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

Dam Bypass Selectivity Report:

Review of Analyses of Juvenile Fish Size Selectivity in Dam Bypass Systems and Implications for Estimating and Interpreting Fish Survival

Members

Courtney Carothers
John Epifanio
Stanley Gregory
Dana Infante
William Jaeger
Cynthia Jones
Peter Moyle
Thomas Quinn
Kenneth Rose
Carl Schwarz, Ad Hoc
Thomas Turner
Thomas Wainwright

ISAB 2021-1
April 12, 2021

ISAB Dam Bypass Selectivity Report: Review of Analyses of Juvenile Fish Size Selectivity in Dam Bypass Systems and Implications for Estimating and Interpreting Fish Survival

Contents

Executive Summary 1
Background 3
Overarching Comment 5
Review Questions for the ISAB 8
References 16

ISAB Dam Bypass Selectivity Report: Review of Analyses of Juvenile Fish Size Selectivity in Dam Bypass Systems and Implications for Estimating and Interpreting Fish Survival

Executive Summary

In December 2020, NOAA Fisheries asked the Independent Scientific Advisory Board (ISAB) to review the scientific findings and subsequent dialogue associated with two published papers (Faulkner et al. 2019, Storch et al. 2021) that investigated fish size selectivity in juvenile bypass systems and its implications for estimating and interpreting juvenile salmonid survival.

It has long been observed that juvenile salmonids that encounter multiple juvenile bypass systems during downstream migration return as adults, on average, at a lower rate than those that have fewer bypass encounters. Two, non-mutually exclusive, hypotheses have been put forth to explain this phenomenon: 1) bypass systems impart some sort of damage or stress that results in mortality, but not until the fish have completed passage through the hydropower system; 2) bypass systems select for individuals that are smaller or have other characteristics that result in a survival disadvantage regardless of passage routes at dams. Addressing the issue of effect of passage history on ocean mortality is important because the current management strategy of maximizing spill is designed to route fish away from bypass systems.

Faulkner et al. (2019) sought to investigate whether differences in length between fish using alternative passage routes might help explain differences in associated adult return rates. They found that smaller fish were more likely to enter juvenile bypass systems than larger fish and that smaller fish were less likely to return as adults. They also found that apparent effects of bypass passage on adult returns were diminished or disappeared when fish length was taken into account. In a comment to the journal, Storch et al. (2021) were critical of the data and approach adopted by Faulkner et al. (2019). In addition, the 2019 CSS report (McCann et al. 2019) had an appendix (Appendix G) that was also critical of Faulkner et al. (2019).

The ISAB considered the following review questions:

1. Was the Faulkner et al. (2019) analysis scientifically sound, and were the data it used appropriate for addressing the question?

2. Were the conclusions drawn by Faulkner et al. supported by their results?

3. Does the ISAB have recommendations to improve the analysis?

4. Are the criticisms raised by Storch et al. comment and the CSS report appendix valid and supported by the evidence and do any of those criticisms weaken Faulkner et al.'s results or conclusions?

5. Was the Faulkner et al. (2021) response to the Storch et al. (2021) comment appropriate and were their criticisms of the Storch et al. methods valid?

6. Is PITPH an effective index of the powerhouse passage of individual fish, and is it valid to use it to draw causative inferences about effect of powerhouse passage on ocean survival?

One of the key reasons for the disagreements among these papers is a scientific problem known as the *ecological fallacy*. The ecological fallacy occurs when it is assumed that relationships observed for groups necessarily hold for individuals or vice versa. Faulkner et al. (2019) looked at the effect of fish length on survival within populations; Storch et al. (2021) looked at the effect of fish length on survival across populations. For example, considering salmonids in general, the larger individuals of smolts going to sea are more likely to return than are smaller ones of the same population in a given year. However, the average smolt length has little or no explanatory power for predicting the marine survival of that year's cohort relative to smolts from other years, and the average marine survival observed among populations is not strongly associated with fish length either.

Faulkner et al. (2019) estimate individual-level effect of length on return probabilities while Storch et al. (2021) estimate the population-level effect of length on return probabilities. Faulkner et al. (2019) then go further and try to interpret why such a relationship between length and bypass probability may occur. Their discussion about possible size-selectivity of bypass structures is interesting but should be considered conjectural and a new hypothesis to test. Similarly, the Faulkner et al. (2019) discussion of the second finding of fish length affecting return probability is of great interest and yet more tenuous than the first finding. The time from recording length to the return of adult fish is now on the scale of years, which includes possible size-dependent mortality (predation), bioenergetics, and involves other habitats (including the ocean). Faulkner et al. (2019) accurately present these as possibilities in the Discussion (which is appropriate in a scientific publication), but the alternatives are not supported by the actual analyses.

Fisheries managers in the Columbia River Basin may be required to make decisions about management actions primarily intended to influence group-level survival (e.g., flow manipulations). In other situations, they face decisions about management actions that are designed primarily to improve within-group survival of individual fish (e.g., local habitat restoration projects). Managers should be cautious about incorrectly assuming that actions that influence population survival will similarly influence survival of individual fish. Likewise, it would be incorrect to assume that improvements that benefit the survival of individual fish will necessarily benefit the survival of the population. Researchers and managers should clearly identify the biological level (e.g., individual, population, metapopulation, community) of observations used in quantitative analyses and the appropriate biological level to which conclusions and recommendations apply.

The reviews raise important questions about the treatment of the data, questions asked, and analytical methods that require a coordinated (with original authors) or third-party comparative approach. Such follow-up analyses would likely add important insights to the data and the relationships of length with bypass encounters and return rates. The effects of the differences between the two analyses can be assessed and even resolved, which would lead to an even stronger set of findings. **Without such an effort, the discussion and arguments will remain unresolved and allow for easy mis-interpretation of the results of each analysis.** Faulkner et al. (2019) have raised a good set of questions and Storch et al. (2021) have provided a thoughtful response; it would be unfortunate and a missed opportunity not to pursue this further.

In conclusion, the original paper, the comment by Storch et al. (2021) (and Appendix G), and the response by Faulkner et al. (2021) provide an opportunity to make progress on the issue of the role of body length in how the fish use the bypass system and may clarify the effect of length on bypass usage and perhaps, return probability. If there is a size-selection effect on bypass probability, then there may also be an opposite effect on powerhouse passage probability (assuming this is not affected by spill passage probability), which suggests the need for yet another analysis.

Background

Faulkner et al. (2019), *Associations among fish length, dam passage history, and survival to adulthood in two at-risk species of Pacific salmon*, sought to investigate whether differences in length between fish using alternative passage routes might help explain differences in their associated adult return rates. They found that smaller fish were more likely to enter juvenile bypass systems than larger fish, and that smaller fish were also less likely to return as adults. They also found that apparent effects of bypass passage on adult returns were diminished or disappeared when fish length was taken into account. Faulkner et al. presented four major findings:

1. There was strong evidence for a negative association between fish length and the probability of bypass at most dams.

2. There was strong evidence for a positive association between fish length and return probability.

3. There was moderate to weak evidence for a negative association between bypass history and return probability, which weakened further when fish length was included in the models.

4. Researchers investigating return probability or bypass probability should include the length of individual fish in their models.

In a published *Comment* to the Faulkner et al. journal paper, Storch et al. (2021) were critical of the data and analysis approach used by Faulkner et al. (2019). In addition, the 2019 CSS report (McCann et al. 2019) included an appendix (Appendix G) that was also critical of the analysis by Faulkner et al. (2019).

Storch et al. (2021) raised five key issues in their review of Faulkner et al. (2019):

1. Concern about different overall survival between the fish tagged at Lower Granite Dam versus those tagged upstream of the dam used in the analysis of Faulkner et al (2019). Survival differences related to release location may confound analyses of length effects. Faulkner et al. (2019) account for the difference in group-level survival by fitting separate models to each group of fish and exploring the impact of fish length within each group of fish.

2. Concern about inadequate sample sizes to test for bypass effects on survival for fish from the two groups above. Low sample sizes would result in difficulties in isolating and detecting the effects of bypass encounters and length on survival, and failure to detect the effect of bypass encounters does not imply no effect of bypass encounters.

3. Concern that the major drivers of bypass probability at dams were ignored in the analysis. In the Faulkner et al. (2019) analysis, cohort-specific effects (e.g., different locations) are subsumed into group level random effects. Year-specific effects (e.g., different flow patterns) are subsumed into year level random effects. Consequently, small effects on survival due to length may appear greater if these other important effects are minimized by these analytical approaches.

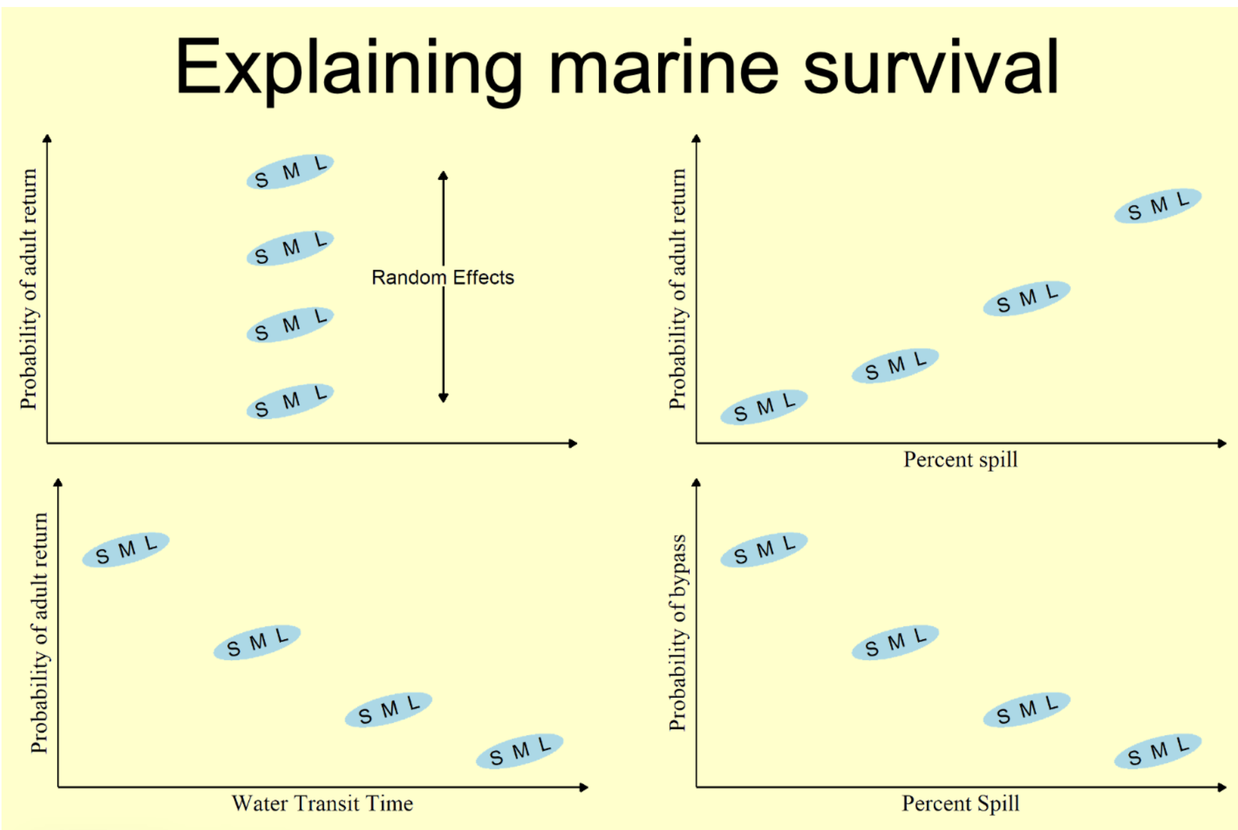
4. Concern that length does not sufficiently explain differences in survival among disparate population groups. This raises the issue that the length effects found by Faulkner et al. have limited generality and suggest that length plays a minor role, overall.

5. Indices of freshwater out-migration conditions continue to explain patterns of marine survival with or without information on individual length. Thus, the effects of length on survival are small relative to the effects of other factors.

A key issue is relevant to concerns 1, 3, 4, and 5 in Storch et al. and accounts for much of the disagreement among the original paper by Faulkner et al. (2019), the responses by Storch et al. (2021), and the rebuttal by Faulkner et al. (2021). This issue is known as the *ecological fallacy problem* where individual-level effects do not apply at the group level and vice versa. For example, considering salmonids in general (i.e., not only the Columbia River system and the species in question here), the larger individuals of the smolts going to sea of a given population in a given year are more likely to return than are smaller ones. However, the average smolt length has little or no explanatory power for predicting the marine survival of that year's cohort relative to smolts from other years, and the average marine survival observed among populations is not strongly associated with body length either.

Overarching Comment

The main analysis conducted by Faulkner et al. (2019) is a logistic regression looking at, for example, the relationship between probability of adult return and length and bypass encounters, after adjusting for other covariates such as the day of release, release site, and random effects for year (which also affect the intercept and coefficient for the effect of day). Consequently, the estimated effect of length is the marginal effect against all other variables (i.e., after correcting for the effects of other variables). In statistical analysis, a marginal effect is the change in the outcome (dependent variable) when a specific independent variable (explanatory variable) changes while all other variables are held constant. The length effect applies within a release group and the analysis attempts to standardize the effect based on the other factors related to the date of release (such as the average size of the fish). So, for each year and release group, there is a marginal effect of length, but when data are combined across years or release groups, there could be no evidence of an effect of length. Consider the simple (exaggerated) examples below:



(Taken from presentation by Storch to ISAB on 2021-03-18)

Within each group (the ovals in the figure), the return probability increases with length (as found by Faulkner et al. [2019]), but when you combine data over multiple groups, the relationship between return probability and length may not be apparent. Similarly, other variables may have a relationship with return probability and their effects may be larger than the effect of length. When the data are broken into groups, the effect of length can be determined more readily because the group "controls" for the impact of the other variables.

Faulkner et al. (2019)'s conclusion about the impact of length on the return probability simply says: "There was strong evidence for a positive association between fish length and return probability," leaving it to the reader to remember this is the marginal impact of length on return probability (that is, the within-group effect shown by each "S-M-L" oval in the figure).

One of Storch et al. (2021)'s concerns about Faulkner et al. (2019) was:

> "length does not explain differences in survival among disparate population groups in the Columbia River Basin"

This is an among-group comparison and not a marginal comparison of the effect of length within each group. Later on, Storch et al. (2021) note that:

> "Positive associations between length and survival at the individual scale may or may not explain patterns of variation at the population level."

Storch et al. (2021) then fit a model that looks at factors that influence bypass or return probabilities across groups and across years, but now the effect of length is across populations rather than within populations. This effect is among the different "S-M-L" ovals. Essentially, the analysis of Faulkner et al (2019) looks within ovals; Storch et al (2021) looks among ovals. It is not surprising that conclusions differ.

This also explains some of the arguments by Storch et al. (2021) about fish tagged at Lower Granite Dam (LGR) versus fish tagged upriver (the ULGR group). Faulkner et al. (2019) fit separate models for the return probabilities for the two groups and found evidence of the marginal impact of length in both groups. Storch et al. (2021) fit a model that combined the two groups into a single analysis and they:

> "found models accounting for tagging site of individuals were substantially better supported than models that ignored that effect."

Storch et al. (2021) found that the return probability for fish tagged at LGR was lower than the return probability for fish tagged at ULGR and concluded that this difference "likely misrepresent[s] survival at the population level." By fitting separate models for the two tagging sites, Faulkner et al. (2019) estimate the effect of length as a within-group (marginal) comparison. By pooling the data from the two release sites, Storch et al. (2021) estimates the effect of length as a combination of within-group and across-group comparisons (semipopulation level) because Storch et al. (2021) used only an indicator covariate for LGR versus ULGR but estimated a common length effect for both populations. A fairer comparison would include an interaction between the effect of site (LGR versus ULGR) and an effect of length, formulated to let each group have separate intercept and slope parameters. Such an analysis would be nearly equivalent to an analysis on the separate groups. In general (and as is discussed more below), it is possible to fit models (random slope models) that allow for among group (population level) and within group (individual level) effects.

Faulkner et al. (2021) responded to this issue with:

> "Instead, we showed that there was strong evidence for an association between length and adult return probability after accounting for other sources of variation, but never made the assertion that length was the primary factor affecting survival."

The key phrase is "accounting for other sources of variation." Because the other sources of variation include year and site effects (either explicitly by including a release-site term in the model or implicitly by fitting different model to LGR and UGLR), the effect of length is measured at a marginal level rather than population level.

Faulkner et al. (2021) responded:

> "Storch et al. contend that a population-level comparison of survival between salmonid populations from different basins somehow invalidates individual-level patterns observed within a single one of those populations. Because fish from the John Day River are slightly smaller on average than fish from the Snake River, but have higher survival, they conclude that results from Faulkner et al. that relate fish size to return probability must be incorrect."

This is the gist of the issue. The ecological fallacy problem unifies these two seemingly contradictory analyses and shows how the analyses are estimating different length effects and are not contradictory. In effect, they address two different hypotheses. Faulkner et al. (2019) estimate the marginal (individual level) effect of length return probabilities while Storch et al. (2021) estimate the group (population level) effect of length on return probabilities.

Similar comments apply about the impact of the number of bypass events. The results of Faulkner et al. (2019) apply marginally to the individual and may not reflect population effects.

The papers by Faulkner et al. (2019), Storch et al. (2021), and Faulkner et al. (2021) could be substantially improved if this distinction about the level of their analyses was made clear. Many readers may not make the distinction between individual (marginal) and population level effects and may make incorrect interpretations due to the ecological fallacy.

Review Questions for the ISAB

1. **Was the Faulkner et al. analysis scientifically sound, and were the data it used appropriate for addressing the question?**

The methods used by Faulkner et al. (2019) were reasonable for the question they actually addressed; however, upon careful examination, the question addressed was quite limited in scope. Whether the length of individuals has an effect on adult return can be formulated as a broad or very narrow question, depending on the details of the analytical method. The analysis in Faulkner et al. (2019) is not a comparison of the effects of length versus the effects of other factors (bypass, number of encounters with dams) that could affect adult returns, even though an initial reading of the paper can leave the reader with the misinterpretation that length is important relative to other effects. The key (as discussed above) is that the analysis of Faulkner et al. (2019) adjusted for the effects of other factors and then looked at whether length is correlated with estimates of adult returns that have been adjusted for other factors. The analysis by Faulkner et al. does not estimate population level effects.

Scientists analyzing the same data may treat the data and formulate response and explanatory variables differently. For example, is it better to analyze data from the two different release locations separately or together? Or should the number of bypass encounters for a fish be treated as a categorical variable (1-2 encounters, 3-5 encounters, 6+ encounters), a yes/no variable (bypass encountered or not), or a linear variable (the impact of number of encounters upon survival is a smooth relationship)? The decisions were documented by Faulkner et al. (2019) and they presented the rationales for these data treatment decisions.

Consequently, the discussion about differences in survival between the two release groups at LGR are moot. It does not matter that one group has a higher/lower survival than other groups when looking at within-group effects of length. The results will be similar if the groups are analyzed separately, or pooled but with a random/fixed effect of group included in the analysis.

There are two (small) additional concerns noted by the ISAB.

Small numbers of "successes" in groups. While the Faulkner er al. (2019) analysis is conducted on information for individual fish, logistic regression (using a logit transformation) can have difficulties when the number of "successes" (e.g., number of adults returned) is zero or small in certain groups. This issue was also raised in Storch et al. (2021) as a criticism of Faulkner et al. (2019).

The problems with estimating very low probabilities on the logit scale is that large difference in the logit can occur when the number of successes only varies slightly. For example,

3% survival of 3/100 has logit of logit(.03) = -3.47
2% survival of 2/100 has logit of logit(.02) = -3.89
1% survival of 1/100 has logit of logit(.01) = -4.59
0.5% survival of 1/200 has logit of logit(.005) =-5.29
0% survival of 0/100 has logit of logit(.00) = -Infinity!

So, if return numbers are small (say averaging around 2%), small changes above and below this overall 2% return are quite asymmetrical on the logit() scale. A small depression in the probability of success gets "magnified" on the logit-scale, so the small change in the probability of success becomes a larger change in the logit value. The analysis uses the logit values as the response variable. Finally, dealing with "infinity" as a response is always problematic.

Faulkner et al. (2019) used the glmer() routine in R to fit the logistic regression, and this routine can deal with a 0% (it actually doesn't estimate -infinity, but arbitrarily stops around -20), but certain combinations of data values can lead to cases where the models do not converge properly without warning, especially if there is a large number of categorical variables. Typical results when the statistical model does not perform well are estimates with non-sensical standard errors that should be caught by a review of the results.

More serious is that small effects of a covariate (e.g., length or bypass) can be "magnified" if the base probability is close to zero. For example, changes in logit(probabilities) when the typical value is around 5% are roughly symmetric around logit(5%), but changes in logit-transformed probabilities when the typical value is around 1% are no longer symmetrical, which can lead spurious results. Because the analysis of Faulkner et al. (2019) breaks data into cohorts, some of which have very few returns or have small sample sizes, there is some concern about maximum likelihood estimates that are known to have a "small sample bias" and overemphasize effects that lead to small probabilities (of return or bypass probabilities) (Nemes, Jonasson, Genell, and Steineck 2009). The small sample bias from maximum likelihood estimates can be removed by using Penalized Maximum Likelihood Estimation (e.g., the Firth [1993] and related methods).

Mixing information theoretic and hypothesis testing frameworks. Faulkner et al. (2019) mix the AIC and p-value paradigms freely (e.g., Table 1 and 2 of Faulkner et al.). This complicates interpretation by the reader, but it is not a major flaw because the two approaches in the Faulkner et al. analysis are consistent in their conclusions. There are some cases where several models appear to fit similarly (based on delta AIC, e.g., Table 2 in Faulkner et al. [2019]). It would be preferable to report model averaged effect estimates rather than relying on p-values from individual models to account for the additional uncertainty of different models giving different estimates of effect sizes. Similarly, model-averaged effect sizes should be reported in Figures 2 and 4 rather than the results from a single model.

2. Were the conclusions drawn by Faulkner et al. supported by their results?

With careful reading, the first two conclusions (that they termed findings) of Faulkner et al. (2019) are reasonably supported by their analytical results. A negative relationship between fish length and bypass probability (at most dams) and a positive relationship between fish length and return probability are supported. A major issue is the apparent small differences in lengths that underlie the contrasts in the analysis. Are these effects ecologically important when placed into context of effects from other factors? Similarly, the narrow scope of inference (i.e., not at the population level) may be overlooked by readers who are unfamiliar with analyses of marginal vs analyses of population effects, and how the choice of statistical model informs one or the other.

Note that findings 1 and 2, as they are stated in Faulkner et al. (2019), did not investigate mechanisms that drive the findings and say nothing about the relationship between bypass history and return probability (which is finding 3). Given the marginal approach of the analyses and that it is statistical (correlative) relationships between fish length and bypass probability (finding 1) are supported by the analysis. Faulkner et al. (2019) then go further and try to interpret why such a relationship between length and bypass probability may occur. Their discussion about possible size-selectivity of bypass structures is interesting but should be considered conjectural and a new hypothesis to test.

The Faulkner et al. (2019) discussion of the second finding of fish length affecting return probability is of great interest and yet more tenuous than the first finding. The time from recording length to the return of adult fish is now on the scale of years, which includes possible size-dependent mortality (predation), bioenergetics, and involves other habitats (including the ocean). Again, the authors provide a nice discussion of possible mechanisms that could lead to the statistical relationship between length and return probability, but the alternatives are not supported by the actual analyses. Faulkner et al. (2019) accurately present these as possibilities in the Discussion, which is appropriate in a scientific publication.

How well the analyses of Faulkner et al. (2019) support the third finding of bypass history affecting return probability is questionable. Results were mixed and of various degrees of explanatory power. There are also issues about the formulation of the variable of bypass history, and the analysis did not account for other, perhaps more important factors affecting return probability. Faulkner et al. (2019) suggested that including length as a covariate reduced the effect of bypass history on the return probability. The authors attempt to unravel such complex relationships, but it is difficult because the analyses have limited power and resolution. Faulkner et al. (2019) present their interpretation in an appropriate manner. The finding is logical (internally consistent) with the analysis when the many assumptions are accepted as reasonable (some are likely not). However, when viewed alone, the finding is more an opinion of a possible relationship then supported by the analysis.

The final conclusion of the paper by Faulkner et al. (2019) is a good summary and is well supported by the results. The authors state that length should be considered in analyses involving bypass and return probabilities. The actual findings stated in the paper have very different levels of confidence and robustness. The paper would have benefited from a clearer distinction of these differences in confidence across the findings, and this would also have reduced the possibility of over-stating the findings.

3. Does the ISAB have recommendations to improve the analysis?

As with many situations of two or more complicated analyses of complex datasets, there is much to learn by doing a follow-up analysis that tests and evaluates the alternative methods and treatment of data. The Faulkner et al. and Storch et al. exchange through formal comments and response (plus this review by the ISAB and Appendix G of McCann et al. [2019]) has nicely clarified the similarities and differences between the two analyses. However, the reviews raise important questions about the treatment of the data, questions asked, and analytical methods that require a coordinated (with original authors) or third-party comparative approach. Such follow-up analyses would likely add important insights to the data and the relationships of length with bypass encounters and return rates. The effects of the differences between the two analyses can be assessed and even resolved, which would lead to an even stronger set of findings. Without such an effort, the discussion and arguments will remain unresolved and allow for easy mis-interpretation of the results of each analysis. Faulkner et al. (2019) have raised a good set of questions and Storch et al. (2021) have provided a thoughtful response; it would be unfortunate and a missed opportunity to not pursue this further.

Any analysis must account for the nesting of individuals within groups within species. It may be possible to model both marginal and population-level responses. Multilevel analysis (e.g., random slope models; Gelman and Hill 2006) enables simultaneous examination of the effects of group-level and individual-level variables on individual-level outcomes (see this brief tutorial on random slope models). Here, each individual group has its own slope (marginal impacts of the covariate), but at the same time the average slope (across populations) measures the population level response. Given the size of the data sets involved (potentially several million records), fitting these models will require the assistance of statistical experts and knowledge about the data.

4. Are the criticisms raised by Storch et al. (in their comment) and the CSS report appendix valid and supported by the evidence, and do any of those criticisms weaken Faulkner et al.'s results or conclusions?

Storch et al. (2021) and Appendix G of the 2020 CSS report raise five primary issues: with the Faulkner et al. (2019) analysis:

1) That there is "evidence of a negative bias in survival associated with the fish tagged at Lower Granite Dam versus those tagged upstream." Storch et al. suggest that this means that this tagging location is not representative of the whole population, although they do not define

what whole population they are considering. They speculate that there may be some tagging effects that might mask bypass effects. However, they do not explain how a difference in tag-group survival (population-level effects) would change the individual effects of length and bypass history on adult return rates.

2) That there were "inadequate sample sizes to test for bypass effects on survival for fish tagged above Lower Granite Dam." Storch et al. (2021) suggest that "limited sample sizes likely precluded detection of an unambiguous bypass effect ..." and use a simulated power analysis to support their argument. However, their description of their power analysis does not seem to be comparable to the structure of the Faulkner et al. (2019) statistical model, so its relevance to the Faulkner et al. analysis is unclear. Regardless, the point by Storch et al. about small sample sizes is valid; however, it applies to only the upriver tagging location which is only part of the overall Faulkner et al. study.

3) That "major drivers of bypass probability at dams were ignored ..." in the bypass probability analysis, including spill proportion and flow. Storch et al. present an alternative statistical model looking at fish length, flow, and spill proportion for selected years, and found that spill proportion has a much larger effect than fish length. This is a legitimate issue to raise, as including spill proportion in the Faulkner et al. model could change their conclusions regarding the marginal effects of length, but it is difficult to know how to apply spill proportions, which can change quite quickly, to the data on individuals.

4) That "length does not explain differences in survival among disparate population groups." This argument is based on a confusion related to the "ecological fallacy" discussed above. Faulkner et al. analyzed individual length as a marginal effect, which does not imply any expected effect of average length in an aggregate population-level comparison.

5) That "indices of freshwater out-migration conditions continue to explain patterns of marine survival with or without information on individual length." Similar to issue 3, Storch et al. suggest that ignoring hydrosystem conditions as predictors of adult survival (as done in Faulkner et al.) could lead to spurious conclusions regarding length and bypass conditions. However, Storch et al. (2021) go on to support this suggestion by using aggregate variables that cannot be directly related to individual bypass histories, and so their model is not easily connected to the Faulkner et al. analysis. It is a legitimate issue to raise, but Storch et al. fail to make the case of a significant problem in Faulkner et al. (2019).

Storch et al. (2021) went further than many reviews and performed analyses to support the issues they raised. As expected with overlap of authors, there was overlap in the comments between Storch et al. (2021) and Appendix G of the CSS 2019 report. The Storch et al. (2021) commentary (and Appendix G) enables the uncertainties of the original Faulkner et al. (2019) analysis to be sharpened and more clearly understood. In a sense, there were many uncertainties in the Faulkner et al. (2019) analysis when viewed alone; Storch et al. (2021) and Appendix G clarified some of the major uncertainties which (whether small or large) can and should be used to help people properly interpret the results of the Faulkner et al. (2019)

analysis. What is needed is an investigation of the well-documented issues in Storch et al. (2021) or those that are more difficult to quantify but have potential to have major effects. Such an investigation might explicitly consider what results from Faulkner et al. (2019) are highly uncertain and what results are robust? How would use of alternative assumptions and associated statistical models affect the effects of length at the marginal and population levels?

5. **Was the Faulkner et al. (2021) response to the Storch et al. comment appropriate and were their criticisms of the Storch et al. methods valid?**

Again, dealing with each of the five issues raised by Storch et al. (2021):

1) Faulkner et al. (2021) acknowledge the difference in survival of fish released at LGR versus above the dam but disagree with the Storch et al. (2021) characterization that the inclusion of both datasets invalidates the Faulkner et al. analysis. The response by Faulkner et al. to this criticism by Storch et al. is adequate and valid.

This issue of two different release groups at LGR is one example of the individual versus population level effects debate, and the difference was summarized in Faulkner et al. (2021):

> "Population-level comparisons of Snake River stocks to those from other basins potentially contribute to the general understanding of salmonid populations in the Pacific Northwest. However, given the true objectives of our investigation, such comparisons really have no place in a critique of our individual-based analyses of fish within the Snake River populations. Storch et al. state '*one should expect that the effect would apply across multiple populations.*' But the effect that Faulkner et al. assert is a relationship between individual length and individual survival. We see no reason that evidence of the individual-level effect necessarily predicts the population-level pattern they looked for."

2) Faulkner et al. (2021) agree that sample sizes for fish tagged upstream of LGR were small but justify procedures that led to selection of this sample. They provide a power analysis that is more closely tailored to their model than the Storch et al. power analysis and find that the power to detect bypass effects was somewhat lower for the above-LGR tag groups than for those tagged at LGR, but not as low as Storch et al. (2021) suggested.

3) Faulkner et al. (2021) suggest that the criticism about not including flow or spill proportion in the bypass probability model was based on a misstatement of their purpose of the analysis. They argue that their method (random effects) for incorporating variation in bypass proportion adequately captures the variation that would be explained by spill proportion or flow, and thus that their analysis of length effects would not be substantially changed by including the other variables. They say the purpose of the Faulkner et al. analysis was not to determine which factors best explain bypass proportion, but rather to look at the effect of fish length after accounting for other sources of variation. This is a valid and adequate response.

4) Faulkner et al. (2021) correctly point out that Storch et al. (2021) misconstrued their conclusions regarding length effects and that population-level comparisons are not relevant to the Faulkner et al. study. This is a valid and adequate response.

5) Faulkner et al. (2021) correctly point out that this criticism (ignoring conditions experienced during migration) is compounded with the "ecological fallacy" issues discussed above; the response of the means of variables is not the same as the response at the individual fish level. Faulkner et al. also point out that the variables used by Storch et al. are not compatible with the individual bypass histories used in the Storch et al. model. This is a valid and adequate response.

The Faulkner et al. (2021) response would be more direct if the main issue, “strong evidence for an association … after accounting for other sources of variation” was more fully explained as an individual within a group effect and not as a population level effect. The phrase “after accounting for other sources of variation” implicitly includes the effect of year, release site, and time of the year.

Further resolution of some of the issues would have been an excellent addition to the response via analyses. The original paper, the comment by Storch et al. (2021) (and Appendix G), and the response by Faulkner et al. (2021) provide an opportunity to make progress on the issue of the role of length in how the fish use the bypass system and may clarify the effect of fish length on bypass usage and perhaps, return probability.

6. **Is PITPH an effective index of the powerhouse passage of individual fish, and is it valid to use it to draw causative inferences about effect of powerhouse passage on ocean survival?**

The Fish Passage Center and the Comparative Survival Study (CSS) have promoted using an index of average cumulative powerhouse passage for groups of fish, which they call PITPH, to capture the effect of passage route taken by juveniles and to estimate the magnitude of delayed mortality in the estuary and ocean. This metric is based on predicted powerhouse passage probabilities from dam passage models and does not track the passage history of individual fish. It is currently being used to guide management decisions regarding the amount of water spilled at federal dams.

This question is perhaps misdirected in using the specific variable “PITPH.” Faulkner et al. (2019) and the comments and responses did not use or discuss PITPH.

The relevant point here includes the more general issue of how to characterize the passage routes and encounters with structures for smolts and how to deal with information that may not be available at the individual fish level but is available at the aggregate level. Faulkner et al. (2021) summarizes it as:

> "There is no ecological or mechanistic reason to conclude that averages of measurements of a process experienced at the individual level are a more appropriate indicator of the experience of the collection of individuals than the individual measurements themselves. It is always better to construct models that represent data at the smallest measured observational unit for which unique information is available and for which processes of interest can be most directly represented. Given that, it is unclear why one would prefer to explain ocean survival of any individual fish using the mean number of bypass events experienced by all fish migrating in a year. The number of bypass events experienced by each of those individual fish is a known and measured variable. If it is in fact the act of passing through a bypass system that is detrimental to ocean survival, then it is much more informative to relate the bypass history of an individual to the survival of that individual."

A more general discussion and evaluation of how to codify the experience of fish with dam-related structures is obviously critical to many analyses and has major implications for management, as illustrated by the exchange between Faulkner et al. and Storch et al. about the many ways to quantify the encounters with bypass in this specific analysis. Comparison and synthesis of alternative, sometimes competing analyses are hindered by similar but different algorithms and bookkeeping when formulating response and explanatory variables. This is unnecessary and clouds understanding of why analyses generate different results.

If there is a size-selection effect on bypass probability, then there may also be an opposite effect on powerhouse passage probability (assuming this is not affected by spill passage probability), which suggests the need for yet another analysis.

References

Faulkner, J.R., B.L. Bellerud, D.L. Widener, and R.W. Zabel. 2019. Associations among fish length, dam passage history, and survival to adulthood in two at-risk species of Pacific salmon. Transactions of the American Fisheries Society 148: 1069-1087. https://doi.org/10.1002/tafs.10200

Faulkner, J.R., D.L. Widener, S.G. Smith, and R.W. Zabel. 2021. Associations among fish length, dam passage history, and survival to adulthood in two at-risk species of Pacific salmon: response to comment. Transactions of the American Fisheries Society 150: 196-206. https://doi.org/10.1002/tafs.10280.

Firth D. (1993). Bias reduction of maximum likelihood estimates. Biometrika 80: 27-38.

Gelman, A., and Hill, J. (2006). Data Analysis Using Regression and Multilevel/Hierarchical Models (Analytical Methods for Social Research). Cambridge: Cambridge University Press. doi:10.1017/CBO9780511790942

McCann, J., B. Chockley, E. Cooper, B. Hsu, G. Scheer, S. Haeseker, R. Lessard, T. Copeland, E. Tinus, A. Storch, and D. Rawding. 2019. Comparative survival study of PIT-tagged spring/summer/fall Chinook, summer steelhead, and sockeye: 2019 annual report. Contract report to the Bonneville Power Administration. www.fpc.org/documents/CSS/2019CSSAnnualReport.pdf

Nemes, S., Jonasson, J. M., Genell, A., & Steineck, G. (2009). Bias in odds ratios by logistic regression modelling and sample size. BMC Medical Research Methodology 9: 56-60.

Storch, A.J., S.L. Haeseker, G. Scheer, J.A. McCann, B. Chockley, T. Copeland, and R.B. Lessard. 2021. Comment: Associations among fish length, dam passage history, and survival to adulthood in two at-risk species of Pacific salmon. Transactions of the American Fisheries Society 150: 189-195. https://doi.org/10.1002/tafs.10279.