Independent Scientific Advisory Board
for the Northwest Power and Conservation Council,
Columbia River Basin Indian Tribes,
and National Marine Fisheries Service
851 SW 6th Avenue, Suite 1100
Portland, Oregon 97204
ISAB@nwcouncil.org

Memorandum (ISAB 2006-6) **August 8, 2006**

To: Bob Lohn, Regional Administrator, and Usha Varanasi, Science Director, National Marine Fisheries Service, Northwest Region;
Olney Patt, Jr., Executive Director, Columbia River Inter-Tribal Fish Commission;
Tom Karier, Chair, Northwest Power and Conservation Council

From: Nancy Huntly, ISAB Chair

Subject: COMPASS Model Development **--** ISAB reply to NOAA Fisheries' request for a reaction to the COMPASS team response

At the request of NOAA Fisheries (NOAA), the ISAB replies to the response of the COMPASS team (detailed in Dr. Rich Zabel's PowerPoint presentation to the ISAB on June 30, 2006) to the ISAB's initial COMPASS Model review (ISAB 2006-2[1]). The points at issue for both the ISAB's report and for NOAA's response were largely confined to statistical usage, over which there remained some differences of opinion that needed further discussion and resolution. Several model team members also responded to Dr. Zabel (letter of May 22, 2006 from the CRITFC, IDFG, ODFW, and USFWS model team members) with our review in hand, and raised many of the same concerns, as well as some additional concerns. CRITFC forwarded this letter to the ISAB and requested that we consider the other team members' comments in developing this memo. In sum, the ISAB is encouraged by the efforts of the COMPASS team to date, and the ISAB hopes its comments will further the team's discussions and development of the model.

Regarding the ISAB's initial review, Zabel felt that there were four major points at issue, and one minor point:

(1) *Data weighting*: The question was whether all the data should be used or whether poorly estimated points should be sacrificed. The issue of a proper weighting scheme was also discussed, if all the data were to be retained. The response to this issue included several points.

(a) Zabel and several team members felt that weighting tended to focus most of the attention on a few data points, thus allotting almost no weight to other data points.

[1] www.nwcouncil.org/library/isab/isab2006-2.htm

Despite these concerns, they agreed to weight, rather than tossing out the poorly estimated points altogether. They also pointed out that their *rejected* data points accounted for almost no collective weight in the weighted regression models.

The other model team members (May 22nd ltr.) agreed with the ISAB that inverse variance weighting was a much better alternative, and voiced objections to data elimination. They also felt that data elimination would prevent useful comparisons of different managerial or natural flow conditions, to the detriment of the overall effort.

Some data are more reliable (smaller variances) than are others, sometimes substantially so, and that the best statistical practice is to weight accordingly. We fully anticipated that the *rejected* data points would receive almost no weight in the weighted regressions, since their variances were large enough to make their inclusion objectionable, had they been employed without appropriate weighting. Weighting that allows for the differences in variance allows one to use all the legitimate data *for what they are worth*, without arbitrarily having to pick which data to use. We agree with their current decision to weight, as do the other team members.

(b) ISAB suggested that the COMPASS model should employ binomial weights. Zabel and all model team members pointed out that Cormack-Jolly-Seber (henceforth CJS) estimates are more complex than is usual in standard binomial treatments and that binomial variance is generally too small to account for all the errors of estimation in such situations.

The point is well taken, but weighting is still entirely appropriate and necessary. Instead of using a binomial form of the variance, however, the revised COMPASS model will now use an appropriate form of the CJS variance (see (2) below) for weighting. We concur with that choice for the next round, though other members of the model team would like to see a more analytical solution, presumably starting from CJS methodology. The ISAB finds that suggestion attractive for the long haul.

(2) *Model form*: The question was whether survival data should be truncated at (or adjusted to be less than) 1.0, or whether they should not be truncated, in view of the CJS origins of the estimates themselves.

(a) The real difficulty here is that the CJS estimates can exceed ‘1’ from real data, and that standard survival models are confined to the interval $p \in [0, 1]$. Survival cannot be greater than ‘1’ or less than ‘0.’ Zabel points out that any survival model that insists on *containment* within the interval is inherently biased. They explore several options that reduce the bias, and after doing that, choose the logistic option. Other members of the team (May 22nd ltr.) point out that the results presented in Table 2 make it look worse than it may be, because there are biases in back-transformation that are not accounted for in this table.

This is tough problem that accompanies the CJS estimation procedure. A careful examination of Figure 6 (page 16) of the response by Zabel and several team members provides some insight into the nature of the problem. Most of the CJS estimates of $p \cong 1$ exceed '1' itself, as explained by Zabel and as expected from experience. What is particularly interesting is that the CJS estimates are actually curvilinear in terms of increasing time, but opening *upward*, rather than *downward*, as would be expected from the negative exponential model favored by several members of the team. Both estimation models (logistic in form) track upwardly opening curvilinearity of the CJS estimates themselves, and neither is doing a particularly good job of it, constrained by the imposition of containment of the estimates within $p \in [0, 1]$.

The issue is whether they are attempting to predict CJS estimates that sometimes exceed '1' or whether they are attempting to predict survival, which is bounded, $p \in [0, 1]$. The ISAB's sense is that they should be trying to do the latter. For that latter purpose, we prefer models for which predicted estimates of survival probability are contained within $p \in [0, 1]$, even if the CJS estimates > 1. It is also worth noting that while avoiding bias is good, small estimation variances are also good. Sometimes, maximum likelihood estimates are severely biased, but they have minimum variance. There are times when it is more appropriate to be concerned with Mean Squared Error, MSE = Variance + $(\text{Bias})^2$. We confess ignorance on the MSE details of this particular situation, but minimizing Bias may not be enough. The other members of the team (May 22nd ltr.) state that it would be good to explore the stochastic aspects further and suggest some sort of AIC, AICC, BIC or DIC model assessment evaluation of the various strategies that could be employed.

(b) Zabel and several members of the team make the point that it makes a difference whether one uses the variance associated with the observed probability of survival or that associated with the model-predicted probability of survival, and that the latter is the better choice. We concur, and that is exactly what we intended to imply in our initial critique. On rereading our earlier recommendation (ISAB-2006-2), it becomes obvious that we failed to make ourselves clear. The point is taken. The other team members (May 22nd ltr.) point out that there is more than one model transformation of survival probability that is *contained* within $p \in [0, 1]$. They refrain from endorsing the ISAB's logistic suggestion, but they agree that *containment* has much to offer.

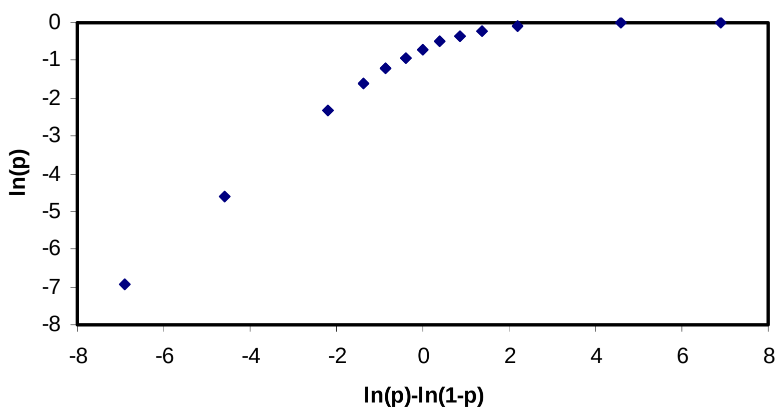
(1) In the end, Zabel and several team members opt for the negative exponential model, based on several considerations: 1) the formal survival model is negative exponential in form, not logistic in form; 2) the logistic models are biased; 3) the two forms of model are linear in each other, so it basically makes no difference. Neither the ISAB nor other team members (May 22nd ltr.) are firmly wedded to the logistic, but while an interim choice of the negative exponential may suffice, the ISAB would prefer a *contained* estimation approach. The other members of

the model team (May 22nd ltr.) can live with the negative exponential for now, but would like to see further exploration of *containment*.

(2) A model is just a model; it is useful to the extent that it helps us think about the problem; it should not be confused with reality. The logistic is also traditional and is statistically *contained*. Our recommendation was based on the urge to avoid predicting $p > 1$, while accommodating *both* positive and negative regression coefficients (β) for time, distance, or any other predictor variables, some of which were an outgrowth of CJS estimates that exceed '1.'

(3) Suffice it that one must take observed data and translate them into model parameters, with which to drive the demographic models needed by COMPASS. One should continue to improve the estimation strategy as one goes along, as normal operating strategy.

(4) $-\ln(p)$ and $-[\ \ln(p - \ln(1 - p)\ ]$ are not entirely linear over the full range of p-values, as the following graph shows:



The issue here is how large a range of p-values one is likely to encounter in practice. If real survival data are restricted to the range [$0.7 < p < 1.0$], the *linearity* claim is probably an adequate 1st approximation, but *not* if the full range of p-values from '0' to '1' is considered.

We gather that most single-reach p-values meet the stipulation, which provides the opportunity to use either the $\ln(p)$ treatment that Zabel and several members of the team prefer the alternative $[\ \ln(p) - \ln(1 - p)\ ]$ treatment suggested by the ISAB, whichever suits the estimation purpose better. The collective COMPASS team should feel entitled to whatever treatment works best for their various purposes, but empiric success should be the criterion, not adherence to any survival or statistical model.

(c) One other point we should make is that Zabel and several team members prefer the use of relative variance CJS weights, given that they prefer the $\ln(p)$ modeling frame, and in view of their choice, that seems a reasonable interim choice. We would also agree with other members of the team that further exploration of an analytical solution is in order, presumably beginning from standard CJS estimation theory.

(d) The issue of how to contain the estimates deserves more attention, especially if the negative exponential treatment is continued. Pending a choice of a *contained* estimation strategy, both the ISAB and several model team members (May 22nd ltr.) prefer that the parameter estimates from the statistical analysis be calculated first, even if they exceed $p > 1$, and that the corresponding demographic parameters for the demographic projection model should be constrained so that $p \leq 1$. Zabel and several on the team seem to prefer disallowing models with $p > 1$, though other team members question that. The demographic projection model is the product; the statistical estimation of p-values is a means to an end.

(3) *Model complexity*: The question was whether the survival model was too complex, as claimed by the ISAB, and whether the model was thus "overfit," or whether it was not. A number of members of the model team leveled even more severe criticism on this matter (see below). Zabel and some on the team have simplified the model, in response to this criticism, but have argued strongly for the retention of a few more elaborate models for particular purposes. This exercise in *thinking it through* is precisely the sort of thing that the ISAB wished to encourage. We do offer some further thoughts:

(a) In the equations at the bottoms of Pages 18 and 19 of the COMPASS response to our call for *intercept* terms, there are still no intercept terms. Rich Zabel explained that a project-specific intercept was in order, best viewed as a function of the distance from one project to the next, hence the $\alpha_0 d$ term on page 18 and the $\alpha_0 d$ and $\beta_0 d$ terms on page 19. However, a large part of the team (May 22 ltr.) takes issue with the use of fish travel times or distance as predictors, preferring water travel times instead, and particularly dislikes time/distance-standardized treatment, since the elucidation of such effects could itself be important. Once standardized, the opportunity to explore such effects is lost. The choice of which terms to include should not be rigidly decided in advance. A modeling platform that is flexible allows the data to speak, rather than allowing *a priori* decisions to channel the effort too tightly.

There is typically a survival cost to handling smolt, a *transplantation* cost, over and above the cost of moving downstream as far they have to go between projects. If *transplantation shock* reduced survival by 1%, then an intercept term of $\gamma = -0.01$ (multiplied by no other variables) should yield a null model of

$$\ln(p = 0.99) = \gamma = -0.01 ,$$

which is an appropriate starting point for further analysis, with additional terms being compared via AIC-type criteria, as outlined in both Zabel's and other team members responses. They will need to estimate γ, rather than simply invoking a fixed value.

(c) Unless one wants model predictions of $p > 1$, it would be good to constrain all of the resulting models to being contained within $p \in [0, 1]$. We commented on this issue in our earlier report, and the other team members (May 22nd ltr.) raise the issue as well. Zabel prefers to disallow models that yield $p > 1$, but the rationale for this is unclear. Both ISAB and other members of the team prefer allowing the statistical estimation to play out as it will, but constraining the resulting parameters in the demographic model to $p \leq 1$. A *contained* estimation procedure would avoid this necessity altogether, but such analysis may require additional exploration.

(d) Zabel and several team members express an aversion to using "expert opinion modeling" **. . .** "at all costs." Our response is that in the absence of compellingly contrary information, "expert opinion" is a reasonable guide to standard practice, though one should never follow it slavishly. A wise modeler follows the evidence (or data) wherever it (or they) may lead, but always *with one eye cocked* on what we think we already know (= expert opinion). The other team members (May 22nd ltr.) point out that the limited quality and applicability of the available data may obviate any but the coarsest of statistical estimation models. Common sense is the most reliable standard, which is basically what AIC-type procedures are attempting to codify.

(4) *Travel time distribution*: The question was whether travel time distributions would underestimate the mode, therefore placing too much probability in the tails. Zabel and several team members took issue with ISAB's assertion, on the basis of more recent information, presenting data that suggest the fit of observed and predicted travel times seems to be relatively close and not particularly biased. Several other team members (May 22nd ltr.) would like further clarification on this point and the contrast between the Zabel and Anderson (1997) and Zabel (2002) outcomes. It seems clear that this point is in need of some further discussion between all members of the model team.

(5) *Outlier data point*: The ISAB had called into question the claim that years of low flow were associated with poor survival, claiming that this conclusion was based on "one anomalous observation." Model team members have countered with data accumulated from earlier reports, citing also a 2003 ISAB review of evidence, interpreted as showing that there appears to be a threshold in flow rate, below which survival is compromised, but above which survival rate is relatively insensitive (see ISAB 2003-1[2]). The point is well taken; the ISAB withdraws its objection with the following caution. The ISAB is suspicious of "broken-stick models" that indicate a biological response exhibits a threshold. In this case the model indicates that below a threshold in flow rate survival is compromised, but above which survival rate is relatively insensitive may be an over simplification due to a lack of data. The ISAB recommends that the model be critically evaluated often as more data are obtained.

[2] www.nwcouncil.org/library/isab/isab2003-1.htm