be used in the analysis of the whole tissue data. But because the timing of closure was determined to be so important in the analysis of the gonad data, we wanted to retain some variable that represented such timing issues. In this case, we decided to lump certain months together to avoid the problem. So, for this analysis, the months of January through April were lumped together, and November and December were lumped together. All other months (May through October) were considered separately from one another.

Given that definition of the explanatory variables, the 301 observations considered here (of which 99 had low DA concentrations) were distributed by area and month-group as follows:

	Area							
Month of Closure	**C**	**E**	**J**	**M**	**NM**	**O**	**S**	**All**
Jan-Apr	1	11	5			2	3	22
May		1		8	3		3	15
Jun		8		10		47	5	70
Jul	3			8		40		51
Aug		11		5		60		76
Sep							12	12
Oct		27		7		3	9	46
Nov-Dec				3	5	10		18
Total	4	58	5	41	8	162	32	310

The forward selection process of covariables to include with the chosen functions of DAYS proceeded as summarized here:

	Base Model (Function of DAYS)					
	$DAYS^2$ (-190.5)		**$DAYS^3$ (-189.5)**		**$DAYS^6$ (-188.8)**	
Step	Var. Added	p-value	Var. Added	p-value	Var. Added	p-value
1	MON-group (-141.0)	1.67E-18	MON-group (-140.7)	3.24E-18	MON-group (-140.4)	5.13E-18
2	AREA (-113.2)	3.54E-10	AREA (-113.4)	5.71E-10	AREA (-113.9)	1.13E-09
3	AVGDEPTH (-109.6)	0.0072	AVGDEPTH (-109.7)	0.0065	AVGDEPTH (-110.1)	0.0059

Notes: In parentheses are the log-likelihood values associated with the corresponding models.

All three base models added the same main effects, which included the bathymetric variable AVGDEPTH, but did not include the concentration level causing closure (LEV). The month-group entered first in each case followed by area and then AVGDEPTH.

Interactions of the function of DAYS with the three main effects shown above as well as with LEV were checked to see if they significantly improved the fit of the models to the data. Neither the interaction with AVGDEPTH nor the interaction with LEV significantly improved the fit. The interactions with area and with month-group could not be carried out because of the separation problem mentioned above. Apparently, when one of those interactions is included in the models, some combination of the DAYS function and the other term (say area) is such that all the Y=1 observations can be separated (or nearly separated) from the Y=0 observations. This would be the case if all the observations above (or below) some value of DAYS were all Y=1 (or all Y=0), but the inclusion of the interaction terms means that that cut-off value could be different for every area. In any case, it was not possible to evaluate the interactions of the function of DAYS with area or month-group, so the models shown in the above table are taken to be the selected models, since LEV and AVGDEPTH did not add significant interactions.

The three base model fits to the proportions derived from the observations are shown in Figures 14-16. Again, as in the case of the analysis of the gonad samples, these base models do not account for the covariables that were determined to be significant contributors to the fit of the model. But they do give an indication of the overall trend in the data and the fact that the initial probability of obtaining a low-concentration sample is low and only increases after some considerable time has passed. Given the higher degree of curvature achievable by the $DAYS^6$ model compared to the $DAYS^2$ or $DAYS^3$ models (compare Figure 16 to Figures 14 and 15), it is not surprising that the former model was the best fitting of the base models. For these graphs, the observed proportions were defined with reference to the following DAYS ranges and observed frequencies of low-concentration samples:

Range of DAYS	Number of Samples	Proportion of Samples with Low Concentration
0	32	0.38
1-58	69	0.30
68-160	73	0.22
163-240	46	0.24
241-300	47	0.26
301-335	43	0.63

The goodness-of-fit statistics for the selected models for the three functions of DAYS are summarized in the following table:

	Statistical Measure		
	R^2	Concordance (%)	H-L Statistic (p-value)
$DAYS^2$ Selected Model	0.420	90.4	1.90 (0.98)
$DAYS^3$ Selected Model	0.420	90.2	5.30 (0.73)
$DAYS^6$ Selected Model	0.419	89.9	6.88 (0.55)

The fit statistics in that table indicate that the selected $DAYS^2$ model appears to give a slightly better fit to the data, based primarily on the Hosmer-Lemeshow goodness-of-fit test result. That is the case even though the base model for $DAYS^2$ had the lowest maximized likelihood of the three models considered. With the addition of the covariables, the $DAYS^2$ model was able to provide a better match to the data, as indicated by the following table of observed and expected numbers:

	Low-Concentration Samples		Higher-Concentration Samples	
	Observed	Expected	Observed	Expected
Group 1	0	0.16	31	30.84
Group 2	1	0.89	31	31.11
Group 3	2	1.32	28	28.68
Group 4	2	1.86	29	29.14
Group 5	3	3.13	28	27.87
Group 6	6	5.94	25	25.06
Group 7	14	14.25	18	17.75
Group 8	18	19.96	13	11.04
Group 9	25	25.05	6	5.95
Group 10	28	26.44	2	3.56

The selected $DAYS^2$ model may be the best option for planning the sampling of whole tissues from closed boxes.

As in the case of the gonad concentration data, only some of the areas are represented in the database. Analogous to the treatment of gonads, we implemented the model selection process when COAST was used in place of AREA. This was done for the function of DAYS that was the best fitting when using area, i.e., $DAYS^2$. The model selection process found that month-group was again the first covariable to enter the model. But in this case, with no AREA variable, LEV was the next variable to enter, followed by AVGDEPTH. COAST was not a significant contributor; apparently

designation of the coast from which a sample was obtained did not supply as much information about the probability of a low-concentration whole-tissue sample as did designation of the area from which it came. However, LEV was significant in this model whereas it was not in the model with AREA; LEV may have supplied some of the information previously contained in the AREA covariable, which would happen if those two variables were correlated to some extent. The final model selection step added the $DAYS^2$ x AVGDEPTH interaction, a term that was not significant in the model that used AREA. So, the final version of the whole-tissue, non-area model did not contain any covariable related to geographical location, but the impact of the bathymetric parameter was more pronounced. The R^2 for this model was 0.36, the concordance was 87.3%, and the p-value for the Hosmer-Lemeshow goodness-of-fit test was 0.087. None of these values are as good as that for the selected $DAYS^2$ model that used AREA. The goodness-of-fit p-value, in particular, indicates a fit that is adequate but not superior. When possible (see the discussion below) the prediction of the probability of collecting low-concentration whole-tissue samples should use the selected model with the AREA covariable.

DISCUSSION

Implementation

The purpose of the analyses discussed above was to help determine when sampling might be done in closed boxes, with the goal of not sampling so soon that there would be little likelihood of collecting enough samples with a low DA concentration (less than 20 μg/g). On the other hand, it would be desirable to not delay sampling when there is a reasonable chance of collecting low-concentration samples. The results presented above can be used in the following manner to help in that regard.

It is assumed for this exposition that the usual practice is to collect two samples from a closed box on any given day, with two such collections at least seven days apart being required for re-opening (assuming all four samples have DA concentrations less than 20 μg/g). The logistic regression models discussed above predict the probability that a *single* sample will have DA concentration less than 20 μg/g, as a function of days since closure and other covariables. Thus, one way to use the logistic modeling results is to determine the earliest day after closure such that the likelihood that 2 samples on that day and 2 samples seven days later will all have low DA concentrations is greater than P, where P is some target probability that is high enough to justify sampling.

Consider, as an example, P = 0.5, i.e., one wants to have at least an even chance that the four samples will have low concentrations. The probability that all four have low concentrations is just the product of the probabilities that each one will have low concentration.[5] A change of seven days will not change the model-predicted probability of low concentration much, so we can simplify the task a bit by finding the day after

[5] This calculation assumes that the observations are independent of one another.

closure where chance of a low concentration is $(0.5)^{1/4} = 0.84$. Such a choice will ensure that the estimated probability of getting four low-concentration samples will be at least 0.5. If some other value of P is deemed appropriate, substitute that value in the equation $P^{1/4}$ to find the day of choice.

The day that gives the calculated probability, $P^{1/4}$, can be found by using the spreadsheet, LOGISTIC_PREDICTIONS.xls, which accompanies this report. That spreadsheet has four pages, two for gonads and two for whole tissues, each of which calculates the probabilities required for the corresponding tissue using the models indicated above to be the preferred choice for that tissue. Each page has a plot that shows the predicted probabilities, as a function of the days since closure and as a function of the covariables included in the model. So, for example, for the gonad concentration model, the user will have to enter the area of the closed box, the month in which closure occurred, and the DA level that caused closure. The plot indicates how the probabilities for getting single low-concentration samples change and the corresponding numerical results can be used to identify a specific time. Those numerical results are in column J, which gives the probability in question, and K, which gives the corresponding day on which that probability occurs.

As an example, suppose that a box is closed in July due to a gonad tissue DA contamination level of 30 μg/g and that the box is in area SM. Figure 17 shows the plot that is obtained from the "Gonads Model" page of LOGISTIC_PREDICTIONS.xls when the user enters the required input (AREA = SM, MONTH=7, LEV=30). Visually, it appears that to have P = 0.5, as in the above example, one must wait until about day 280 after closure to get the desired single-sample probability of a low-concentration gonad sample, 0.84. Looking at the spreadsheet set up for that input, one can confirm that, indeed, it is predicted to take 278 days to get to that probability of a low-concentration sample.

That procedure works well when the probability of getting a low-concentration sample increases as time since closure increases. There are cases, as illustrated in Figure 18 for the hypothetical case of a closure in a box in area NM in June due to 22 μg/g of DA in gonad tissues, where the probability is predicted to decrease with time. In those instances, the model predicts that sampling ought to occur as early as possible to maximize the chance of reopening. However, it may never be the case that a P even as high as 0.5 could be attained. In the example shown in Figure 18, the maximal P would be about 0.08, if sampling was done 7 and 14 days after closure (where 0.08 is derived by looking at the spreadsheet and noting that for NM in June, with a 22 μg/g closure level, the probability of getting a low-concentration sample is about 0.534 on both day 7 and day 14 after closure; $(0.534)^4 = 0.08$).

Similar examples and derivations can be obtained for whole tissues using the "Whole Tissues Model" page of LOGISTIC_PREDICTIONS.xls. For those calculations, area and month of closure input is required, but in addition the average depth of the closed box must be input as well (but not the DA level causing closure), based on the logistic regression model selection process that identified the covariables that appeared to

significantly improve the fit of the model to the observations. Average depth for most boxes can be found in the database that was created in the first task of this project.

The modeling that was done for the analyses reported here used a fixed date of closure and referenced subsequent results (about whether a suitably low concentration was observed) to that date of closure. That was appropriate in the sense that the goal was to determine the likelihood of getting a low-concentration sample at any time after closure, where the presence of any samples collected earlier in the closure period were ignored. That is, we were interested in estimating the probabilities as if there were no prior samples collected.

Operationally, for implementation in the field however, it would appear that the best course of action would be to update the predictions whenever subsequent collection in a closed box resulted in samples with DA concentrations at or above 20 μg/g. That is, if a box was closed and then subsequently (some number of days later) samples were collected in that box with at least one of those samples having DA concentration of at least 20 μg/g, then for future planning of collection, the data input into the logistic regression models should reflect the most recent information regarding closure, and the "days since closure" should be calculated from the date of that most recent data collection. For example, suppose a box is closed in June because of a DA concentration of 25 μg/g. That box has samples collected from it in July, and one of those samples has DA concentration of 34 μg/g. Then, in terms of using the logistic regression results for planning subsequent sample collection, one should input "July" as the month and "34" as the level causing closure. That is so because, unlike the process by which the logistic regression models were developed, one can not ignore the most recent samples and determine what might happen if they did not exist.

The models have been developed for certain ranges of the covariables. For example, the level of DA in gonads causing closure ranged from 20 to about 140 μg/g; for whole tissues the corresponding range was about 20 to 112 μg/g. Model estimates were based on fitting data within those ranges of "closure concentrations." Care must be exercised when using the model with inputs outside this range. For any modeling effort, extrapolation beyond the range of the fitted observations can be problematic, so users must be aware of that model predictions may be more uncertain when such extrapolation is necessary.

Besides the level causing closure, the only other extrapolations that might be necessary for the gonad concentration modeling is to other areas not included in our analyses. It is for this reason that we provided the alternative analysis that used COAST as a covariable in place of area. In practice, it might be best to use the COAST-based model to predict the probability of a low-concentration sample from an area not specifically included in the data set of gonad tissue data (areas H, J, M, NM, O, and SM were represented in the fitted data set). Implementation in such instances can be accomplished using the page in LOGISTIC_PREDICTIONS.xls called "Gonads (COAST) Model." All boxes can be assigned to a coast (east – being areas E, M, S, and O – or west), so there should be no extrapolation when using that model for predictions.

Similarly, LOGISTIC_PREDICTIONS.xls provides a page called "Whole (COAST) Model" that can be used when the area for a box under consideration is not one of those included in the fitted data (which were C, E, IS, J, M, NM, O, and S). The box-specific input, average depth, ranged from 11 to 87 meters in the fitted data; the user should be aware of possibly greater uncertainty in model predictions when making predictions for boxes that have average depth outside that range. Moreover, as discussed above, the whole-tissue "COAST" model looks to have substantially less predictive ability than the selected whole-tissue model. In fact, despite the fact that we have identified the page as the "Whole (COAST) Model" (by analogy to the gonads page and because of the process by which the model was selected), the "Whole (COAST)" model does not include the COAST variable (nor does it include any variable related to geographical location).

Follow-On Investigations

The previous discussions mentioned uncertainty in model predictions. These uncertainties were from identifiable causes, such as extrapolating beyond the data used to fit the models. But there are other sources of uncertainty that have potentially greater impact. Those are often referred to as "model uncertainties" and have to do with the limitations imposed by the models chosen and the inherent uncertainty of estimating parameters of those models from a sample of observations.

Identification and quantification of such uncertainties is an effort worthy of additional follow-up. We have not discussed or considered confidence limits on the parameters of the models (e.g., the coefficient for $DAYS^x$ in our models) that would affect the predicted probabilities of obtaining a low-concentration sample. Nor have we considered functions of DAYS other than the power functions presented above. While those functions do represent a range of curve shapes (e.g., having different curvatures as nicely illustrated in Figures 1-3 and 14-16) they do not include all the possible shapes. Of particular interest might be models that are not monotonic, for which the predicted probability can increase and then decrease again as time since closure increases, or *vice versa*.

Logistic regression modeling offers the opportunity to identify the most important observations and those that may appear to be outliers. These model "diagnostics" can be important for determining how to improve or modify the modeling effort and for interpretation of the results. Such diagnostics have not been included in the analyses to date; their consideration could add to the usefulness and to the determination of the "robustness" of the predictions that have been derived.

The only interaction terms included in the models were those involving the function of DAYS and the covariables. Other interactions, among the covariables themselves, could be examined. These would not affect the slope of the logistic regression curves, but they could make the intercepts more "case sensitive." By that we mean, for example, that instead of having an intercept determined by a MON term (that

would apply across all areas) and an AREA term (that would apply across all months), one could get intercepts terms specific for each month/area combination. It would be interesting to see if such interactions significantly improve the model fits.

These models have considered the time of closure only in terms of the month in which it occurred. This was consistently one of the most important covariables that entered into the models. But perhaps there are other representations of the closure date (based on knowledge of some associated meteorological or bathymetric variables) that might better account for the patterns of DA concentrations in closed boxes. This may be particularly important for the analysis of the whole tissue concentrations, where some grouping of months was required to avoid the separation problems discussed above (Y=1 observations completely or nearly separated from Y=0 observations by one or more of the explanatory variables); in lieu of the grouping that was used here (January through April in one group and November and December in another) there may be a better way to accomplish that grouping that is informed by such other considerations.

One note about the time of closing relates to the use of a "year" covariable. These analyses intentionally did not include year. Such a variable might have helped to explain more of the variability exhibited among the observations because year might correlate with other factors that could influence the course of DA concentrations, such as temperature, days of sunshine, or other types of environmental disruption. However, in terms of predictions, such a covariable would have been useless – all future applications of these modeling efforts will be for years not included in the database, so there would be no way to make sense of such a variable for any implementation.

Rather, as noted in passing above, direct input and use of additional bathymetric or meteorological data (ones that might be correlated with "year") would be the way to extend the models and potentially improve model predictive power greatly. Then, those inputs, if deemed to be significant contributors to model fit, could be used in any specific implementation. Moreover, in light of the importance of the box-specific bathymetric measurement, average depth, for predicting the probability of a low-concentration whole-tissue sample, additional examinations of the available bathymetry data (or alternative data that might be recommended) with respect to the gonad concentrations might be warranted. These and other possible extensions to the modeling should be examined for feasibility.

The application of the logistic regression modeling to the whole tissue data must be viewed as more uncertain generally than the modeling for the gonad tissues. There were many more gonad samples (by nearly a factor of 6) in closed boxes and these samples were over more years of observation. Although the selected models for whole-tissue concentrations appeared to fit the data well (as judged by the R^2 and concordance values as well as the goodness-of-fit test), the constraints on that modeling imposed by the separation issues discussed above have restricted the interactions that were able to be considered. Continued updating of the whole-tissue model should be pursued and would be expected to resolve some of these issues as more data become available.

Finally, a note in relation to additional data. Since the completion of the analyses, more data have been received from FSAS covering the period from April through early October of 2004. Those data have not been included in the analyses reported here. Clearly, they could serve the purpose of extending the whole-tissue concentration data set that was mentioned in the preceding paragraph. In fact, for both gonads and whole tissues, those additional observations could be used in two ways.

First, they could be used as a "test set" to determine how well the model predictions match observations that have not been used to estimate parameters of the models. This is an important aspect of validation that is lacking in many modeling contexts, so the opportunity presented by the additional data collection should be taken advantage of.

Second, the additional data can be used to update the estimates of the model parameters and, one would expect, allow better prediction through more accurate parameter estimation. A continuing effort to incorporate such re-estimation might be considered a valuable investment and would certainly optimally integrate new data with previously collected information.

To incorporate the new data, either for use as a test set or for updating model estimation, those new data points need to be entered into the database that was created in an earlier task and that was extended through the early part of 2004 for this analysis. On-going updates of the database, with the option to revisit the analyses described here, are needed to maintain the best possible basis for making decisions about tissue collection in closed boxes, as discussed here, and for addressing larger issues of surveillance schemes that can be defined that meet requirements of the EU, provide adequate protection of the health and safety of the consumers of shellfish, and maximize the use of the scallop fisheries along the coast of Scotland.

CONCLUSION

Logistic regression analyses have been successfully applied to the DA concentration data in both gonad and whole tissues of king scallops. The modeling steps have identified functions of DAYS and covariables that appear to satisfactorily match the patterns in the database of observed concentrations.

It is worth reiterating that the modeling approach that has been adopted recognizes (in fact, is predicted upon the fact) that there is variability in the data. Not all scallops (even ones co-located in the same box) have the same DA concentration; changes in DA concentrations are not in lock-step; sampling a small proportion of scallops in a box yields values that may differ from the true box-specific mean. That is why a probabilistic approach is appropriate: the models are predicting the likelihood of low-concentration samples, and that likelihood is a result of the variability. If, all the variability could be accounted for, then we could get perfect predictions. But in all real-life systems, including this scallop sampling system set up by FSAS, that level of

prediction can never be attained. What we can hope for is that the sources of variability are identified and appropriately included in our model. Rather than abandoning hope in the face of variability, the perspective of our modeling approach is that such variability can be accounted for and predictions (albeit in terms of probabilities) can be provided. It is in this sense that the models presented here are considered successful: the predicted probabilities satisfactorily match the observed probabilities (obtained as a consequence of the various sources of variability) of obtaining low-concentration samples.

Those models can be used, as discussed above, to predict the likelihood of obtaining a low-concentration sample at various times after closure of a box; from such predictions, sampling plans that take account of the probability of successful reopening can be formulated. As discussed previously, the spreadsheet tool that has been provided with this report yields such predictions on a closure-specific basis. That is, the predictions depend on the factors associated with any given closure: the area in which it occurs, the month in which it occurs, and the DA concentration causing closure. And, for implementation of the results of the logistic regression analyses, such closure-specific factors should be considered on a case-by-case basis.

Nevertheless, some sense of the overall model predictions with respect to the timing of resampling efforts can be identified (see the figures for specific instances). Although there is a chance that low-concentration samples will be obtained at any time after closure, for the most part, the predictions are such that sampling early after closure will result in a small chance of successful reopening. That is a consequence of the low predicted probability of obtaining a single low-concentration sample and the fact that four consecutive low-concentration samples are needed for reopening. In fact, the 0.84 probability of a single low-concentration sample (yielding, as discussed in the examples given above, the moderate "50:50" chance of reopening a box closed due to a gonad concentration of 20 μg/g or more) would not be attained in many areas or months of closure until at least 60 days after closure, and often that level is not predicted to occur until much later than 60 days after closure.

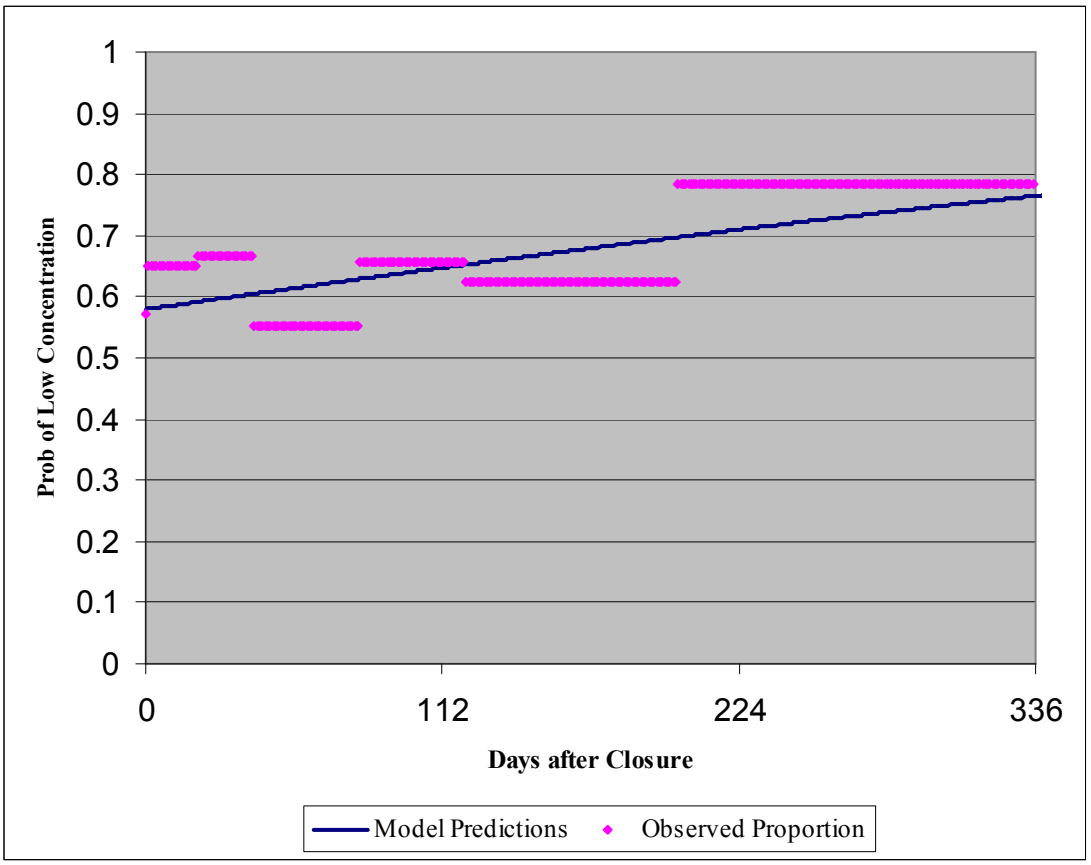
From the stand-point of allocating scarce resources, sampling a closed box when the likelihood of successful reopening is low may be perceived as a waste of effort. The most likely result would be no change in status. If resources are taken away from monitoring other boxes (either closed boxes that have a better chance of successful reopening, or open boxes that may pose a risk to public health if high DA concentrations are missed because of lack of sampling) then premature resampling of a closed box with a low probability of successful reopening may in fact be detrimental both to the scallop fishery and to protection of public health.

From the standpoint of the analyses completed here or others that might follow, sampling of closed boxes at all times has been beneficial. That is, that the patterns of DA concentration changes at all times since closure are better represented and the modeling can benefit from having observations to constrain the model fitting. Indeed, from a modeling perspective, continued (if lessened) sampling at short times after closure would be useful. But, as just argued, from the perspective of developing sampling plans that are

protective of public health as well as of the viability of the scallop fishery, such sampling does not appear to be optimal.

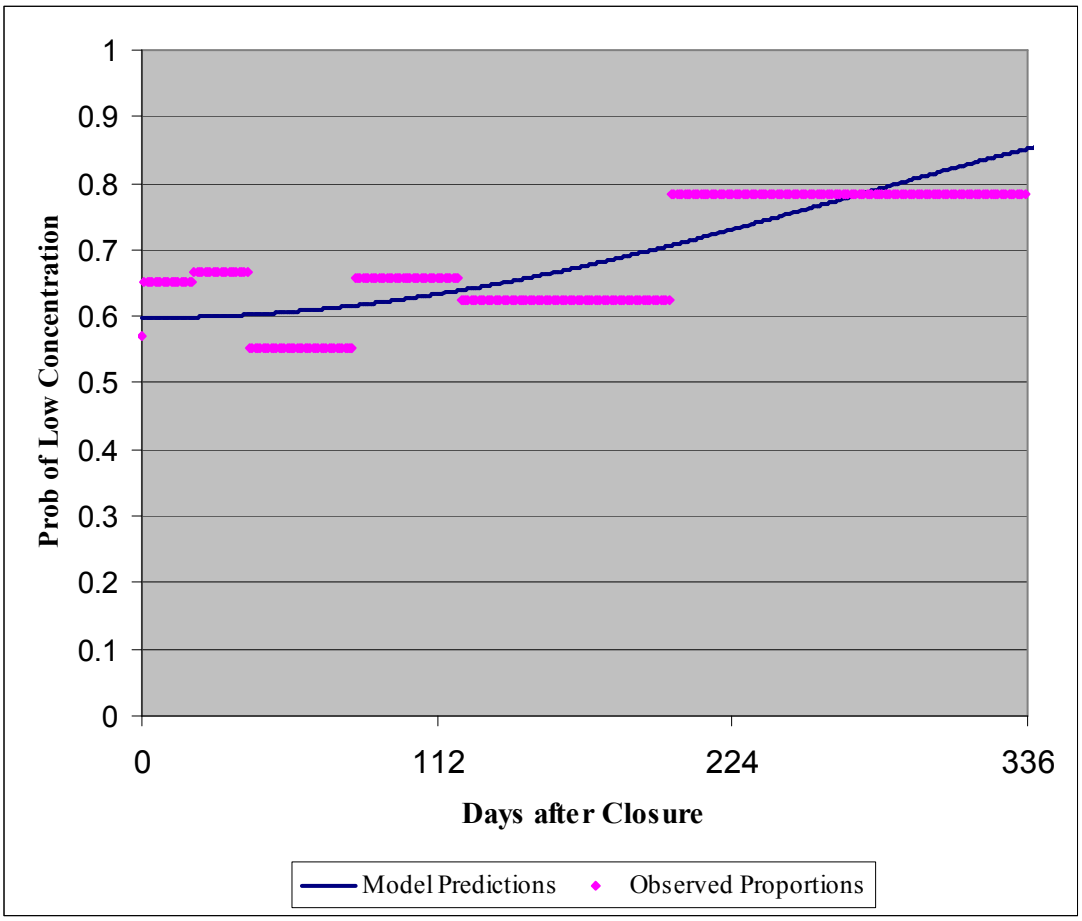
Again, each individual closure should be evaluated on a case-by-case basis using the spreadsheet tools provided with this document. But to the extent that sampling at short times after closure is not predicted to yield a high probability of a successful reopening, sampling efforts may be more efficiently directed. The analyses presented here, and the predictive tools that have been prepared as a result, can therefore form a key component, along with economic and health-protection considerations, of any sampling plan that FSAS intends to formulate.

Figure 1: Observed and Predicted Proportion of Low Concentration Gonad Samples; DAYS[1] Base Model



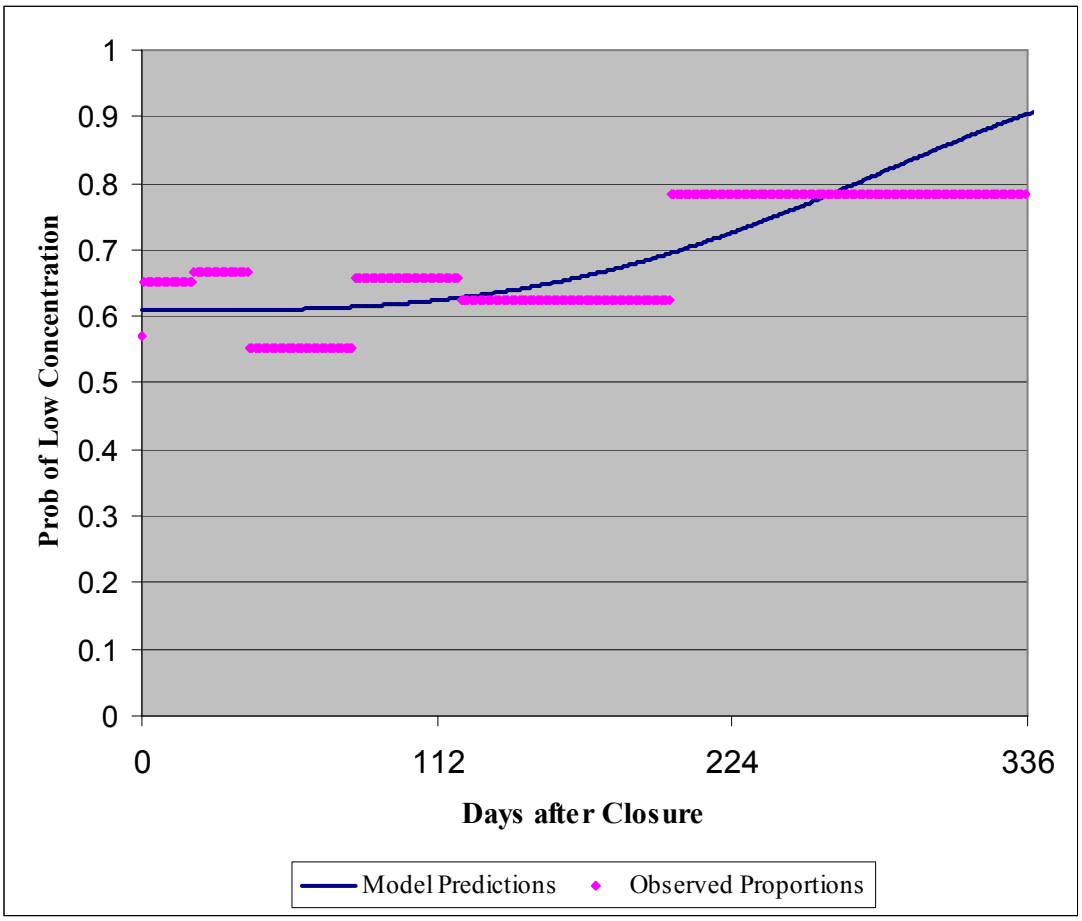
Notes: Observed proportions were derived for each of seven DAYS ranges (0, 1-19, 20-40, 41-80, 81-120, 120-200, 201-335) as the number of samples with low DA concentration (less than 20 µg/g) in that range divided by the total number of samples in that range.

Figure 2: Observed and Predicted Proportion of Low Concentration Gonad Samples; $DAYS^2$ Base Model



Notes: Observed proportions were derived for each of seven DAYS ranges (0, 1-19, 20-40, 41-80, 81-120, 120-200, 201-335) as the number of samples with low DA concentration (less than 20 µg/g) in that range divided by the total number of samples in that range.

Figure 3: Observed and Predicted Proportion of Low Concentration Gonad Samples; $DAYS^3$ Base Model



Notes: Observed proportions were derived for each of seven DAYS ranges (0, 1-19, 20-40, 41-80, 81-120, 120-200, 201-335) as the number of samples with low DA concentration (less than 20 µg/g) in that range divided by the total number of samples in that range.

Figure 4: Model Predictions Compared to Observed Values and Proportions for Area H, Month of Closure April; Best Model Using DAYS[1] for Gonad Samples

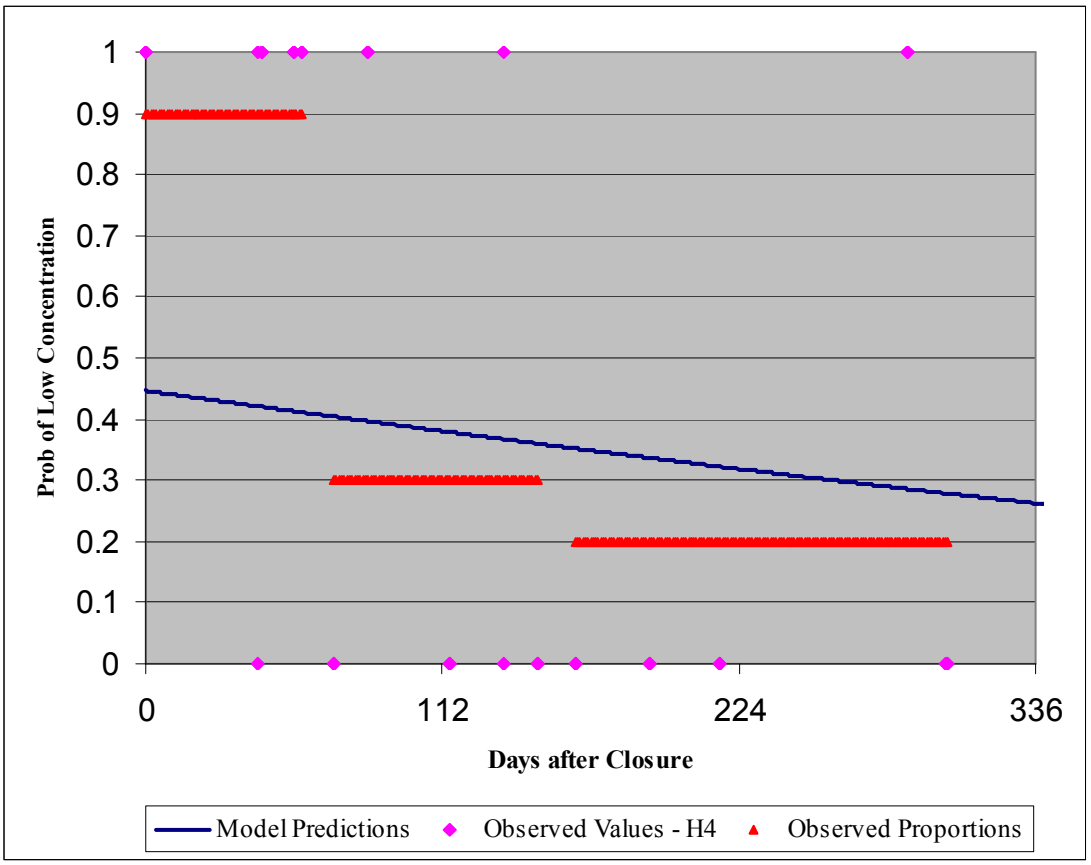


Figure 5: Model Predictions Compared to Observed Values and Proportions for Area H, Month of Closure April; Best Model Using $DAYS^2$ for Gonad Samples

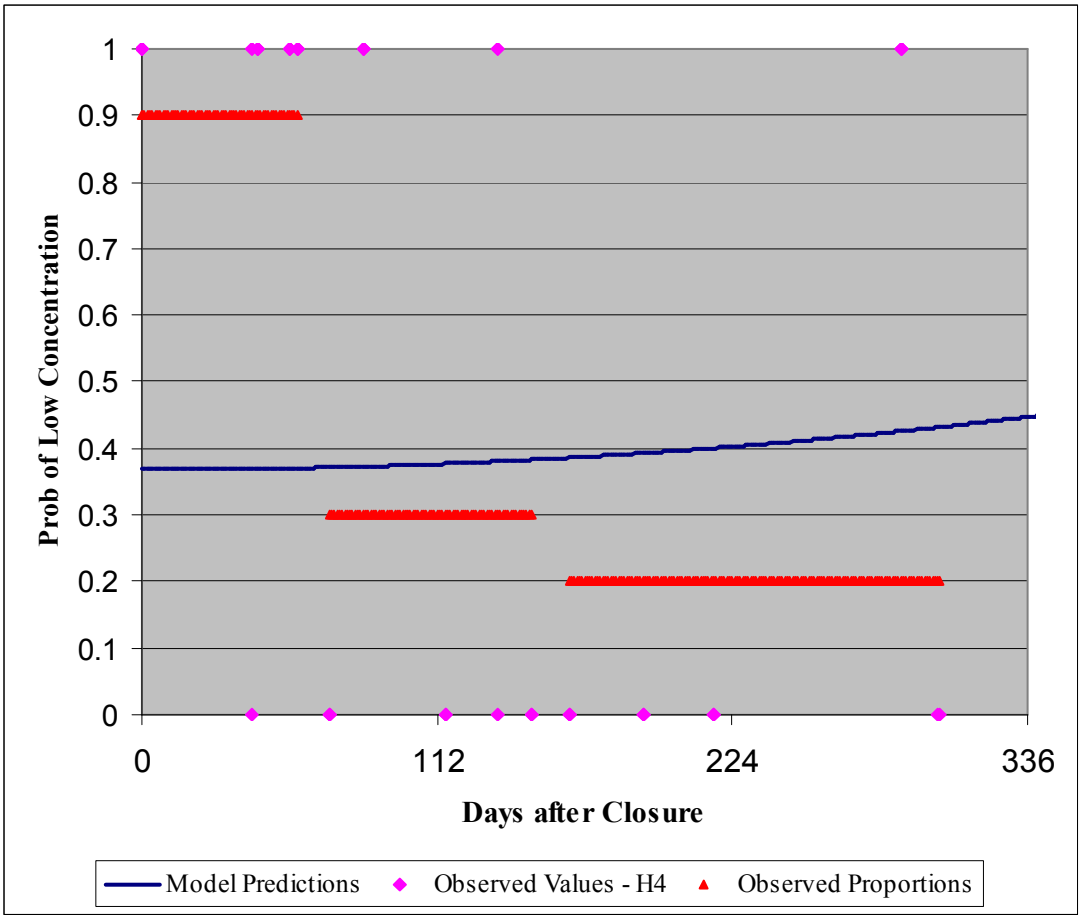


Figure 6: Model Predictions Compared to Observed Values and Proportions for Area H, Month of Closure April; Best Model Using $DAYS^3$ for Gonad Samples

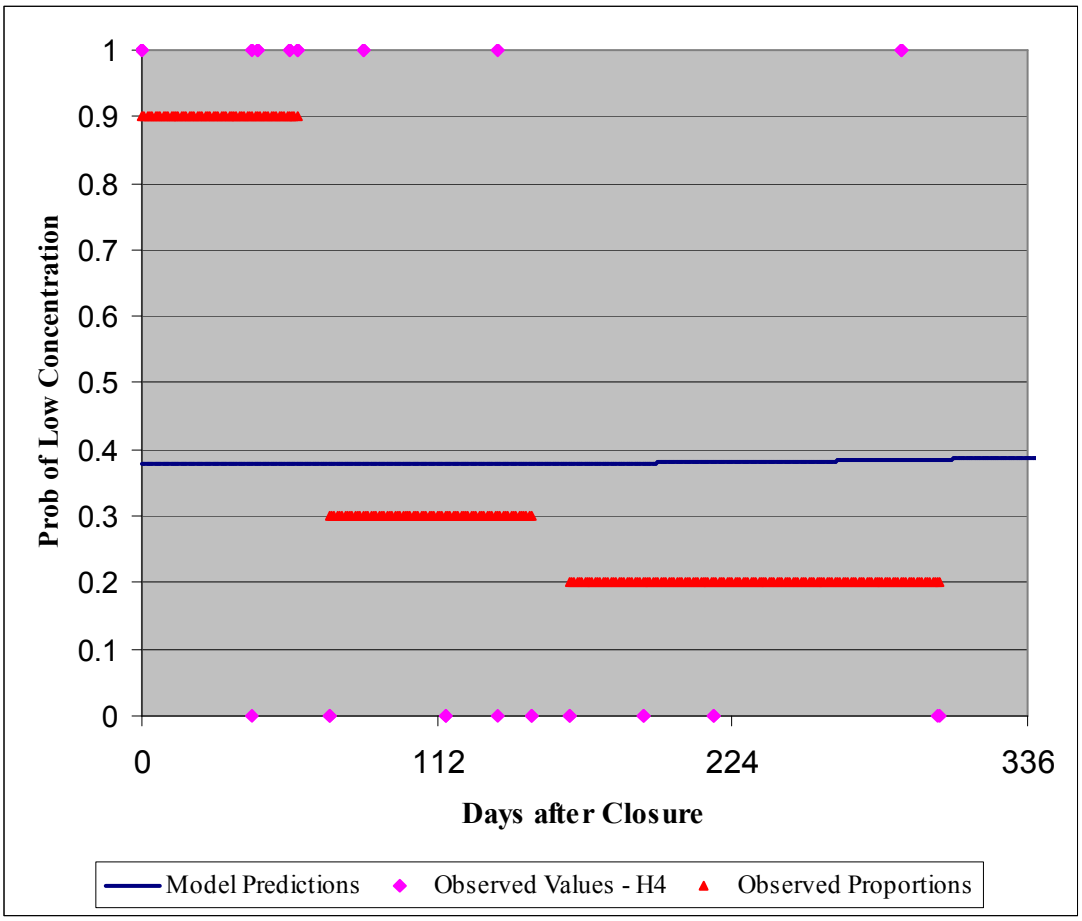


Figure 7: Model Predictions Compared to Observed Values and Proportions for Area O, Month of Closure July; Best Model Using DAYS[1] for Gonad Samples

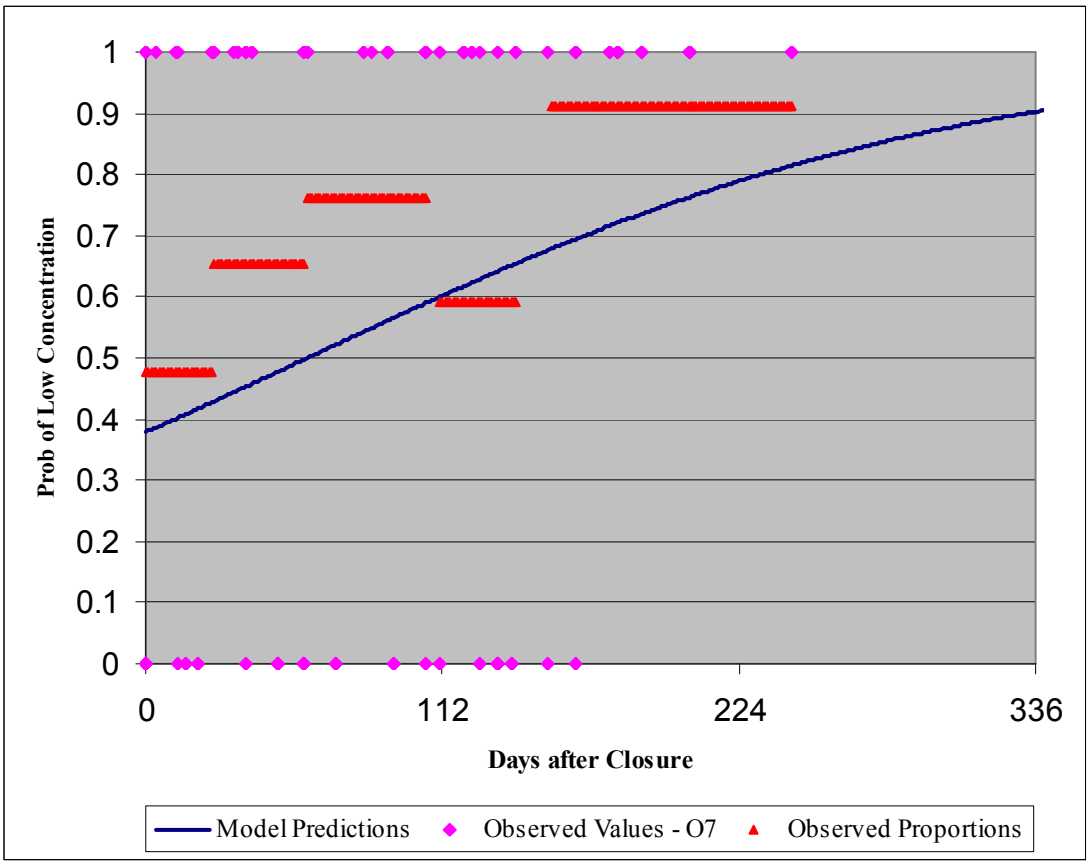


Figure 8: Model Predictions Compared to Observed Values and Proportions for Area O, Month of Closure July; Best Model Using $DAYS^2$ for Gonad Samples

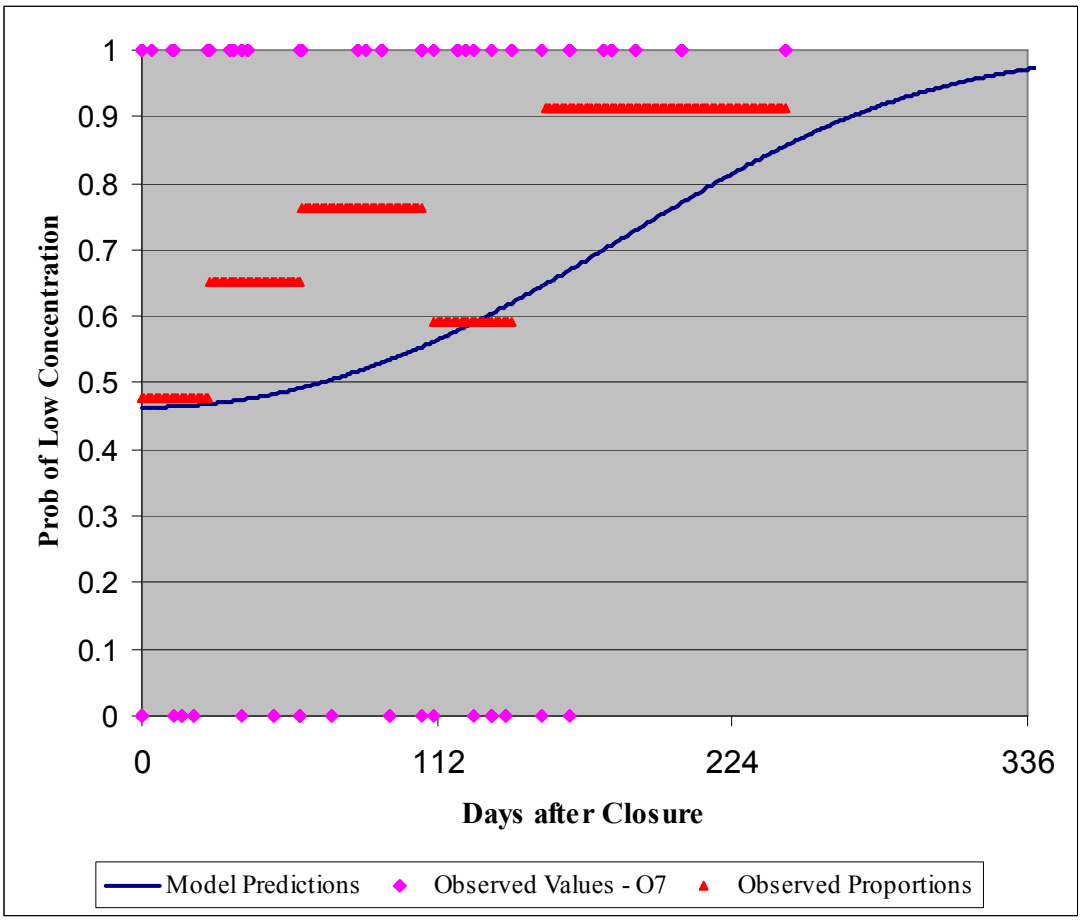


Figure 9: Model Predictions Compared to Observed Values and Proportions for Area O, Month of Closure July; Best Model Using $DAYS^3$ for Gonad Samples

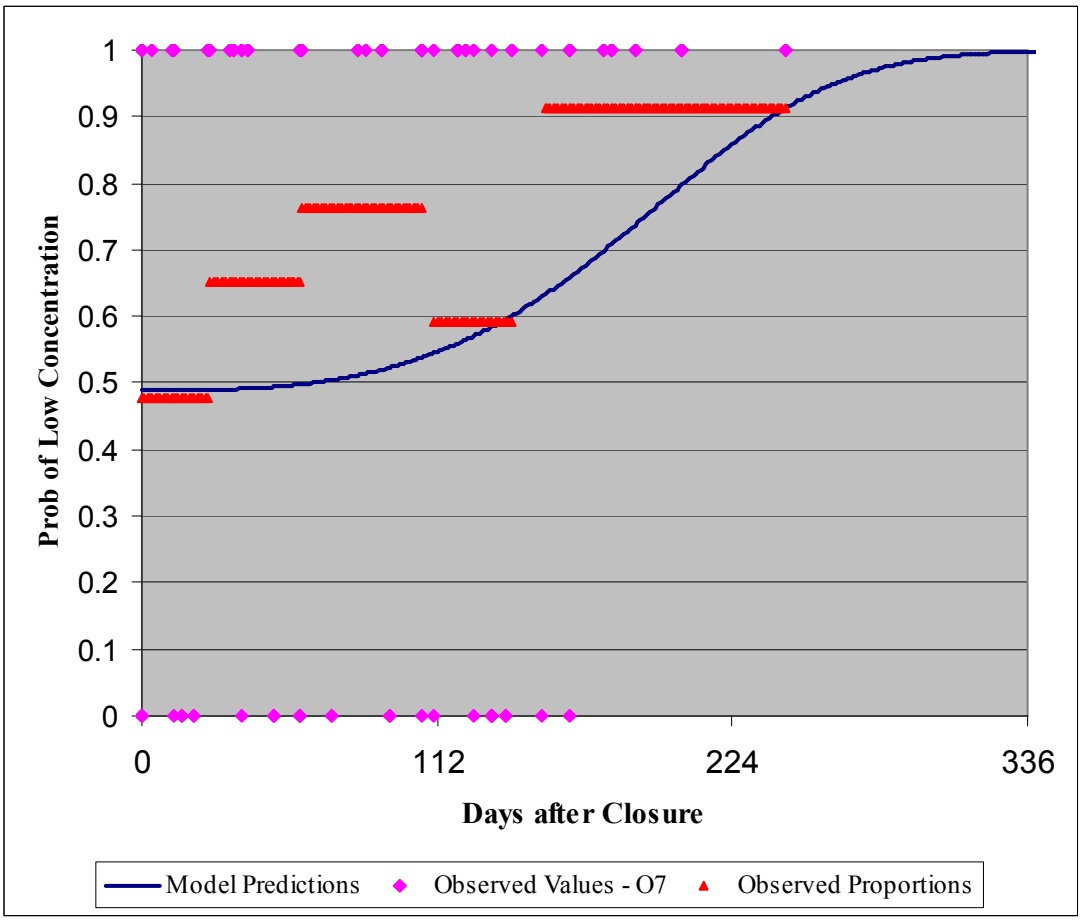


Figure 10: Model Predictions Compared to Observed Values and Proportions for Area SM, Month of Closure October; Best Model Using DAYS[1] for Gonad Samples

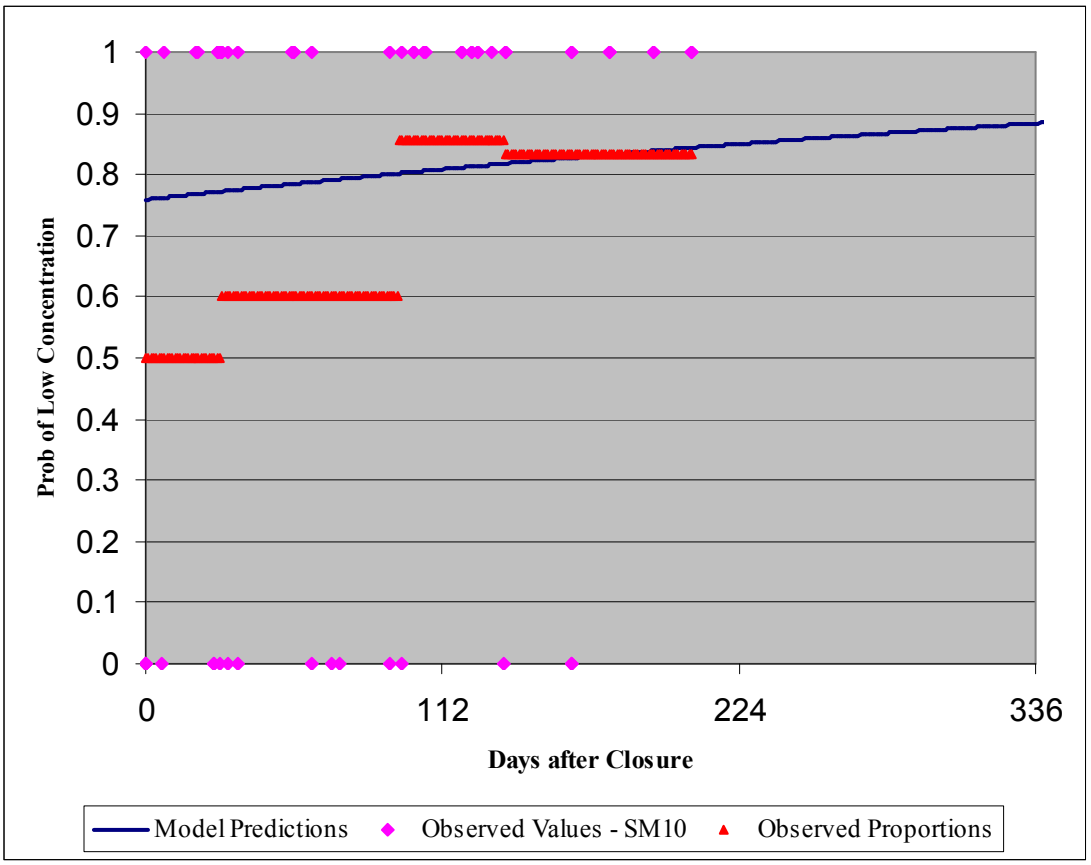


Figure 11: Model Predictions Compared to Observed Values and Proportions for Area SM, Month of Closure October; Best Model Using $DAYS^2$ for Gonad Samples

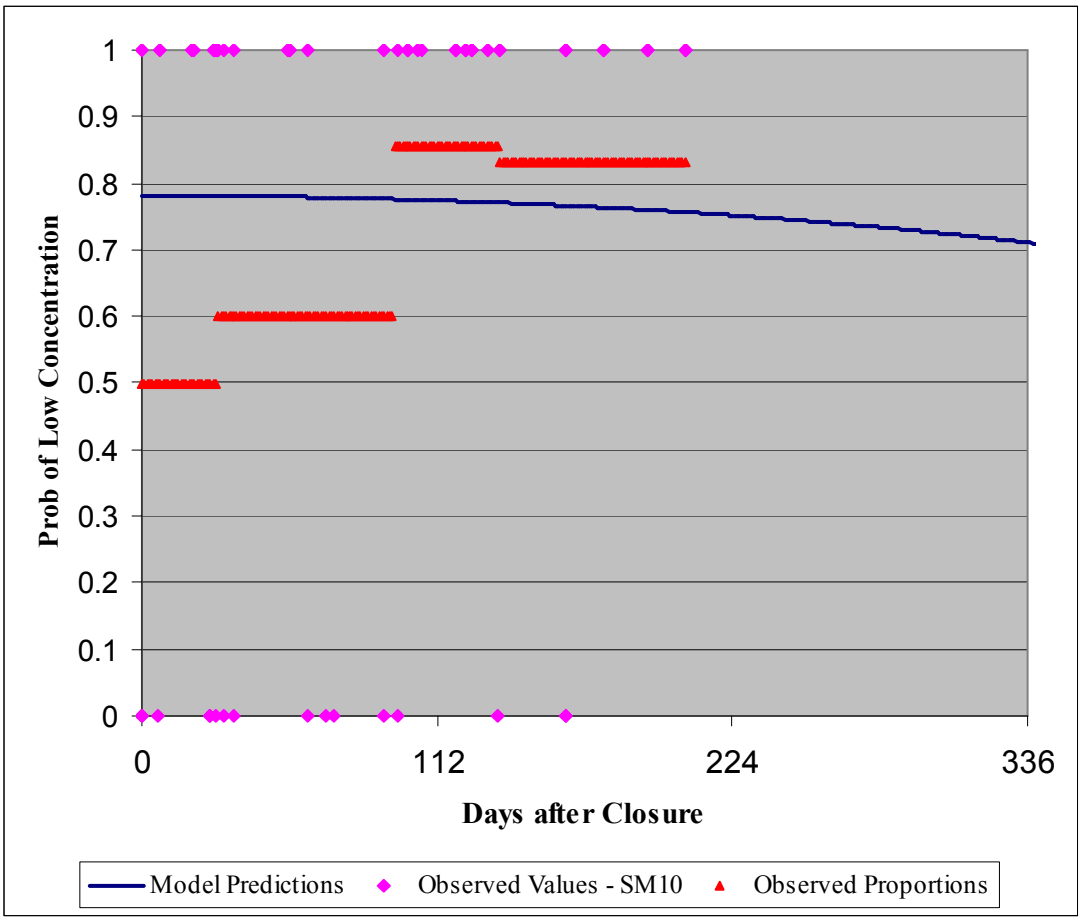


Figure 12: Model Predictions Compared to Observed Values and Proportions for Area SM, Month of Closure October; Best Model Using $DAYS^3$ for Gonad Samples

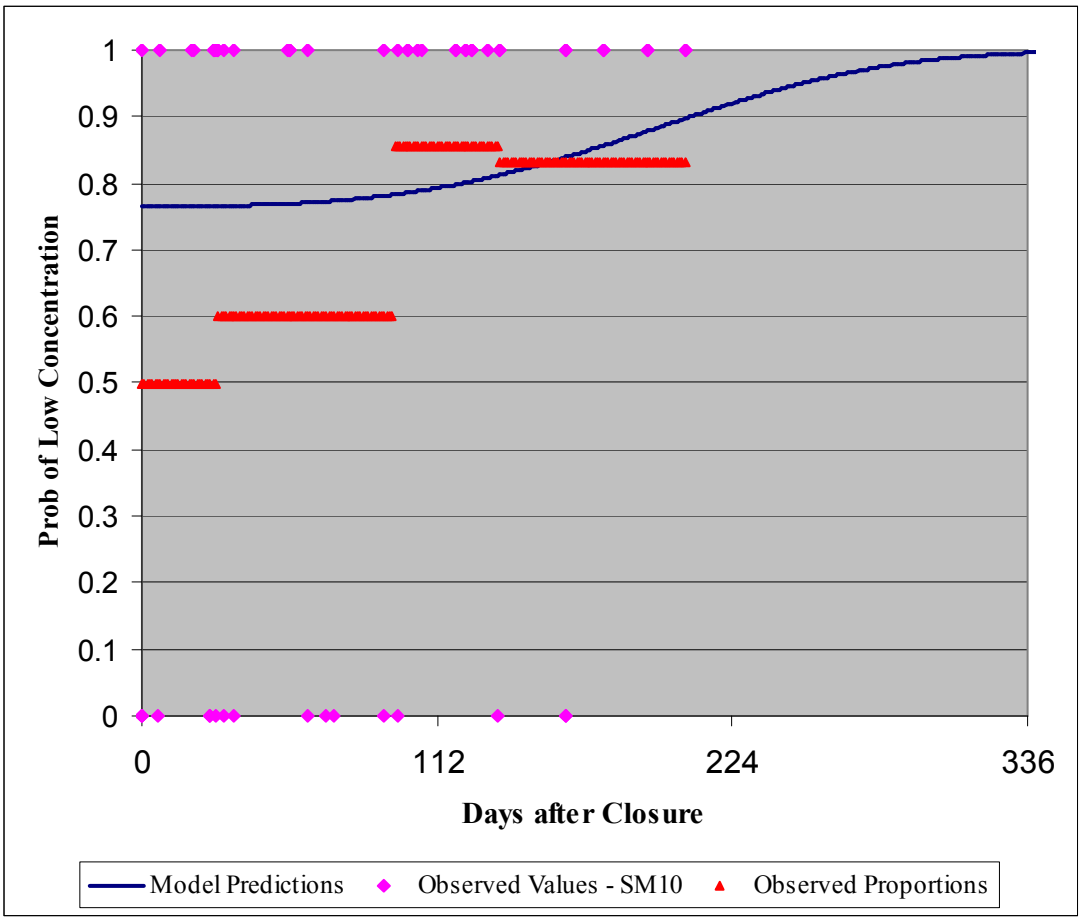


Figure 13: Model Predictions Compared to Observed Values and Proportions for Area SM, Month of Closure October; Model Using DAYS^3 and COAST for Gonad Samples

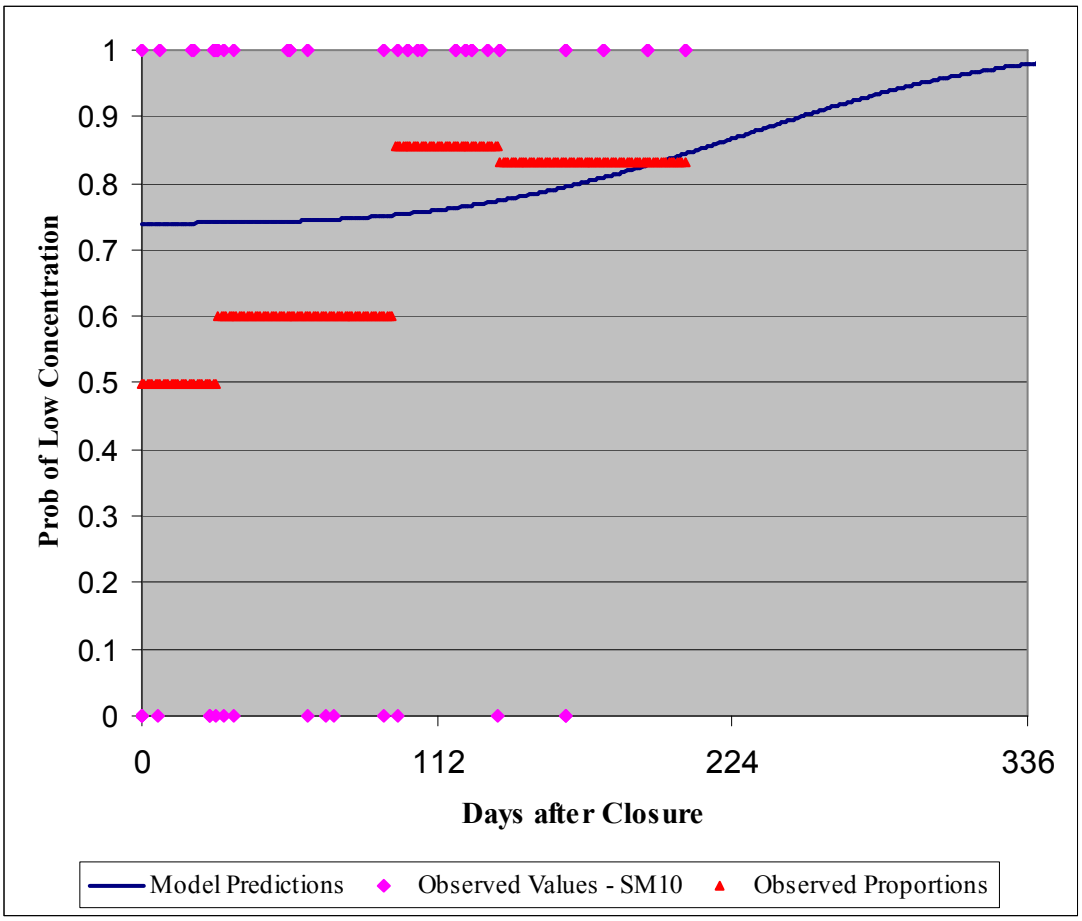


Figure 14: Observed and Predicted Proportion of Low Concentration Whole-Tissue Samples; $DAYS^2$ Base Model

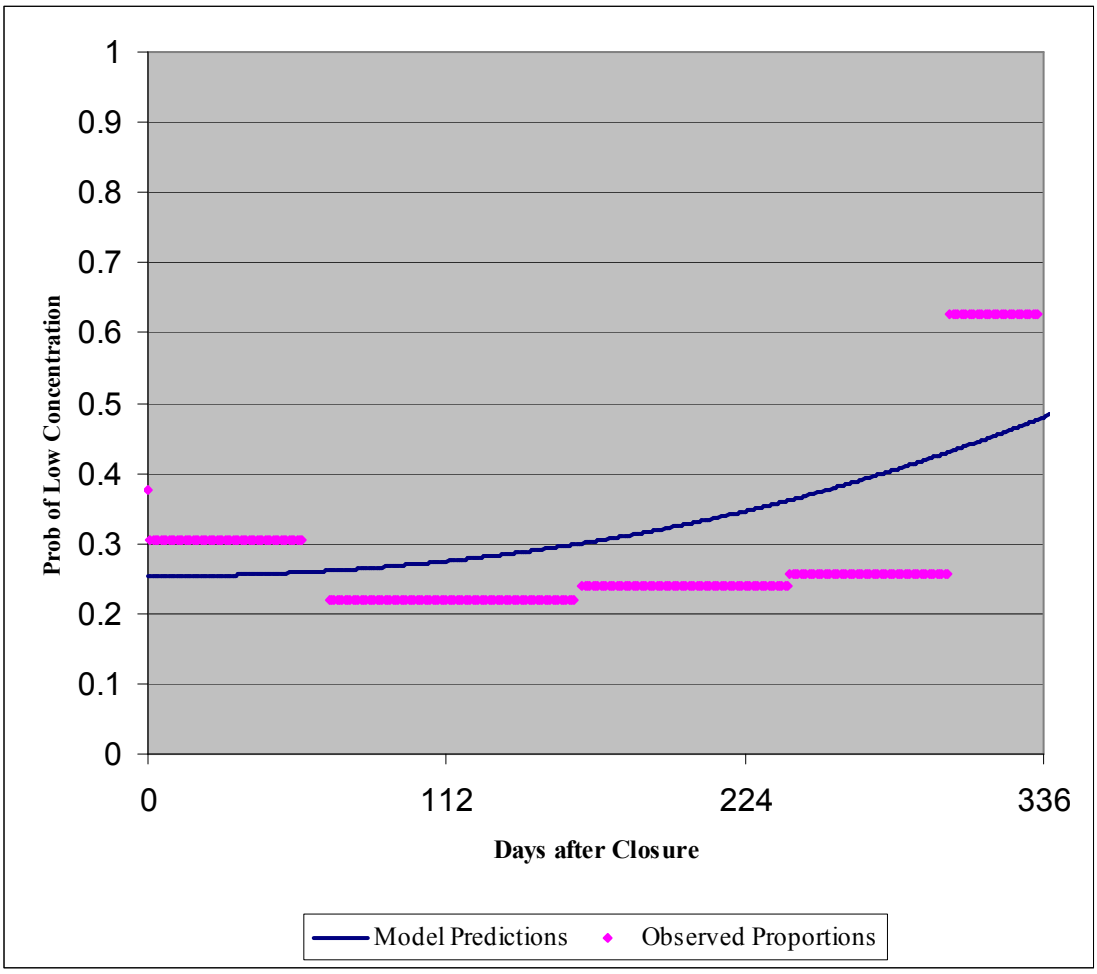


Figure 15: Observed and Predicted Proportion of Low Concentration Whole-Tissue Samples; $DAYS^3$ Base Model

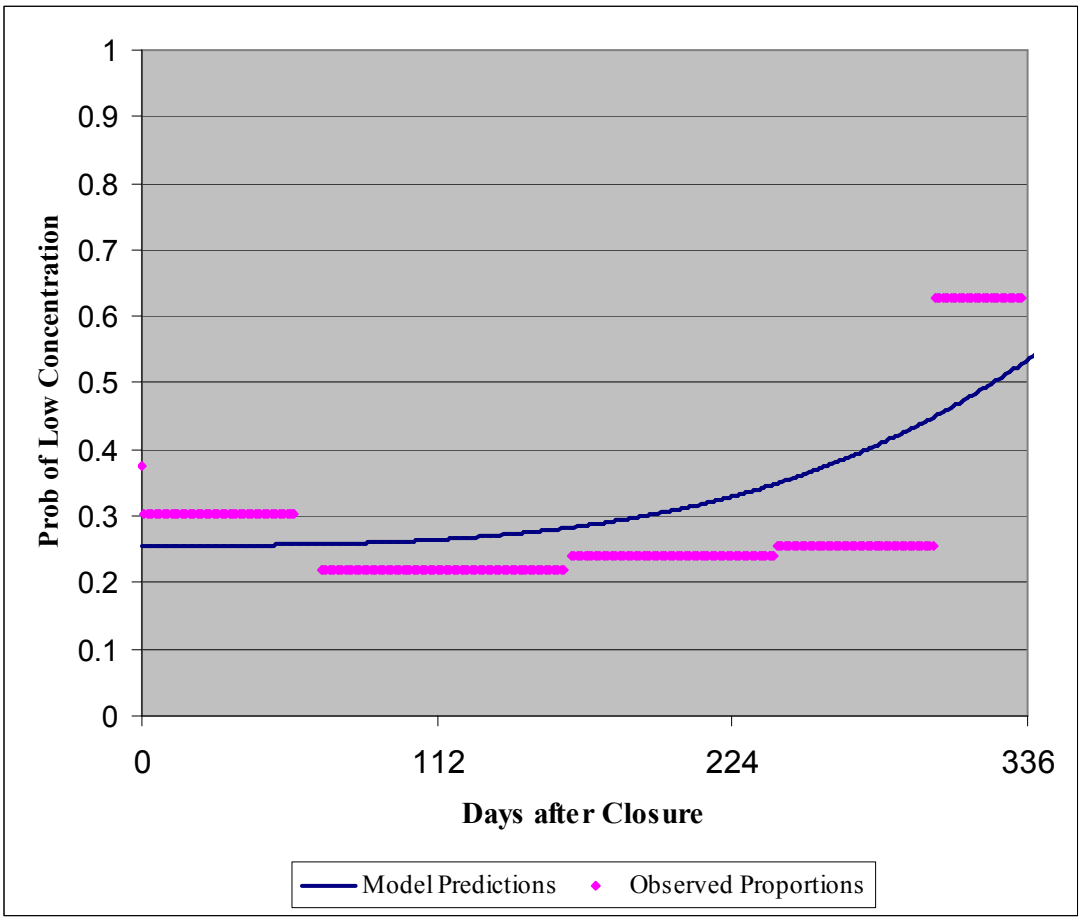


Figure 16: Observed and Predicted Proportion of Low Concentration Whole-Tissue Samples; DAYS^6 Base Model

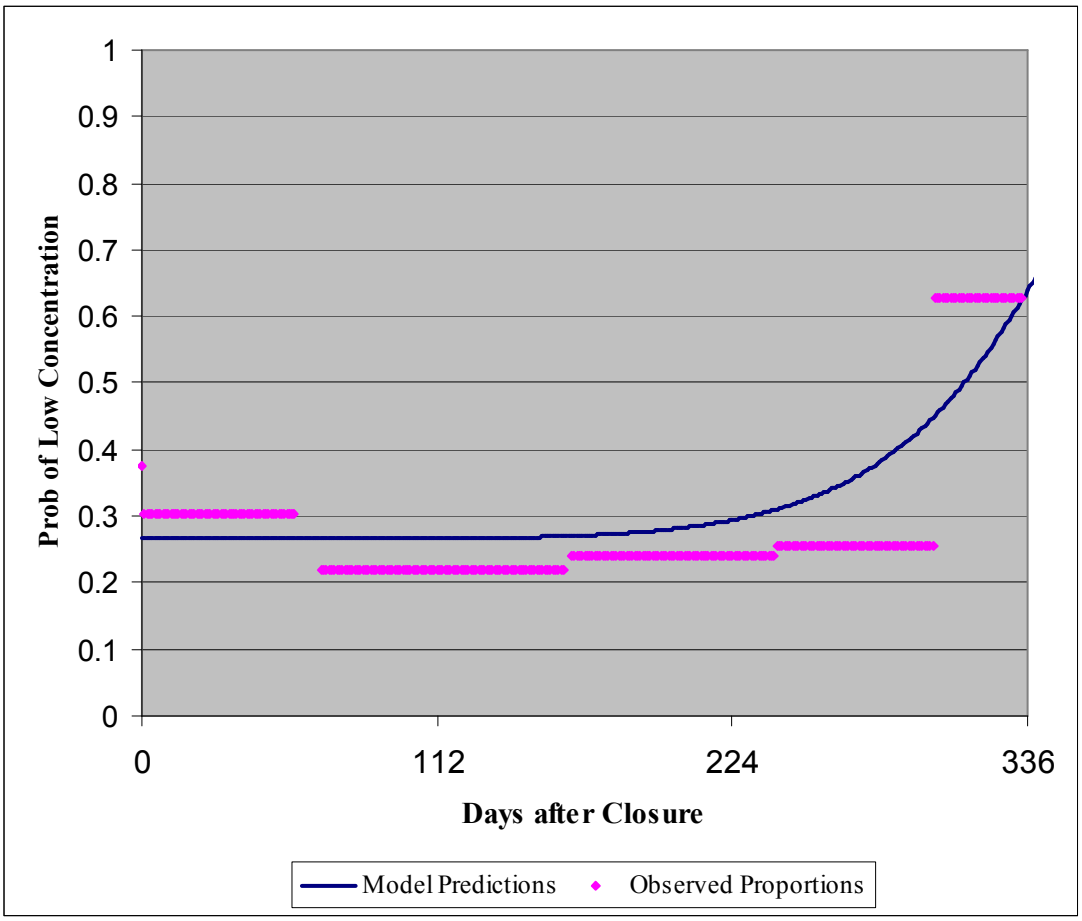


Figure 17: Model Predictions for a Hypothetical Closure Occurring in Area SM in July, Due to a Gonad DA Level of 30 μg/g

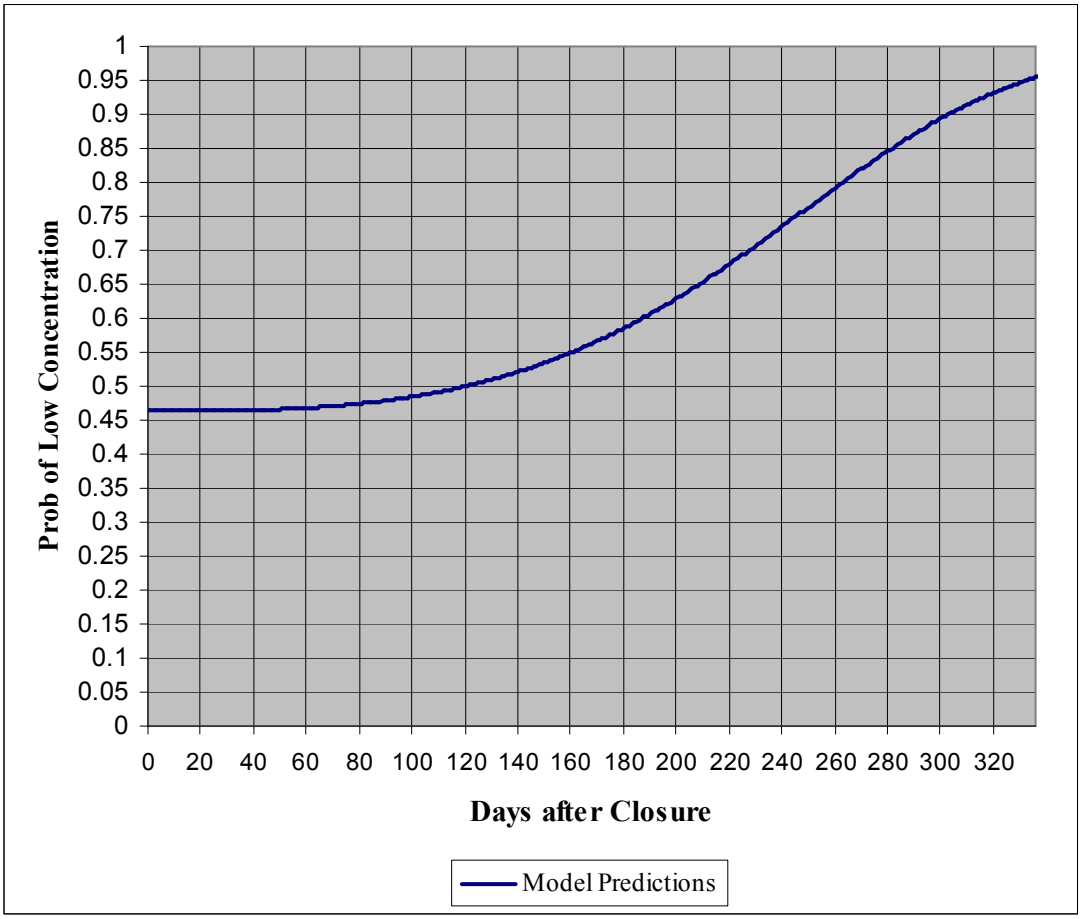


Figure 18: Model Predictions for a Hypothetical Closure Occurring in Area NM in June, Due to a Gonad DA Level of 22 μg/g