Review of Papers Pertaining to Salmon Survival in Relationship to the Closing of the Delta Cross Channel Gates and Export Pumping

Bryan F.J. Manly
Western EcoSystems Technology Inc.
Cheyenne, Wyoming
bmanly@west-inc.com

Contents

Summary

- This review examines two reports of analyses of the results of mark-recovery experiments on salmon smolt migrating through the lower Sacramento River system to the ocean. The mark-recovery experiments have taken place for more than 20 years. Releases of tagged juvenile salmon from hatcheries have been made at various locations in the river system, with some recoveries downstream from trawl fishing at Chipps Island, and some recoveries two to five years after release from ocean fishing. Based on certain assumptions, the mark-recovery data allow the estimation of survival rates through different portions of the river system. These survival rates can then be related to the conditions when the salmon smolt were migrating, as measured by a large number of potential covariates such as the water temperature at the time of release.

- The review concentrates on how the survival rate is related to two management variables: (a) the effect on smolt survival of the closure of the delta cross channel gates about 6.5 miles south of Courtland, and (b) the effect on smolt survival of state and federal export pumping.

- Newman and Rice (2002) consider the analysis of the results obtained from the river and ocean recoveries of 101 groups of marked juvenile chinook salmon from the Feather River Hatchery. Releases were made in the spring over the years 1979 to 1995, in seven release areas, with a great deal of variation in the number of releases for year-area combinations.

- Newman and Rice used a model that assumes that the trawl fishing and ocean fishing recoveries of salmon follow independent over-dispersed Poisson distributions. This is a reasonable model given the nature of the data.

- Overall, the Newman and Rice paper does not give clear evidence of any effects of the delta gates being open or of an effect of exports. The regression coefficients for export related variables are not significantly different from zero at the 5% level. Also, although the delta gate being open seems to just have a significant effect, this could be due to the standard error not being estimated very well because it seems that the estimated standard errors of regression coefficients are subject to sampling errors of up to about 10-15%.

- There are two versions of the Newman (2003) paper, which are almost or completely the same. One is an unpublished report (Newman, 2003a), and the other is a published paper (Newman, 2003b). Comments are provided here on the contents of the published paper. The paper considers the analysis of paired release-recovery data for salmon, grouped in such a way that for each group there are several upstream releases and one downstream release, with all releases made at about the same time. Recoveries from the upstream releases were made by trawl fishing at Chipps Island,

and from ocean fishing two to five years later. Downstream releases were made below the trawl fishing area and therefore only ocean recoveries were possible. Most of the data from the 61 upstream releases were included in the data set of 101 upstream releases analyzed by Newman and Rice (2002). Upstream releases were made near the city of Sacramento (about 50 miles above the trawl fishing), near Courtland (about 38 miles above the trawl fishing), and near Ryde (about 30 miles above the trawl fishing).

- Three methods of analysis were considered by Newman (2003). The first method was based on a multinomial model, of a type often used for describing release recovery data. This method can be expected to produce reasonable estimates of model parameters, but with estimated standard errors that are definitely too small because ocean recovery numbers are estimated rather than known exactly but the sampling errors are not allowed for, although they will be large. The second method, called pseudo-likelihood, is better in this respect, but unfortunately over-dispersion is not properly allowed for with the ocean recoveries of downstream releases although there is no doubt that this over-dispersion exists. Therefore, the estimated standard errors for this method will also be too small, to an unknown extent, again making clear inferences impossible. The third method of analysis involves a hierarchic Bayesian model. This includes many assumptions that cannot be verified, and in some cases are definitely not correct. In particular the ocean recovery numbers are treated as being exact rather than estimates. This model appears to give clear evidence of a negative effect of the gates being open and exports. It is possible that the results from this model are reasonable, but it is suggested that before they can be accepted the model requires to be tested properly. This would involve simulating data under a variety of scenarios, with the assumptions of the model both correct and incorrect. The trawl and ocean sampling processes would need to be incorporated into the simulations. If the model is found to be robust to assumptions then the estimates from the model can be considered seriously. Thus the analyses of Newman (2003) may give some indications of the negative effects of the cross channel gates and exports on the survival of salmon in the Sacramento river, but it does no more than that without the robustness of the hierarchical model being studied further.

Introduction

This review examines two reports of analyses of the results of mark-recovery experiments on salmon smolt migrating through the lower Sacramento River system to the ocean. The mark-recovery experiments have taken place for more than 20 years. Releases of tagged juvenile salmon from hatcheries have been made at various locations in the river system, with some recoveries downstream from trawl fishing at Chipps Island, and some recoveries two to five years after release from ocean fishing. Based on certain assumptions, the mark-recovery data allow the estimation of survival rates through different portions of the river system. These survival rates can then be related to the conditions when the salmon smolt were migrating, as measured by a large number of potential covariates such as the water temperature at the time of release.

Although the relationship, if any, between smolt survival and any covariate is of potential interest, this review concentrates on how the survival rate is related to two management variables. The particular questions addressed are:

- What do the reports indicate about the effect on smolt survival of the closure of the delta cross channel gates about 6.5 miles south of Courtland?

- What do the reports indicate about the effect on smolt survival of state and federal export pumping?

- If smolt survival seems to be affected by the closure of the delta cross channel gates, what is the apparent magnitude of the effect?

- If smolt survival seems to be affected by export pumping, then how much change in survival is apparently associated with different amounts of change in export pumping?

These questions are phrased in terms of indicative and apparent effects because the nature of the data are such that any relationship between survival and a covariate, however strongly supported by an analysis, may be partly of fully due to the affects of unknown and unmeasured covariates. The problem is that although the mark-recovery data are obtained from what can be considered as experiments, the analysis is really on observational data because of the large number of uncontrolled variables when releases took place. Newman and Rice (2002, p. 8) make the same point.

This situation is not in any way unusual. It is just important to keep in mind that the results of an analysis on observational data may always be upset to some extent by the confounding effects of covariates that are not included in the model used for the analysis.

Newman and Rice (2002)

General Comments

The paper considers the analysis of the results obtained from the river and ocean recoveries of 101 groups of marked juvenile chinook salmon from the Feather River Hatchery. Releases were made in the spring over the years 1979 to 1995, in seven release areas, with a great deal of variation in the number of releases for year-area combinations.

Most of this paper is devoted to the consideration of a model that assumes that the number of trawl recoveries and the number of ocean recoveries of salmon follow independent over-dispersed Poisson distributions. This model seems reasonable in general. The authors provided estimates with and without the use of a special type of estimation called ridge regression. Ridge regression is not used much in practice. Here the reason for using it is said to be to stabilize estimates. Generally, the difference between the estimates with and without ridge regression are not that great, although the standard errors for coefficients estimated by ridge regression are slightly less than they are using standard estimation. As the use of ridge regression has apparently not been used before with the type of model considered by Newman and Rice, I believe that some sort of simulation study of the properties of the method in this context is needed before estimates based on this method can be said to be preferred to estimates obtained using the usual methods.

Following the main analysis and the discussion of the results, some alternative methods of analysis are briefly discussed. Some of the modeling described here seems to me to be somewhat ad-hoc. However, as these analyses received little attention by the authors they are not really important as far as the main thrust of the paper is concerned.

In general I think that the analysis presented by Newman and Rice is reasonable for the data. The model is sensible, and although I have criticized the use of ridge regression, this is only a detail that does not influence the results much. Also, I think that the approach to modeling ocean recoveries is a little unusual, but the final estimates of regression coefficients and their standard errors would be about the same for an alternative approach that I suggest below.

Overall, the paper does not give clear evidence of any effects of the delta gates being open or of an effect of exports. To begin with the regression coefficients for export related variables are not significantly different from zero at the 5% level. Also, although the delta gate being open seems to just have a significant effect, this could be due to the standard error not being estimated very well. As shown below, it seems that the estimated standard errors of regression coefficients are subject to sampling errors of up to about 10-15%.

Finally, I have to agree with the last sentence in the paper, which notes the need to collect mark-recovery data with specific manipulations of the river system in order to properly clarify the effect of these manipulations on the survival of smolt.

Specific Comments

(1) It is explained at the start of Section 2 that the total ocean recoveries of marked fish are estimated by sampling the marine catch at different times and places. For each time and location it is then possible to estimate the total catch of marked fish of different ages. Adding the contributions up from the different times and places then makes it possible to estimate the total number of marked fish recovered from any release. This is all straightforward in terms of standard sampling theory. What is not clear is why Newman and Rice then say that the estimated total recoveries can be written approximately as

$$\hat{y}_o \approx \sum\sum\sum e_{atp}\, y_{atp}, \qquad (1)$$

where the summations are over age (a), the time period (t), and the landing area (p), e_{atp} denotes the reciprocal of the sampling fraction for the atp combination, and y_{atp} denotes the catch for the atp combination. The equation for estimation should be exact, not approximate. This leads me to wonder if there is some approximation involved here that is not discussed.

(2) The covariates used seem generally reasonable. I presume that the discussions between interested groups that led to this choice means that it is a good choice given the need to keep the number of covariates as small as possible. My only concern in this area is the use of the ratio of exports to total flow to account for any effects of exports. This is discussed by Newman (2003) who used the median exports instead of the ratio of exports to the total flow. In terms of the effects of exports it seems that more consideration is needed about the ways that this is allowed for in the modeling, considering the possibility of a quadratic term, for example.

(3) In connection with the ocean recoveries, it is not clear to me why the sampling at this stage was not treated the same way as the trawl sampling in the river. In both situations there is an estimated sampling fraction. For a river recovery it is f_r, while for an ocean recovery it is $f_o = 1/e_o$ as defined by equation (1) of Newman and Rice. In the final model for μ_r it is assumed that

$$\mu_r = R\, S\, p = R\, S\, f_r q,$$

i.e., it is assumed the expected number recovered is equal to the product of the number release (R), the probability of surviving to the trawl recovery area (S), and the sampling probability (p), where the sampling probability is equal to the fraction of the river sampled (f_r) times some constant (q). The model for river recoveries then assumes that the observed number recovered follows an over-dispersed Poisson distribution, with the known product R f_r serving as what is usually called an offset. This is a completely standard formulation. The equivalent approach with ocean recoveries would be to assume a mean of

$$\mu_o = RS(1 - p)\pi \approx RS\pi,$$

for the actual ocean recoveries, where the approximation is reasonable because p will be so close to zero. Then μ_o is approximately the product of the number released, the probability of surviving to the trawl sampling area, and the probability of surviving in the ocean and then being recovered. The known approximate sampling fraction f_o can be introduced by assuming that $\pi = f_0\,\theta$, where f_0 is the sampling fraction for the ocean fishery and θ is the probability of surviving in the ocean and being recovered with 100% of the fishery sampled. Then

$$\mu_o = R\,S\,f_o\,\theta,$$

which matches the model for the river fishery. The known values R f_o would give the offset. Rather than using this obvious approach, Newman and Rice assumed that the much larger estimated values $\hat{y}_o$ follow over-dispersed Poisson distributions. One effect of this is that their heterogeneity factor c_o will be much larger than what would be obtained by my formulation. Indeed, if the mean value of the expansion factor e_o is 4.4 then I would anticipate that the heterogeneity factor c_o for my model would be approximately the value estimated by Newman and Rice divided by 4.4, i.e. 42.49/4.4 = 9.7. This is close to the value for river recoveries, and indeed with my formulation only one heterogeneity factor may be needed. Thus I believe that the model used by Newman and Rice has very much exaggerated the heterogeneity in ocean recoveries. I do not think that the estimates and standard errors would change much with my formulation, so I am not claiming that their method is wrong, just that it is strange that they did not treat the sampling of river and ocean recoveries in the same way.

(4) In Section 3.5 Newman and Rice state that the heterogeneity factors c_r and c_0 are estimated by dividing deviances by the error degrees of freedom. This is a standard approach, but it is sometimes not appreciated that the estimates may not be very precise. With one heterogeneity factor the standard errors of estimates of regression coefficients are proportional to the square root of the heterogeneity factor. With two heterogeneity factors such as used by Newman and Rice the situation is more complicated, but still it can be expected that the same type of proportional relationship applies. To get some idea of the level of sampling error that may be involved with estimating heterogeneity factors, a small simulation study was carried out, as described in Appendix A. This only involved the estimation of one heterogeneity factor, and is therefore simpler than what is needed to properly

examine the situation for Newman and Rice's model. Still, it indicates that there is a possibility that the estimated standard errors for regression coefficients are too large or too small by perhaps 10-15%. Unfortunately, looking at Figure 2 of Newman and Rice, the two borderline regression coefficients are for the delta gate, and the ratio of exports to inflow. The delta gate effect is just significant at about the 5% level on this figure, but it could easily be the case that it would not be significant if the true heterogeneity factors were known. Similarly, the export to inflow ratio is not significant at about the 5% level on Figure 2, but might become significant if the true heterogeneity factors were known.

(5) The residual diagnostics used in Section 4.3 are reasonable. It is noted by Newman and Rice that residual plots indicate that there are some unaccounted for year effects. This is unfortunate because if there are important covariates that are missing from the analysis then this may be affecting the regression estimates that are present.

(6) As some of the regression coefficients are clearly not significant I am a little surprised that the sequential model selection procedures (backwards and forwards) kept all of the variables in the model. This is another area where I would like to see some empirical evidence from simulations that the method has good properties with data and models as considered here. I have not used BIC myself, but I know from experience that AIC (which is a similar method) can have poor properties under some circumstances.

Newman (2003)

General Comments

There are two versions of this paper, which seem to be almost or completely the same. One is an unpublished report (Newman, 2003a), and the other is a published paper (Newman, 2003b). The review here is based on the published paper. In particular, the page numbers quoted below are from the published paper.

The study considers the analysis of paired release-recovery data for salmon. That is to say, release-recovery data were grouped in such a way that for each group there are several upstream releases and one downstream release, with all releases made at about the same time. Strictly speaking the data are better described as 'matched' rather than 'paired' because there were a total of 61 upstream releases and only 19 downstream releases. Therefore there were on average about three upstream releases matched with each downstream release. Recoveries from the upstream releases were made by trawl fishing at Chipps Island, and from ocean fishing two to five years later. Downstream releases were made below the trawl fishing area and therefore only ocean recoveries were possible. Most of the data from the 61 upstream releases were included in the data set of 101 upstream releases analyzed by Newman and Rice (2002), as discussed above.

Upstream releases were made near the city of Sacramento (about 50 miles above the trawl fishing), near Courtland (about 38 miles above the trawl fishing), and near Ryde (about 30 miles above the trawl fishing).

The reason for matching the upstream and downstream releases was the idea that the ocean recoveries from the downstream releases would serve as controls to reflect the ocean survival and recovery probabilities for fish that did not have to survive a passage of 38 or more miles in the Sacramento river before entering the ocean. The matching of the downstream releases to upstream releases made at about the same time may then allow to a large extent for changes in the survival conditions over time that are not easily accounted for by measured covariates.

Three methods of analysis were considered. The first method was based on a multinomial model, of a type often used for describing release recovery data. This method can be expected to produce reasonable estimates of model parameters, but with standard errors that are definitely too small. The problem here is that ocean recovery numbers are estimated rather than known exactly but errors of estimation are not allowed for although they will be substantial. The second method, called pseudo-likelihood, is better in this respect, but unfortunately over-dispersion is not properly allowed for with the estimated ocean recoveries of downstream releases although there is no doubt that this over-dispersion exists. Therefore, the estimated standard errors for this method will also be too small, to an unknown extent, again making clear inferences impossible.

The third method of analysis involves a hierarchic Bayesian model. This is clearly the model favored by Newman. However, there are many assumptions in the hierarchic model that cannot be verified, and ocean recovery numbers are treated as being exact rather than estimates. It is possible that the results of the model are reasonable, and that the effect of incorrectly assuming that ocean recovery numbers are exact rather than estimated is taken account of by assuming distributions for p, S and π. However, the assumed distribution for ocean recoveries is not correct and I believe that effects of this and other assumptions need to be tested. This would involve simulating data under a variety of scenarios, with the assumptions of the model both correct and incorrect. The trawl and ocean sampling processes would need to be incorporated into the simulations. If the model is found to be robust to assumptions then the estimates from the model can be considered seriously.

The effects of the cross channel gates being open is estimated to be negative with all three models, and significant under the multinomial and Bayesian hierarchic model. As the standard errors for the multinomial model are definitely too low, the significance for that model can be ignored. The effect is very significant for the hierarchic model, but in the absence of a study of the robustness of this model it is hard to know how seriously to take that. A similar comment can be made about the effects of exports, which is also estimated to be negative with all three models.

Overall, therefore, the analyses of Newman (2003) give some indications of negative effects on survival associated with the cross channel gates being open and exports, but the extent to which this evidence can be accepted depends on the robustness of the model, which has not been properly studied as yet.

Specific Comments

(1) As was the case with Newman and Rice (2002), it is stated in Section 2.1 that the number of ocean recoveries is estimated from a stratified sample, with estimates that can be approximately written in terms of weighted sums of observed recoveries in different years, periods within a fishing season, and landing area. As I noted above, these should be exact rather than approximate equations. This therefore suggests that there is some approximation involved here that has not been explained.

(2) It is stated that in Section 2.2 that the fact that ocean recoveries are estimated rather than observed makes the tri-binomial product model questionable, and over-dispersion likely. This is an under-statement. Over-dispersion will certainly occur. Indeed if the observed recoveries are multiplied by about 4.5 to estimate the total number of recoveries, as indicated in Section 2.1, then the variance will be multiplied by 4.5^2. If an observed counts follow Poisson distributions, which may be a reasonable assumption, then the mean and variance will be the same, say μ. Thus the mean of the estimated total recoveries will be about 4.5μ, and the variance will be about $4.5^2\mu$, and the heterogeneity factor will be about 4.5. As noted in my review of Newman and Rice (2002), I think it is unfortunate that the models used are not based on observed ocean recoveries, thus avoiding at least one source of over-dispersion. As it is, the tri-binomial product model is clearly not appropriate and there seems little point in even using it to analyze the data.

(3) The reason for using the pseudo-likelihood model is to allow for over-dispersion in the river and ocean recovery data. As noted in (b) above, this is necessary, particularly for the ocean recovery estimates because they are likely to have a heterogeneity factor of about 4.5 because of the method of estimation. Unfortunately, it is stated on page 164 that over-dispersion cannot be estimated for the ocean recoveries, and the heterogeneity factor was fixed at 1.0. This seems unsatisfactory, particularly as the over-dispersion associated with the estimation process could have been allowed for by taking into account the sampling fractions for ocean recoveries, which are presumably known. Or, better still, the actual numbers recovered could have been modeled. At any event, the estimated standard errors for some estimates at least will tend to be too small for the pseudo-likelihood model.

(4) The problem with the hierarchic model is that there are so many unverified assumptions made. This makes it difficult to know what to make of the final results from the analysis. Here are comments on specific assumptions:

(a) It is assumed (page 165) that y_{ut} and $\hat{y}_{uo}$ are trinomially distributed, and $\hat{y}_{do}$ is binomially distributed, with their joint probability given by equation (3.1, given the values of p, S and π. This is certainly not true as the ocean recoveries are estimated in such a way that the variance is likely to be about 4.5 times higher than this assumption would suggest. It is not clear what effect, if any, this has on the analysis.

(b) When trawl capture rates (p) are release-specific the prior distribution is assumed to be uniform between 0 and 0.01, with the upper rate based on the trawl effort measure. It seems unlikely that the assumption of uniformity is true. No doubt data on the trawl effort are available so that this assumption can be checked.

(c) It is said in the text that the ocean recovery rates (π) are assumed to be uniform between 0 and 0.08. However, it is shown to be uniform between 0 and 0.1 in equation (3.8). Again there is no reason to believe that this is true.

(d) The upstream survival rate (S) transformed to log{S/(1 - S)} is assumed to follow a normal distribution with a variance σ_S^2 that is exponentially distributed with a mean of 0.001. These seem to be strange assumptions, but it is stated on page 166 that the net result is to make S approximately uniformly distributed between 0 and 1. It seems very unlikely that this is true, but it can perhaps be argued that this is a pessimistic assumption in the sense that the true survival rates are probably concentrated in a narrower range. It is not clear why the assumption of a uniform distribution for S was not used directly.

(e) When the trawl capture rates are constant except for a change in 1988, a normal distribution is assumed for log{p/(1 - p)} with a variance that is exponentially distributed with a mean of 0.001. Again this is intended to make p approximately uniformly distributed between 0 and 0.01.

(f) The mean values for the exponential distributions for σ_S^2 and σ_p^2 were selected in what is described as a non-standard but pragmatic manner. This involved analyzing three subsets of the data and finding the values of the variances that gave relatively small differences from the model for all of the data. As Bayes' theorem (the basis for the hierarchic model) does not apply if prior distributions are chosen based on the data, I am not sure what the effects of this type of procedure might be. It may be that it leads to optimistic estimation in terms of errors, i.e. makes the estimates in general appear to be more accurate than they really are.

(g) The prior distributions for the coefficients of covariates are assumed to be normally distributed with specific variances. It is not very clear how these variances were chosen, although apparently from the text this involved trying various alternatives. Also, all of the prior distributions of the coefficients of covariates are assumed to be independent. Again, it is hard to know whether this is a reasonable assumption or not.

(5) There is a section in the paper on sensitivity analysis (page 13). Unfortunately this concentrates on one particular assumption (the ocean recovery rate π is the same for upstream and downstream releases made at the same time). There is a serious need for a real sensitivity analysis of the model. This would involve simulating data under a range of conditions, with the assumptions made in the model not necessarily being correct. The process of estimating ocean recoveries should be included. If the model can be shown to be robust to assumptions then it is a reasonable method for analyzing data.

References

Newman, K.B. (2003a). Modeling paired release-recovery data in the presence of survival and capture heterogeneity with application to marked juvenile salmon. Unpublished report data January 16 2003.

Newman, K.B. (2003b). Modeling paired release-recovery data in the presence of survival and capture heterogeneity with application to marked juvenile salmon. *Statistical Modeling* 3: 157-77.

Newman, K.B. and Rice, J. (2002). Modeling the survival of chinook salmon smolts outmigrating through the lower Sacramento River system. *Journal of the American Statistical Association* 97: 983-93.

Appendix A: Simulation Study of the Estimation of Heterogeneity Constants

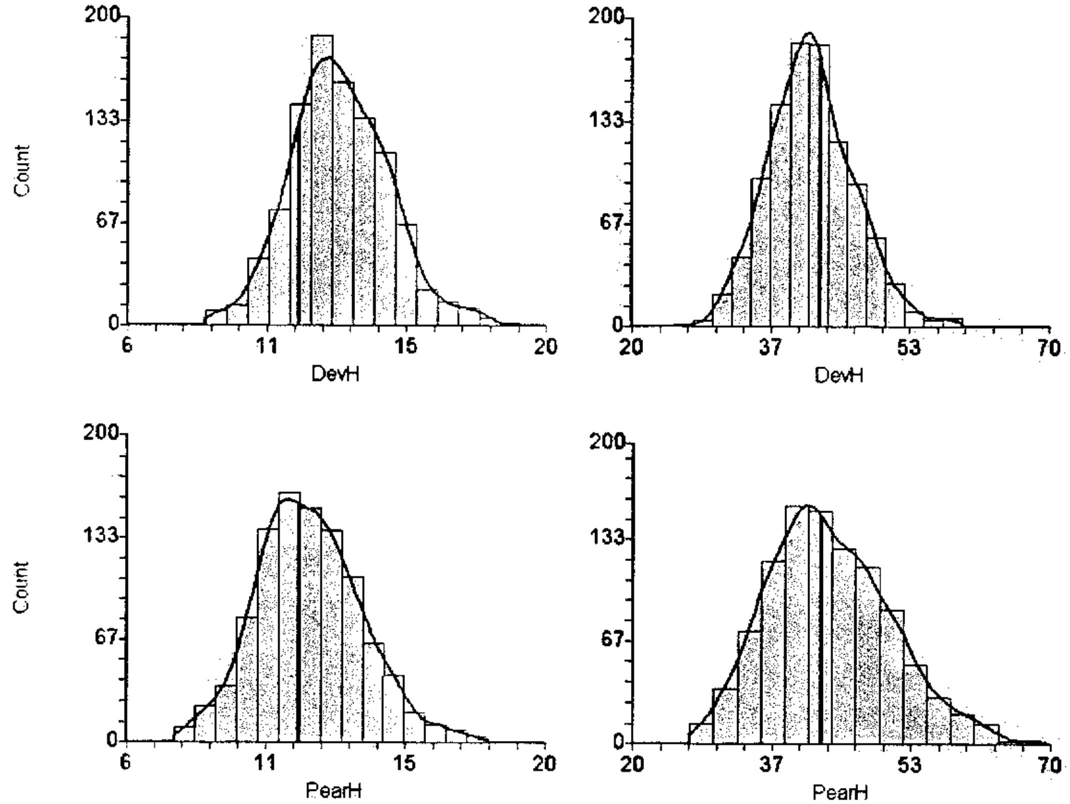
To examine the level of uncertainty that there may be in estimates of heterogeneity constants, data were generated for a log-linear model similar to that for the river and ocean recoveries of salmon. Data were generated for 122 river recovery frequencies, similar to the frequencies in the paired data on Newman's website. Thirteen covariates were used with values as given in the paired data, with each case being used twice to get the 122 rows of data. These variables were given regression coefficients similar to those in the Newman and Rice (2002) paper, where possible. Another 17 variables were introduced with zero regression coefficients and exponentially distributed X values. This then gave values for 30 covariates for the 122 data cases. Corresponding counts were then simulated from over-dispersed Poisson distributions. The regression coefficients were analyzed in the usual way, and the heterogeneity factor was estimated by the deviance divided by the error degrees of freedom, and also by the Pearson chi-squared value for the goodness of fit, divided by the error degrees of freedom.

The process of generating data and estimating the heterogeneity factor was repeated 1,000 times to approximate the sampling distribution for estimates based on the deviance and the Pearson chi-squared value. Two heterogeneity factors were used in separate simulations. These were 11.90 and 42.49, corresponding to the river and ocean recoveries of Newman and Rice (2002). Figure A1 below shows the generated sampling distributions for the two methods for estimating the heterogeneity factor.

When the true heterogeneity factor was 11.90, the mean estimate using the deviance was 13.11, with a standard deviation of 1.93. In this case 95% of the estimates were within the range from 10.04 to 16.63, corresponding to from 84% to 140% of the true value. The average is therefore a bit too high, with over-estimation more likely than under-estimation. In the usual situation where there is only one heterogeneity factor the standard errors of estimates are multiplied by the square root of the factor. Sampling errors in the heterogeneity factor are then likely to give standard errors that range from 92% to 118% of the true value. When the Pearson chi-squared method is used for estimating the heterogeneity factor the mean estimate is 12.12, with a standard deviation of 1.75, and 95% of the values within the range from 8.83 to 15.93. Here there is less bias, and sampling errors in the heterogeneity factor are then likely to give standard errors that range from 86% to 116% of the true value.

The conclusion from these results, is that if the heterogeneity factor is estimated using the deviance then there is some chance of under-estimating the standard errors of regression coefficients, but a higher probability of over-estimation. There is less bias with the use of Pearson chi-squared for estimation, but also a higher chance of under-estimating the standard errors of regression coefficients, which is particularly undesirable.

Figure A1 Sampling distributions for estimated heterogeneity constants. The two graphs on the left are for the situation where the true heterogeneity constant is 11.90, as indicated by the vertical red line, with DevH denoting the distribution with estimation using the deviance, and PearH denoting the distribution with estimation using Pearson's chi-squared statistic. The right-hand graphs are for the situation where the true heterogeneity constant is 42.49.



When the true heterogeneity factor was 42.49 and the deviance was used for estimation, the mean of the estimates was 41.44, with a standard deviation of 5.14, and 95% of the estimates within the range from 31.83 to 52.33. In terms of the effect on the estimation of the standard errors of regression coefficients, the 95% range corresponds to the estimated standard errors varying from 86% to 111% of the true values.

Finally, when the true heterogeneity factor was 42.49 and Pearson's chi-squared was used for estimation the mean was 43.88, the standard deviation was 8.63, and 95% of estimates were within the range from 30.69 to 60.79. The 95% range then corresponds to the estimated standard errors varying from 85% to 120% of the true values.

The reason for considering estimation of the heterogeneity factor using the Pearson chi-squared statistic was the idea that this might work better than the use of the deviance. The simulation gives no support to that idea, although a larger study may show that it is better under certain circumstances.

The main purpose for carrying out the simulation described here was to show that there is error in the estimation of heterogeneity factors, and that this should be taken into account when interpreting the results of the Newman and Rice (2002) analysis. The simulations indicate that estimated standard errors may have errors of up to ± 15% of the true values.