Noakes response to Connors August 10, 2011

There were a number of significant flaws in Connors' analyses and conclusions (Technical Report 5B) and I touched on some of those in my report (Technical Report 5C). I apologize if my explanations were not clear and I provide some additional clarification below.

I'll start with the farm salmon production time series ('Farm') which Connors uses as a 'proxy for pathogen exposure'. It is clear from his report, his comments on my criticism of his report, and the manner in which the data are included in his model that Connors assumes 'pathogen exposure' is proportional to farm salmon production. It is, however, neither sufficient nor acceptable to use farm salmon production as a 'proxy for pathogen exposure' simply because as Connors notes 'it is the only source of information we have'. Connors spends considerable time qualifying and justifying the use of the various sockeye salmon time series and other variables in his model but employs farm salmon production in his model (as a proxy) without any critical assessment or review. The onus is on Connors to clearly demonstrate that this assumption (using farm salmon production as 'a proxy for pathogen exposure') is appropriate. Connors has to clearly demonstrate that farm salmon production (as a proxy for pathogen exposure) is consistent with the fish health data (evidence) available from the farms (detailed data are available since 2003) and that his assumed relationship (proportionality) is consistent over the entire time series. Based on the discussion below, I believe the evidence clearly shows that Connors assumption with respect to farm salmon production as a proxy for pathogen exposure is not reasonable and consequently his analyses and conclusions are without basis.

Let's start by considering the 4 'high risk' diseases identified by Kent (IHN, BKD, furunculosis, and vibrio) and sea lice. There have been documented outbreaks of IHN in farmed fish in the past (1992, 1995-97, and 2001-2003; http://www.agf.gov.bc.ca/ahc/fish_health/IHNV.htm) and no IHN has been detected in farmed salmon since 2003 (not surprising since Atlantic salmon are now routinely vaccinated for IHN). The test for IHN is quite accurate (on the order of 80 or 90%) so the probability of not detecting IHN in any of the hundreds of fish tested over the years (since 2003) if it was indeed present is extremely small. Also, without an active outbreak of disease (IHN), it's highly unlikely that there is any carrier state for IHN and the farmed fish are unlikely to be shedding IHN virus into the environment (K. Garver pers. comm. and expert opinion from other fish health scientists, Noakes 2011). Thus, the available evidence suggests that IHN disease outbreaks (and consequently the potential for exposure to IHN pathogens) have been infrequent and sporadic since the early 1990s and essentially absent since 2003. IHN pathogen exposure is therefore clearly not proportional to farm salmon production at least during the period since 1992. Thus, based on the available evidence farm salmon production clearly cannot be used as a proxy for 'IHN' pathogen exposure.

With respect to BKD, based on the available fish health data from farms as well as discussions with industry veterinarians and fish health scientists BKD is a problem for Pacific salmon (chinook and

coho) but it is not a significant issue for Atlantic salmon. The mix of Atlantic and Pacific salmon has varied over time but recently the percentage of Pacific salmon farmed has decreased steadily from about 30% in 2003 to less than 10% in 2010 (Figure 3, Korman 2011). The number of farms with BKD problems has also decreased in recent years with very few farms (as few as 1 farm in each of the last 3 years) within or close to the migration path for Fraser River sockeye in recent years (Noakes 2011). The location of the farms experiencing BKD disease outbreaks (or any other disease) is clearly very important as farm location directly influences potential exposure to pathogens. There are no farm specific data for BKD prior to 2003 (at least none available to the Project 5 researchers) but given BKD is endemic and widespread it is reasonable to assume that BKD infections were common in farmed Pacific salmon prior to 2003. Whether BKD infections were proportional to Pacific farm salmon production in the past is debatable but it's clear that total farm salmon production cannot be used as a 'proxy for (BKD) pathogen exposure' for the salmon farming industry as a whole given that it (BKD) impacts only a small fraction of the production (Pacific salmon) who's proportion of the total production has varied widely over time. Thus, it is not reasonable to use farm salmon production as a proxy for 'BKD pathogen exposure'.

Vibrio has been detected in farmed salmon sporadically between 2002 and 2010 (8 cases in total) and most of the cases of furunculosis (34 of 56) have been from farms on the West Coast of Vancouver Island (Noakes 2011). The incidence of these two diseases on farms located near or along the migration path for Fraser River sockeye salmon is very small (~3 or fewer farms per year for both diseases combined) and it would be unreasonable to use farm salmon production as 'a proxy of pathogen exposure' for these two diseases for the entire industry. This is not entirely unexpected since farmed salmon have been routinely vaccinated for these two diseases since the mid-1990s. Presumably the efficacy of the vaccines for these two diseases is not zero so the incidence of these diseases (and the resulting pathogen exposure) is also not consistent over the entire time series. The use of vaccines represents an intervention in the time series when the incidence and dynamics of these diseases changed. The same argument also applies to IHN since a vaccine for this disease is routinely used in the industry. Thus, farm salmon production is clearly not a proxy for vibrio or furunculosis 'pathogen exposure' over the entire time period.

With respect to sea lice, Marty et al. (2011) demonstrated that the best predictor for sea lice on farmed salmon was the number of Pacific salmon returning to spawn the previous fall. Thus, the average number of sea lice per fish is not strictly proportional to production but conditional on the number of Pacific salmon returning the previous year as well as several other environmental variables (such as salinity and temperature). Also, since 2003 sea lice levels on farmed salmon have been actively managed (artificially constrained) and thus there has been a structural change in the time series. Again, it is therefore inappropriate to use farm salmon production as a proxy for sea lice abundance (infestation) or 'pathogen exposure' associated with sea lice.

Some have also suggested that the live (apparently healthy) fish on the farms are shedding vast numbers of pathogens but that is a very questionable argument at best. If there were vast numbers of pathogens (enough to cause disease in fish migrating past the farms) then surely there would be an impact on the fish in the net pens. Also, the argument does not mitigate or explain the structural changes in the disease (or pathogen) time series resulting from the introduction of vaccines or in the case of sea lice the mandatory treatment to keep lice numbers below a specified threshold. It is also important to remember that these are endemic diseases and salmon farms are not the only source of pathogens (wild and hatchery salmon being substantial and obvious sources of these pathogens as well). Those choosing to use the shedding of pathogens from live (healthy) fish argument must apply the same process equally to farmed, wild, and hatchery fish particularly given the high incidence of some of these diseases in wild fish (Kent et al. 1998).

Weighing all of the available evidence, it's clear that farmed salmon production is not a reasonable or appropriate 'proxy of pathogen exposure' for any or all of these 4 'high' risk diseases or sea lice. Connors assumption is not consistent with the detailed fish health data (evidence) available from the salmon farms and the use of vaccines and the treatment of lice have resulted in structural changes (interventions) in almost all of the time series that cannot be ignored and cannot be modelled with the available data. Therefore, the use of farm salmon production as a 'proxy for pathogen exposure' in Connors model is not justified. In his response to my criticisms of his paper, Connors suggests that farm salmon production may be a 'poor proxy for pathogen exposure to wild sockeye' and I believe that to be the case. Testing this assumption (i.e. whether farm salmon production is proportional to 'pathogen exposure' or some other metric of disease and consistent with the observed fish disease data from the farms) should have been the first step in the modelling process.

Given the problems with using the farm salmon production time series as 'a proxy of pathogen exposure', Connors' analysis is questionable (at best) but I'll ignore the problems identified above and address some of the modelling issues that have been raised. Connors uses a linear statistical model that also incorporates one or more nonlinear interaction terms (the nonlinear terms produce a nonlinear curved response surface). The interaction terms purposely introduce multicollinearity into the model (i.e. the independent variables are highly correlated with the products of the individual variables) which can result in erratic behaviour of the model as well as other statistical problems. Considerable care needs to be taken when interpreting the results and there is some evidence of erratic behaviour with the reversal of some relationships in the model resulting from minor changes in the data. One problem that warrants attention is how Connors incorporates the independent variables (i.e. 'SST', 'Farm' and 'Pink') into his model (equation 4, page 14, Connors 2011). For example, the farm salmon production time series ('Farm') has a strong trend and exhibits significant autocorrelation (Noakes 2011). While Connors used survival anomalies (from a Ricker or Larkin model) to examine relationships with shorter sea lice and disease records, that does not appear to be the case in this long-term analysis (page 15, Connors 2011). As formulated, Connors' model (equation 4 or a simplified or restricted model) may

identify spurious correlations (reflected in the regression coefficients) between 'Farm' and the $\log_e(R_i/S_i)$ time series some of which will also have a strong trend and autocorrelation (see for example Figure 2 and Appendix 2 in Noakes 2011). The same likely applies to the SST and Pink salmon time series (I expect both these time series are autocorrelated, pink salmon at lag 2, and both may also exhibit trends) as well as the various products of these three variables. There is certainly a strong upward trend in the pink salmon time series. The multicollinearity introduced through the interaction terms further complicates interpretation of the results. Even ignoring the problems with using farm salmon production as 'a proxy of pathogen exposure' it's not clear what useful information can be gleaned from these models given the significant potential for spurious correlations. The inability to incorporate an autocorrelation component in the model is also a significant problem.

I am very familiar with the AIC having used it extensively for many years to fit and select a variety of models and I understand Connors' use of the AIC to examine the fit and parsimony of potential models. Many statistical programs use the AIC to automatically rank models but an important caveat when using the AIC (or BIC or other model ranking or selection criteria) is that you should only compare models that pass diagnostic checks (i.e. statistically significant fit and reasonably satisfy the model assumptions). Ensuring all of the models being considered pass diagnostic checks avoids statistically questionable results and speculation. The result may be several candidate models with no one model being the 'best' in which case the analyst would then use the AIC as well as any other information available to determine which model (or models) to use. Models containing interaction terms are more difficult to access because of the multicollinearity but of the single variable models considered by Connors (Table 6, page 21), only the model containing SST potentially passes diagnostic checks – potentially because of the likelihood of a spurious correlation due to trends and autocorrelations in the SST (see the discussion above). The discussion about possible interactions is also highly questionable given farm salmon production is clearly not a reasonable 'proxy for pathogen exposure' and problems with the model discussed above. The AIC can be a useful tool for model selection but it's not a particularly meaningful discussion in this case given the significant problems with the data.

Pink salmon may influence Fraser River sockeye salmon (although there is no strong evidence to support this assumption). There is no apparent 2-year cycle (pattern) in the $\log_e(R/S)$ for Fraser River sockeye which you would expect to see if pink salmon abundance was a significant issue. Chum salmon as well as other stocks of sockeye clearly compete on the same trophic scale and should be considered along with many other covariates. There are also many other fish populations (such as Pacific herring and Pacific hake stocks in the Strait of Georgia) that could also influence Fraser River sockeye salmon before they pass any salmon farms and these should be considered in the future but these are well beyond the scope of the terms of reference for Project 5.