STUDY DESIGN FOR MONITORING EFFECTIVENESS OF SOUTH FORK CLEARWATER RIVER VELOCITY BARRIER MITIGATION

Technical Report – *FINAL DRAFT*

Prepared for:

Miranda Main (Nez Perce Tribe)

Prepared by:

Lucius Caldwell

Ryan Flaherty

Kai Ross

September 11, 2019

TABLE OF CONTENTS

List of Figures

List of Tables

BACKGROUND AND CONTEXT

During fall 2016 – winter 2017, Cramer Fish Sciences (CFS) and Northwest Hydraulic Consultants (NHC) completed a study for the Nez Perce Tribe (NPT) that evaluated hydraulic conditions in a reach near MP 28 of the South Fork (SF) Clearwater River (Idaho) hypothesized to present a barrier to adult steelhead migrating upstream (hereafter, "barrier reach"). The results of this work indicated that velocities within the barrier reach likely impede upstream passage for adult steelhead at flows greater than approximately 1,000 cfs (approximately 600 cfs at the Elk City gauge) (Timm et al. 2019; Timm et al. 2017).

CFS and NHC then completed a subsequent study for NPT during fall 2017 – winter 2018 that evaluated hydraulic effects associated with proposed velocity barrier mitigation (precision rock breaking to remove large boulders). The results of this second study indicated that the rock breaking under consideration would likely improve passage through the barrier reach for adult steelhead (Timm et al. 2018).

In light of these findings, NPT is now considering proceeding with velocity barrier mitigation work and has contracted CFS to support this work by designing a study to monitor and evaluate biological effectiveness of this mitigation. For our purposes, we generally confine biological effectiveness to the organismal level, and take it to mean improved adult steelhead passage, defined in terms of proportion or duration of successful passage attempts. Although potentially important, we have not discussed additional biological endpoints, such as reproductive success, that may be considered as indicators of population level success of the project. The remainder of this report summarizes our findings and recommendations for conducting a study to evaluate the biological effectiveness of future mitigation work in the SF Clearwater MP 28 study reach.

DEFINING SUCCESS & OUTLINING ANALYTICAL METHODS

The first step in evaluating "effectiveness" of future mitigation is defining success in terms of a positive biological effect. As a starting point, we suggest constraining monitoring and evaluation (M&E) work to identifying improvements for adult steelhead migrating upstream through the location of the mitigation work ("the study reach"). In addition, we discuss physical improvements that may be construed as evidence of success in the absence of a clear biological signal.

Biological Effectiveness

Comparing Proportion of Success

A hypothesis of biological success based on improvements in the proportion of successful passage of the barrier could then be described verbally as follows: "For flows that currently create barrier velocities ($Q_{barrier}$), the proportion of adult steelhead successfully migrating through the study reach ("passing") will be greater after mitigation (Φ_{post}) compared to before (Φ_{pre})." This starting hypothesis can be formalized as follows:

H1: For $Q_{barrier}$, $\Phi_{post} > \Phi_{pre}$

An associated null hypothesis would posit no difference in passing at barrier flows before and after mitigation work ($\Phi_{post} = \Phi_{pre}$). This null hypothesis can be formalized as follows:

$H1_0$: For $Q_{barrier}$, $\Phi_{post} = \Phi_{pre}$

Evidence supporting this null hypothesis, in the form of probability distributions for the proportion of successful passage among groups of fish migrating pre- and post-mitigation would look something like Figure 1. Note that the probability distributions for the two groups overlap appreciably, indicating that they do not meaningfully differ. Even if statistical tests indicated a significant difference, the effect size would be modest.

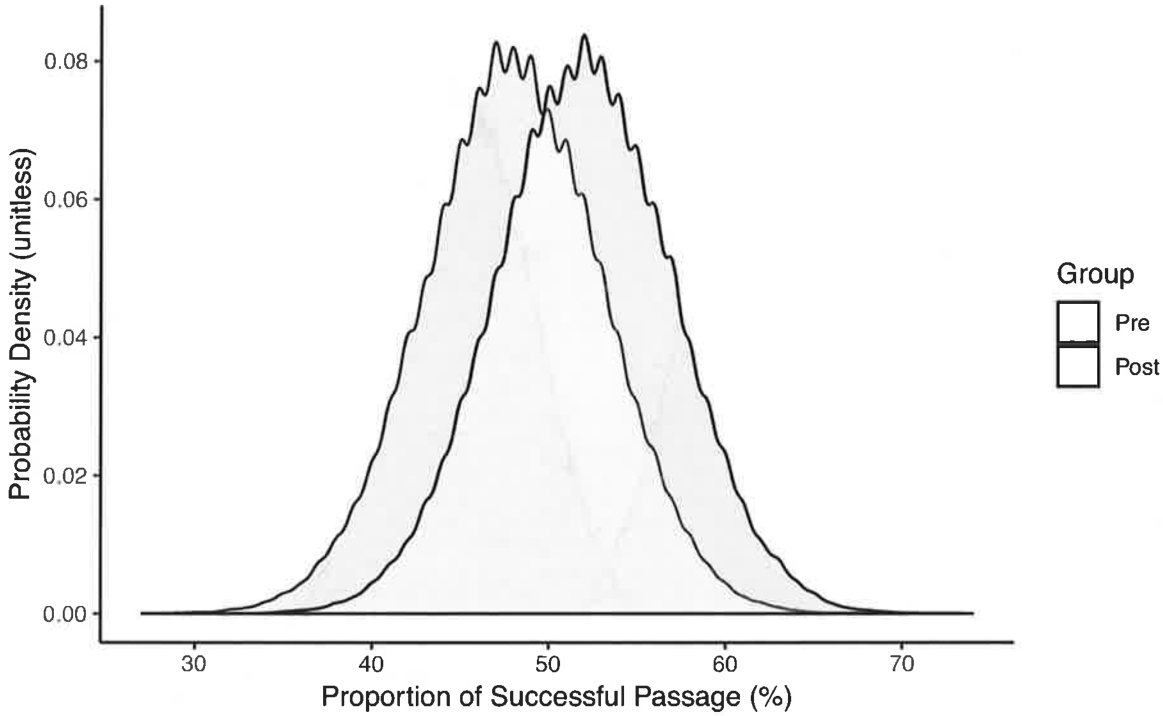

Figure 1. Simulated probability distributions of the proportions of successful passage if mitigation did not lead to meaningful improvement.

On the other hand, evidence refuting this null hypothesis of equal proportions of successful passage pre- and post-mitigation would look something like Figure 2. Here, the probability distributions do not overlap appreciably. Importantly, the amount of overlap can be quantified and used to draw insights about the difference in successful passage at barrier flows among the pre- and post-mitigation groups, in terms of both effect size (e.g., "a 50% (20 percentage point) increase in successful passage") and confidence (e.g., "$p < 0.01$").

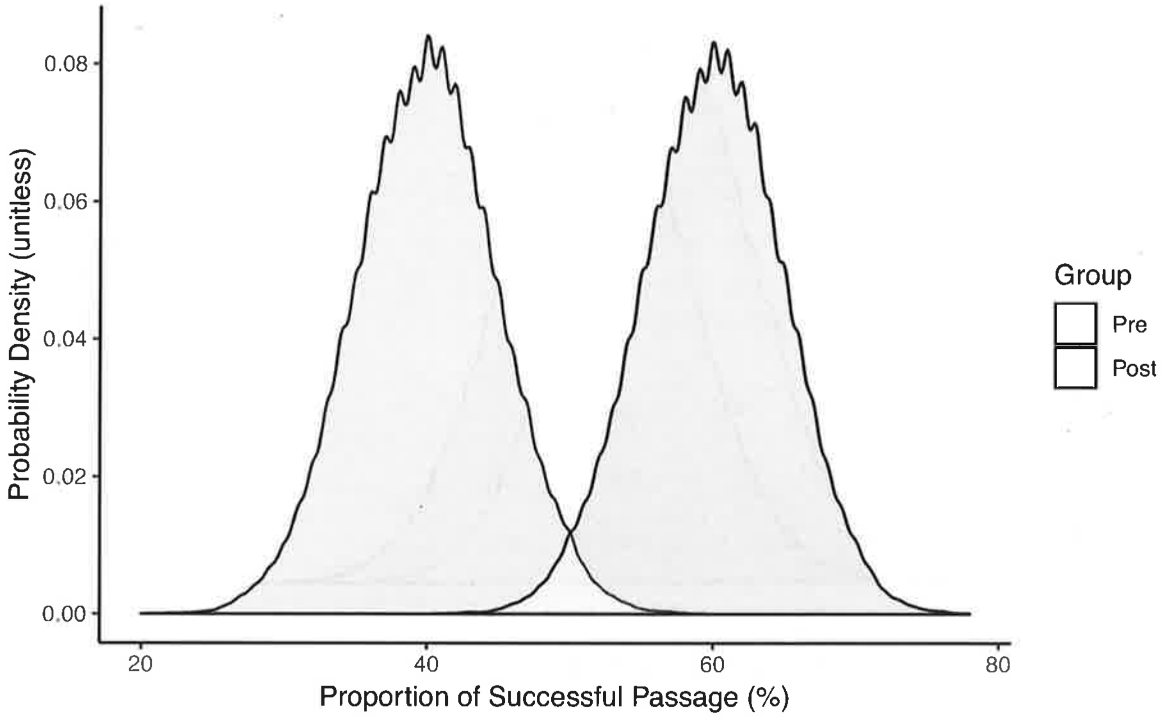

Figure 2. Simulated probability distributions of the proportions of successful passage if mitigation led to meaningful improvement.

While intuitive and analytically simple, this approach requires specifying the critical minimum barrier flow above which to evaluate successful passage. Under current conditions, previous modeling results indicate that velocity barriers to upstream migration emerge when discharge in the reach exceeds 1,000 cfs (which occurs approximately annually), and velocity impediments emerge at discharge as low as 100 cfs (Timm et al. 2019).

However, evidence indicates that steelhead successfully pass the barrier reach at these moderate flows. Steelhead behavioral data from previous bio-telemetry (tracking) projects conducted by NPT indicate that steelhead pass the study reach when estimated discharge in the study reach is as high as approximately 700 cfs, and that some fish may pass at discharges greater than 1,000 cfs (Timm et al. 2017). Therefore, since passage success at moderate flows is already reasonably high, proportion of successful passage may not be expected to appreciably improve at low to moderate flows.

Similarly, it may not be feasible (or desirable) for mitigation work to completely eliminate the emergence of barrier velocities under very high flow conditions. Thus, passage success also may not detectibly improve at very high flows. Given the uncertainty around what the critical flow is, and the importance of accurately defining this value to ensure that the above-described approach yields reliable results, this simple starting metric of comparing the proportion of successful passage among groups of pre- and post-mitigation adult steelhead up-migrants may not perform optimally.

One alternative way to evaluate success across a range of flows would be to iteratively step through a series of analyses comparing the proportion of successful passage, spanning the range of flows observed during steelhead migration. However, this approach is inefficient, time consuming, and difficult to interpret. Iterative analyses require multiple runs of an otherwise identical analysis, altering a single parameter (the threshold flow used to group passage attempts). Although automatable to improve efficiency and time requirements, the presentation and interpretation of such a mass of results would necessitate a time investment unlikely to yield valuable returns. Also, this approach forces lumping passage attempts into flow bin categories ("factorization"). This reduces study power by reducing the amount of information present in the dataset and reducing functional sample size for each sub-analysis. Finally, conducting such an iterative series of analyses would introduce uncertainty by requiring comparisons among groups of results. Many tests would be necessary to determine critical inflection points where passage markedly decreases as flows increase.

Modeling Proportion of Success

A simpler and more statistically robust approach would be to model passage outcome (binary, pass or fail) as a function of flow (continuous). This modeling approach would accomplish essentially the same objective of inferring a difference in the relationship between flow through the barrier reach and passage success among groups of fish that pass before and after mitigation. As the model's dependent output would be binary, and the independent predictor variable input(s) would include at least flow (continuous), this relationship would be most appropriately modeled using a generalized linear model with a logit link, i.e., a logistic regression.

Under this approach, steelhead tracked through the barrier reach would be categorized into treatment groups (pre- and post-mitigation), and the relationship between flow and passage success could then be modeled using one of two broad strategies. Passage could be modeled separately for each of the groups, after which model parameters would be statistically compared among the two groups to infer differences in passage success across observed flows (e.g., using ANOVA-based F-tests). Or, passage could be modeled for both groups together, by including group as a predictor variable, after which the single model would be evaluated to determine if "group" was a significant term (e.g., using likelihood ratio tests, ΔAIC, or parameter specific F-tests). If desired, additional terms could be tested for inclusion in the model structure, to control for ancillary factors known or hypothesized to contribute to passage success (e.g., date of attempt, fish size, fish sex, etc.).

The hypothesis of effect for this modeling approach is similar to hypothesis H1, described above, except it operates across the entire set of observed flows, rather than being specified for a subset of barrier flows. Further, the goals of the modeling approach include not only determining differences in passage before and after mitigation, but also quantifying these differences in order to describe the shape and other key attributes of these relationships. A model-based biological hypothesis of success can be described verbally as, "The proportion of adult steelhead that successfully pass the study reach is a function of flow and group, and the influence of flow differs among groups." Formally, this can be expressed as:

H2: $\Phi \sim Q$ * Group

An associated null hypothesis, describing no effect of mitigation, could similarly be described verbally as, "The proportion of adult steelhead that successfully pass the study reach is a function of flow but *not* group." Formally, this can be expressed as:

$\mathrm{H2}_0$: $\Phi \sim Q\ (\neg\mathrm{Group})$

Evidence supporting this null hypothesis, in the form of logistic curves describing the relationship among flow and successful passage for steelhead navigating the study reach pre- and post-mitigation would look something like Figure 3. Here, the curves describing successful passage are nearly identical in shape (slope) and position (inflection points), indicating that the relationship between flow and passage does not differ meaningfully among the groups.

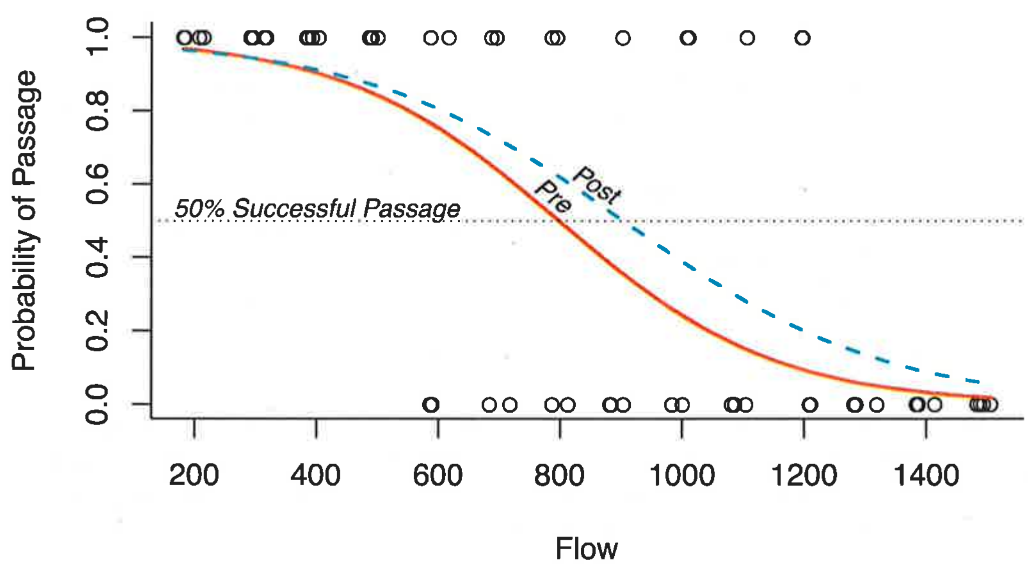

Figure 3. Simulated logistic regression curves relating the proportions of successful passage to flow, if mitigation did not lead to meaningful improvement.

Evidence refuting this null hypothesis of no difference in the relationship between flow and passage among the groups, in the form of logistic curves, would look something like Figure 4. Here, the curves relating flow and successful passage exhibit appreciably different slopes and inflection points. Importantly, these differences can be quantified and used to draw insights about the difference in the relationship between flow and the proportion of successful passage before and after mitigation, in terms of both effect size (e.g., "a 20% increase in successful passage at flow x" or "a 500 cfs increase in the 50% passage flow") and confidence in the statistical result (e.g., "$p < 0.01$").

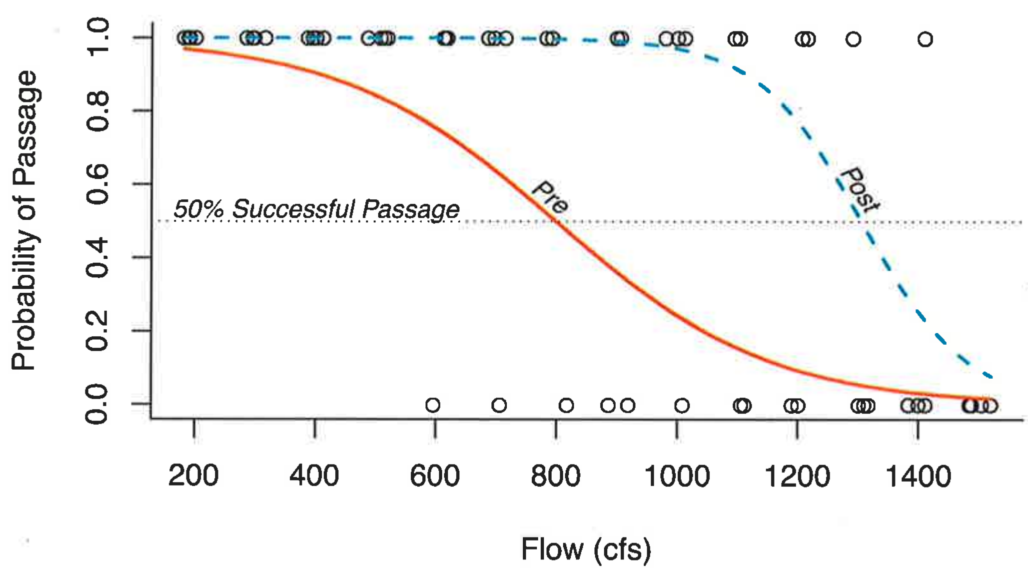

Figure 4. Simulated logistic regression curves relating the proportions of successful passage to flow, if mitigation led to meaningful improvement.

Comparing Duration for Successful Passage

In addition to improving the proportion of fish that successfully pass the study reach, mitigation work may also enable faster passage times. The time required for successful passage of a barrier influences physiological condition of migrating fish (Caudill et al. 2007; Keefer et al. 2004; Naughton et al. 2005), and physiological condition influences spawning (Caldwell et al. 2013; Jenkins et al. 2019; Schneider 2004). For these reasons, duration required for successful passage attempts is frequently included as a target metric for regulatory compliance associated with dam passage (e.g., tailrace residence time) (PacifiCorp and Cowlitz County PUD No. 1 2010; PacifiCorp and Cowlitz County PUD No. 1 2016). Thus, even if the proportion of upmigrating fish that successfully pass the study reach does not improve as a result of mitigation, if the time required for them to pass successfully does improve, this may be construed as project success, since those fish are presumably in better condition and more likely to successfully spawn.

A hypothesis of biological success based on improvements in the time required for successful passage of the barrier could then be described verbally as follows: "Among adult steelhead that successfully pass the study reach, the mean time required to successfully pass the barrier (T) will be less after mitigation (T_{post}) compared to before (T_{pre})." This starting hypothesis can be formalized as follows:

H3: $T_{post} < T_{pre}$

An associated null hypothesis would posit no difference in passing at barrier flows before and after mitigation work ($T_{post} = T_{pre}$). This null hypothesis can be formalized as follows:

H3$_0$: $T_{post} = T_{pre}$

Evidence supporting this null hypothesis, in the form of probability distributions for the mean duration required for successful passage among groups of fish migrating pre- and post-mitigation would look something like Figure 5. Note that the probability distributions for the two groups overlap appreciably, indicating that they do not meaningfully differ. Even if statistical tests indicated a significant difference, the effect size would be modest.

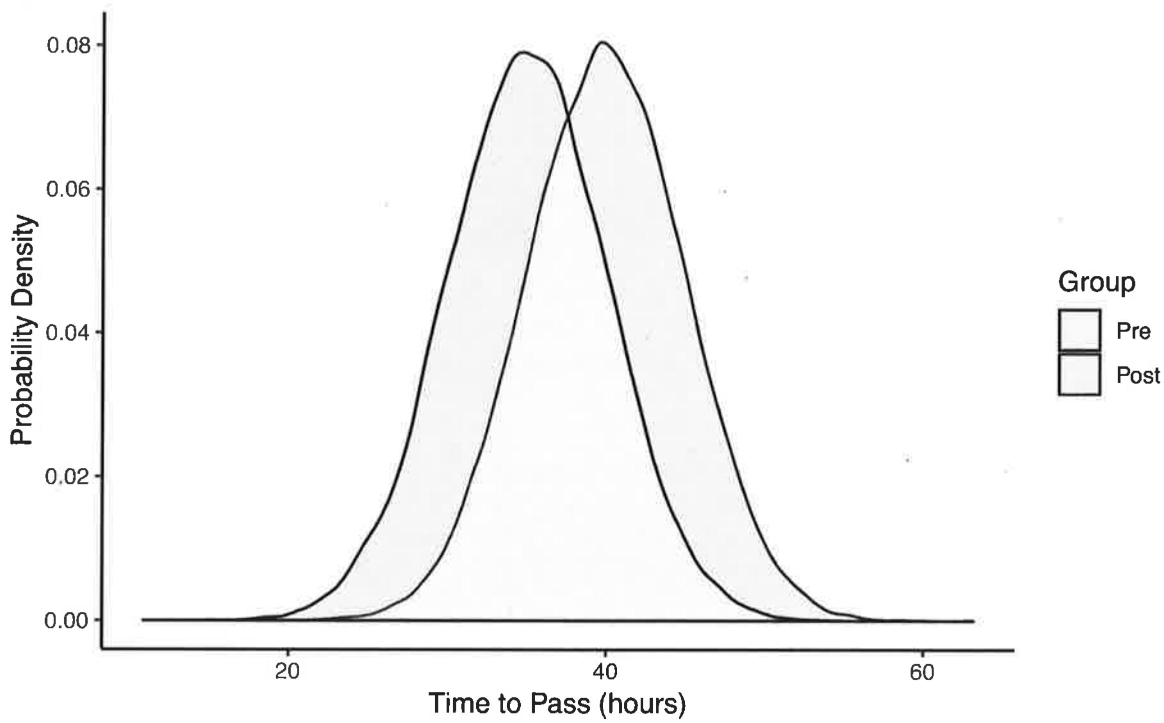

Figure 5. Simulated probability distributions of the time for successful passage if mitigation did not lead to meaningful improvement.

On the other hand, evidence refuting this null hypothesis of equal duration for successful passage pre- and post-mitigation would look something like Figure 6. Here, the probability distributions do not overlap appreciably. Importantly, the amount of overlap can be quantified and used to draw insights about the difference in time for successful passage among the pre- and post-mitigation groups, in terms of both effect size (e.g., "a 20 hour decrease in mean time to successful passage") and confidence (e.g., "$p < 0.01$").

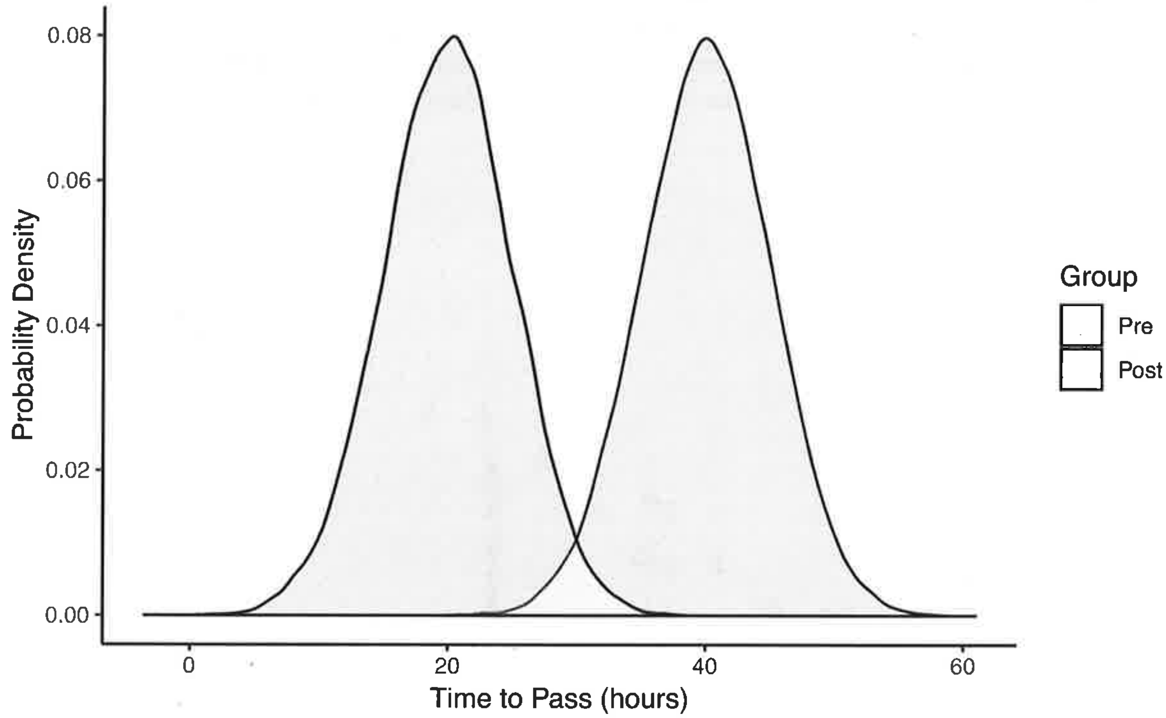

Figure 6. Simulated probability distributions of the time for successful passage if mitigation led to meaningful improvement.

Physical Effectiveness

Biological effectiveness monitoring directly evaluates the project outcome in terms of the specified goal of improving fish passage. However, the signal of a biological response may take years to emerge. Fish migration behaviors integrate many factors unrelated to velocity within the barrier reach. These factors include fish size, physiological condition, and energetic status (Brownscombe et al. 2017), which are in turn influenced by climatic conditions such as ocean temperature, ocean currents and upwelling conditions, regional hydrological conditions, and many other uncontrollable variables (Crozier 2014). As a result, it may be desirable to monitor the physical effectiveness of mitigation work within the barrier reach, for example by directly evaluating hydraulic conditions before and after mitigation.

A hypothesis of success based on hydraulic conditions would compare physical emergence of barrier velocities at varying flows among pre- and post-mitigation conditions. For example, "At a given flow, Q_i, fewer/less spatially extensive/less intense velocity barrier patches emerge within the study reach after mitigation work (VB_{post}) compared to before mitigation work (VB_{pre}).

H4: For Q_i, $VB_{post} < VB_{pre}$

An associated null hypothesis would again posit no difference in the emergence of velocity barriers among varying flows, before and after mitigation work ($VB_{post} = VB_{pre}$)

H4₀: For Q_i, $VB_{post} = VB_{pre}$

Evidence supporting this null hypothesis, in the form of scatterplots describing the relationship between flow and the number of barrier patches in the study reach among pre- and post-mitigation migrants would look something like Figure 7. Here, the series overlap appreciably, indicating that the relationship between flow and passage does not differ meaningfully among the groups.

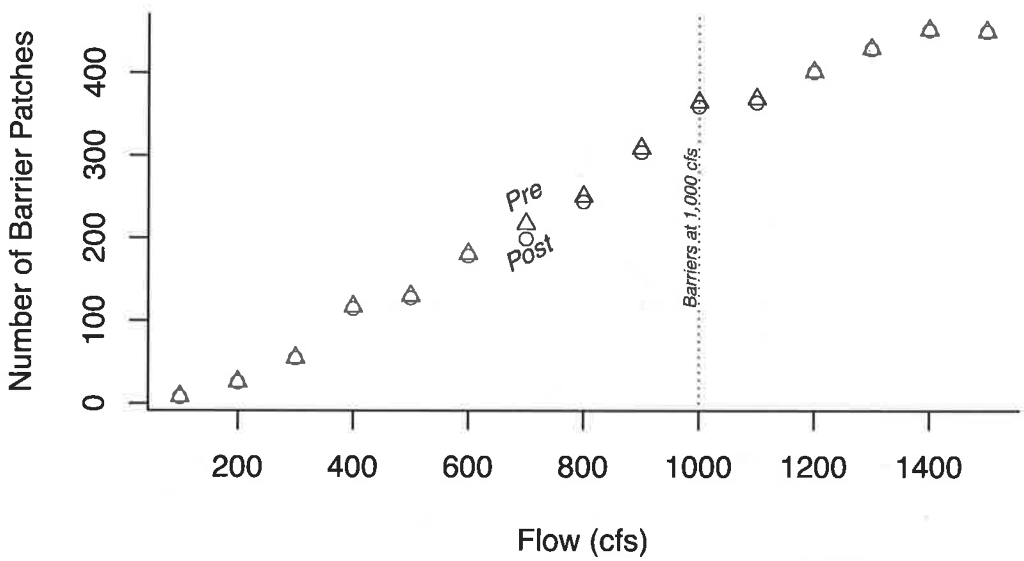

Figure 7. Simulated scatterplot comparing the emergence of barrier patches to flow, if mitigation did not lead to meaningful improvement.

Evidence refuting this null hypothesis of no difference in the relationship between flow and passage among the groups would look something like Figure 8. Here, the series relating flow and the number of barrier patches within the study reach appreciably differ in position. Importantly, these differences can be quantified and used to draw insights about the difference in the relationship between flow and the emergence of barriers before and after mitigation, in terms of both effect size (e.g., "an 85% reduction in the number of barriers at flow x") and confidence in the statistical result (e.g., "$p < 0.01$").

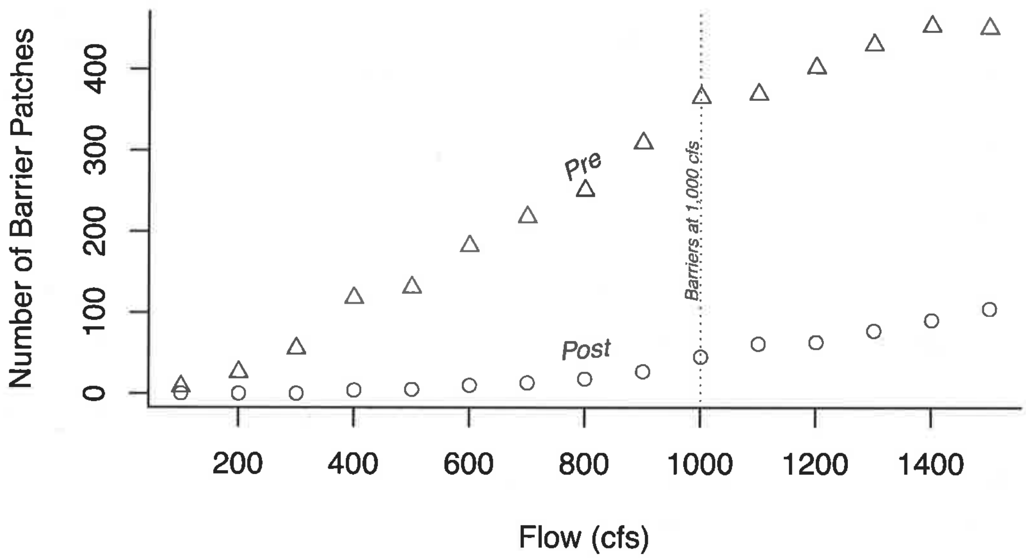

Figure 8. Simulated scatterplot comparing the emergence of barrier patches to flow, if mitigation led to meaningful improvement.

TRACKING TECHNOLOGIES

Initially, we reviewed a comprehensive suite of available biotelemetry and marking technologies, including acoustic telemetry, elastomer tags, fin marks, and others. Many of these exhibited one or more critical shortcomings, precluding them from further consideration. For example, acoustic telemetry is not ideal for tracking fish through the study area because of the need for acoustically quiet conditions, which are not present in the study reach. Similarly, elastomer tags and fin marks, while inexpensive, both require capturing fish below the study site, handling and marking at the time of the initial capture, then recapturing fish above the site at a sufficient efficiency to reliably infer that a lack of recapture meant a failure to pass.

After this initial review, we identified two technologies that are well suited to evaluate steelhead passage through the study reach. Above the others, radio telemetry (RT) and passive integrated transponders (PIT) stand out. Both of these technologies offer strengths and limitations. Here, we outline those qualities and provide insight into the logistics associated with installing, maintaining, and operating RT and PIT systems.

Radio Telemetry Summary

An RT system consists of the following three general components:

1) The transmitter (tag), which is inserted into the study species and emits a unique radio signal
2) The antenna system, which captures the signal emitted by the transmitter, then relays this signal via coaxial cable to a receiver
3) The radio receiver system, which decodes the signal emitted by the transmitter and detected and transmitted by the antenna system, then stores a record of this detection within a datalogger

Modularity of this system (i.e., separation of transmitter, antenna, and receiver) provides flexibility in the design of an RT array. For example, receivers can be paired to multiple antennas or otherwise configured to detect and compile transmitter detections across tunable spatial extents, depending on project and study goals.

RT systems excel in shallow (water depths <10m), swiftwater (lotic) environments, they are durable and otherwise suitable for long-term deployments, and the necessary equipment is generally located on dry land, facilitating relatively easy setup, maintenance and operation. Notable limitations of RT systems include a substantial decrease in detection efficiency at water depths >10 m, visibility of the antennas which may attract vandalism, receiver sensitivity to electromagnetic (EM) interference, tag collisions when many tags occupy the same space, and a system efficacy that depends on site geomorphology and station placement.

Passive Integrated Transponders Summary

The general components of a PIT system are similar to those of an RT system: tag, antenna, and receiver. The PIT tag is a radio frequency identification (RFID) device that is inserted into the study species, and which consists of an encapsulated microchip transponder (combined receiver and transmitter). When energized (e.g., by the field of an antenna), the PIT tag transmits a radio signal that communicates a unique alphanumeric identification code. A PIT antenna both energizes (charges) the tag and detects the signal. PIT antennas either alternate between charging and listening (half-duplex) or simultaneously charge and listen (full-duplex). Finally, a receiver and/or data logger (these are frequently combined) receives, decodes, and stores the information in the signal detected by the antenna.

One key difference between the two technologies is how the signal from the tag is relayed to the receiver. RT tags are battery powered and actively transmit a signal that can be detected by a receiver. Conversely, PIT tags lack an internal battery power—hence, passive. Instead, PIT tags rely on an electrified antenna (essentially a tuned electromagnet) as a power source to charge the microchip.

The lack of battery is PIT technology's greatest strength, allowing for small tags that operate for the entire life of the fish. Furthermore, PIT tags are relatively inexpensive allowing studies to tag greater numbers of fish. Limitations of PIT technology include high up-front cost (installation can be more than $100k for larger systems), small detection range (generally <1m), potential for tag collision, and a detection range that depends on tag orientation relative to the receiver.

RT-PIT Comparison

There are a number of important differences between RT and PIT systems that are worth considering for design of the current study. These include above described strengths and limitations associated with performance, environmental conditions, and logistics, as summarized below in Table 1.

Table 1. Strengths and limitations of RT and PIT biotelemetry systems.

	Technology	
	Radio Telemetry	*Passive Integrated Transponders*
Strengths	Shallow water (<10m)	Long life
	Effective under conditions of low conductivity	Inexpensive tags
	Suitable for long term deployment	Small tag size
	Relatively inexpensive install	
Limitations	Deep water (<10m)	Costly to install new sites
	Sensitive to interference	Short detection range (<1m)
	Vandalism	Difficult to maintain (underwater)
	Tag collision	Tag collision
	Effected by geomorphology, station placement	

One of the most important distinctions among PIT and RT technology is cost. Tagging and tracking project costs include expenses for tags, antennas, and receivers, along with operational costs associated with power, data, and personnel. To help quantify the relative costs for these technologies, CFS contacted RT and PIT vendors to develop estimates associated with setting up RT and PIT systems (Table 2).

Table 2. Cost estimates for crucial elements of RT, PIT and Data telemetry systems.

Technology	*Manufacturer*	*Unit*	*Cost (USD)*
RT	Sigma Eight	Receiver	$3k/ea
		Antenna Switch	$800/ea
		Antenna (Yagi)	$115/ea
		Tags	$175/ea
	Lotek	Receiver (switching built-in)	$3k - $6k
PIT	Westfork Environmental	Array with two dual-antenna sites	$123k - $150k
	Biomark	Tags	$2/ea
Data telemetry	Various	Remote satellite wifi hotspot	$2k - $5k
		Satellite network plan	$40/mo

Other Considerations

A key goal of this study is to increase the sample size of tagged fish returning to the study reach. One simple way to accomplish that is to tag as many fish as possible. PIT tags are less expensive than RT tags, so would seem to be the better solution, at first glance. Increasing the number of PIT tagged juveniles originating above the study reach may boost NPT's ability to subsequently identify (and then RT tag) adults at Lower Granite Dam (LGD) that are destined for the study reach (discussed in more detail below, in the *Sample Size* section). Importantly, this is a benefit of additional PIT tagging that would accrue even if a PIT array is not installed to track adult steelhead through the study reach. On the other hand, installing new PIT arrays through the study reach to use PIT detections to evaluate barrier passage would carry steep costs for initial setup and installation. However, as an added benefit of a PIT array in this area, PIT detections may prove useful for additional ongoing projects in the area (e.g., to help estimate outmigrating smolt abundance and survival).

NPT has expressed interest in exploring the potential to remotely access and download data captured by the RT and PIT arrays, using a data network. Both RT and PIT data can be remotely accessed via a LAN, cellular, or satellite network that communicates either directly with the receiver or with a PC that is connected to the receiver. Unfortunately, cellular network coverage is non-existent at the study site, and the nearest site that could offer a LAN (Idaho DOT maintenance building) is over a kilometer away from the proposed RT sites. An additional option would be to use a remote satellite data uplink that would include a satellite WiFi hotspot at each site and monthly subscription to network plan. Costs for this have not been explored at this time.

TRACKING ARRAY

Technology

In this section, we describe a suggested spatial layout for receivers associated with a hybrid RT-PIT array that brackets the study reach. It is our opinion that combining the two technologies may offer substantial benefits for the current project as well as for ongoing regional NPT efforts. For example, by implementing two tracking technologies, it would be possible to infer relative detection efficiencies associated with RT receiver sites. However, if one technology is going to be pursued, we strongly recommend RT. Thus, sample size estimates and analyses presented in subsequent sections presume an RT-array.

Fixed versus Remote Receivers

Previous NPT steelhead tracking projects in the SF Clearwater area have implemented both handheld (remote) and fixed RT receivers. Remote receivers offer flexibility and the ability to precisely position detections, but fixed receivers are preferred for their ability to autonomously and continuously scan for the presence of fish from a consistent location. With any receiver, failure to detect a fish does not necessarily imply absence of that fish: while a positive detection confirms presence, the lack of a detection does not always refute presence. Fish can evade detection for many reasons, including tag collisions from co-located transmitters, excessive background noise, tag failure, or receiver malfunction. However, by continuously monitoring for the presence of transmitters from a single location, fixed receiver arrays are the better option for inferring failure to pass, an important component of this study.

Array Layout

NPT staff have identified locations along the Idaho State Highway 14 (ID 14) corridor bracketing the study reach that are well-positioned to track adult steelhead upmigrating through the study reach. These include the "Velocity Barrier Bottom" and "Velocity Barrier Top" sites indicated in the map below (Figure 9).

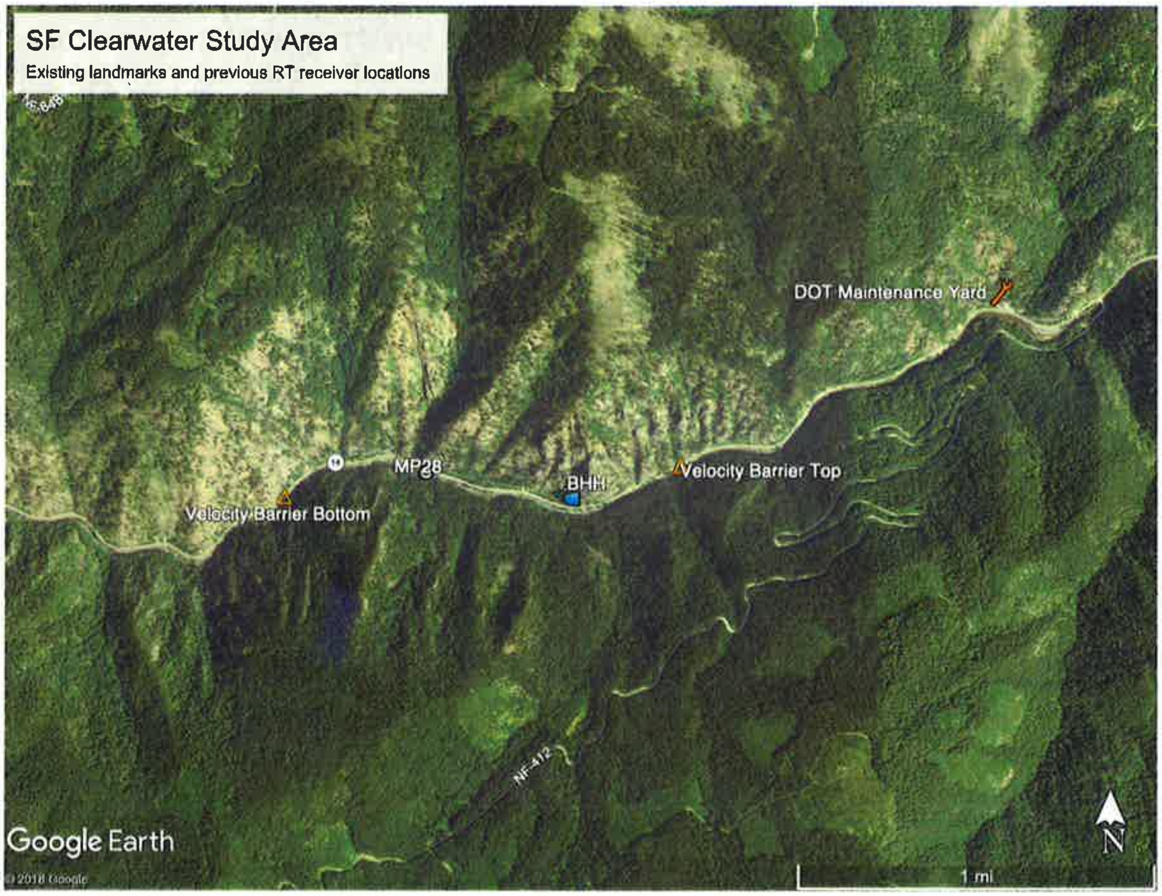

Figure 9. Map of study reach, including locations of riverside landmarks and previous RT receiver sites installed and operated by NPT. MP28 is the location of mile post 28 on Idaho State Highway 14. BHH is the Bird House Hole, a roadside landmark at the approximate location of the crux of the velocity barrier.

These previous RT receiver sites ("Velocity Barrier Bottom" and "Velocity Barrier Top") provide an excellent starting point for future fish tracking and evaluation of fish passage success through the study area. NPT staff has experience setting up receivers at these locations, these sites have demonstrated effectiveness for tracking fish movements through the study reach, and there are previous data collected from these locations that could inform estimates of pre-mitigation passage success. At a minimum, we suggest re-installing RT receivers at these previous locations.

In addition to these previously established locations, we suggest including two additional RT receiver sites within the study reach array (Figure 10). These sites would be positioned as close as feasible to the proposed mitigation site. The purpose of these additional sites would be to narrow the spatial extent over which passage success is evaluated, in order to more precisely quantify changes in the proportion of successful passage that result from future mitigation work, rather than differences in background environmental conditions.

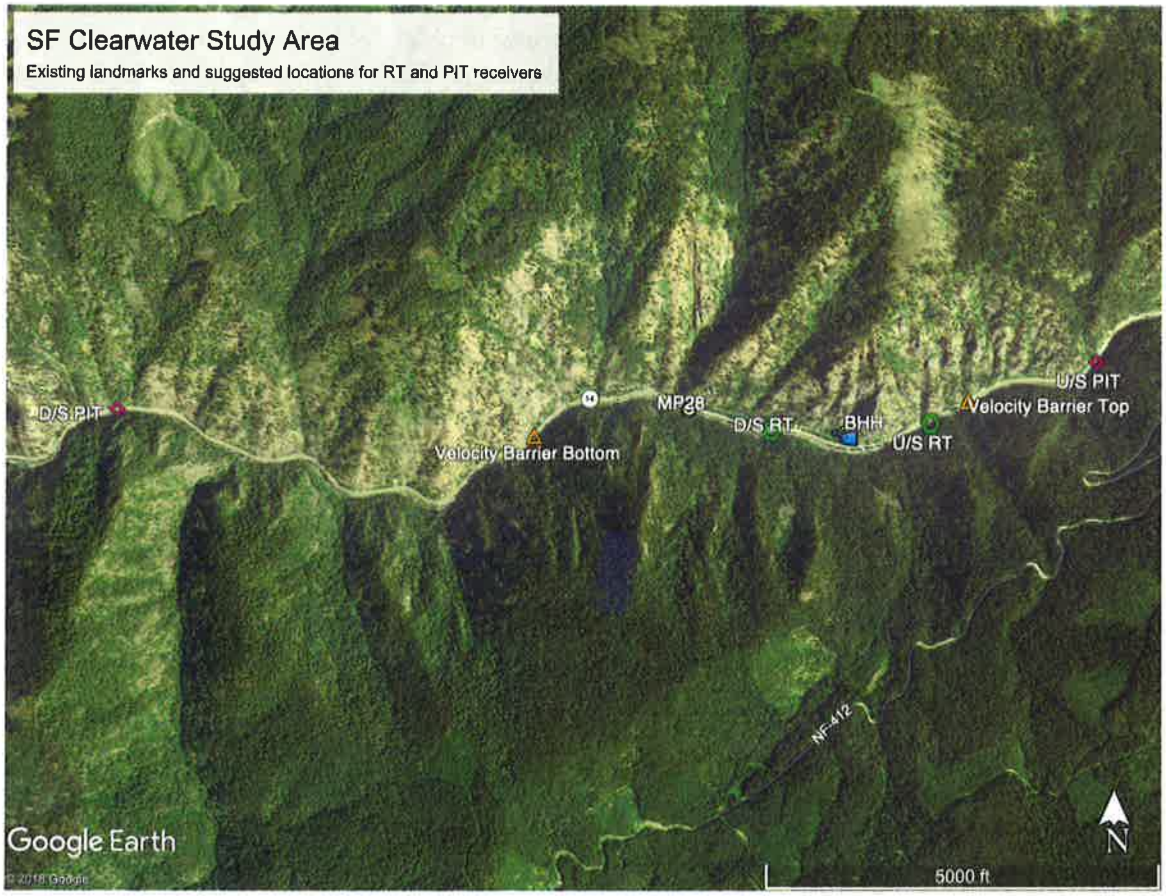

Figure 10. Map of study reach, including locations of riverside landmarks and previous RT receiver sites, along with suggested locations for additional RT receivers and PIT arrays. "D/S PIT" and "U/S PIT" are suggested locations for installing PIT antenna arrays downstream and upstream of the barrier, respectively. "D/S RT" and "U/S RT" are suggested locations for installing RT receivers.

Moreover, keeping the previous RT receivers and adding two new sites, rather than simply relocating receivers from the old sites to the new proposed sites, would provide multiple benefits. First, having four receivers provides a measure of redundancy. This redundancy would ensure that fish are not missed by a single receiver, improving overall project detection efficiency and bolstering sample size. Second, this array layout would provide more defensible grounds for including and excluding fish in calculations of the proportion of success. Given the limited spawning habitat between the previously established "Velocity Barrier Bottom" site and the suggested "D/S RT" site, it can reasonably be inferred that fish passing the Velocity Barrier

Bottom site are attempting to reach spawning grounds upstream of the barrier. This additional spatial resolution of behavioral insights would help support future decision making around the adaptive management question "Was enough work completed, or is more required to meet goals?" Third, this layout enables quantifying the duration of passage attempts over a larger spatial scale, which may more accurately reflect a biologically meaningful signal.

If including PIT tag technology is desired, we have also located two potential sites for PIT arrays. Siting these arrays is more challenging than siting RT receivers, for three important reasons. First, PIT antennas require fairly shallow water in order for antenna read range to effectively cover enough of the water column that detection efficiency is reasonably high. Second, PIT antennas perform best when located in low gradient reaches with an approximately flat bed. Steep gradients create fast and turbulent water, which impedes PIT detection efficiency. Non-uniform bed profiles impose logistical and technical challenges associated with constructing and maintaining the antennas. Third, compared to RT receivers, PIT receivers require a much greater source of power, to energize the antennas. In remote settings, PIT antenna power is frequently provided by solar arrays. Unfortunately, much of the study reach is shaded by steep ridges to the south during much of the steelhead migration season, meaning that solar capacity is limited and may be insufficient to consistently power PIT antenna arrays of the size necessary to span the river in the study reach. If solar is indeed insufficient, then it still may be possible to draw power from the electrical distribution lines along ID 14, although the cost for this likely would be substantial.

If PIT tracking technology is desired in the area of the study reach, we suggest constructing antenna arrays to bracket the study site RT array, at the approximate locations noted in Figure 10. These sites are among the few where installing a PIT array would be feasible. In addition, bracketing the RT sites would provide redundancy in the design, and would enable calculation of estimates of both RT and PIT site detection efficiency.

SAMPLE SIZE

The purpose of this section is to provide information that will help NPT in planning and executing a study to determine biological effectiveness of mitigation work. As an introduction, a few points are worth noting. The relationship between sample size and the ability to reliably detect an effect (statistical power) is well-known (Gotelli and Ellison 2012; Kruschke and Liddell 2018; Quinn and Keough 2002) and described by power curves (Carey and Keough 2002; Durlak 2009). These functions relate the number of study subjects (n) to a reliably detectible effect size, for previously determined acceptable levels of false positives (α), and false negatives (β) or statistical power ($1-\beta$). For the current study, a false positive would be an inappropriate rejection of the null hypothesis of no effect, when in fact there *was* no effect. This would lead to the inaccurate conclusion that mitigation improved some biological outcome. A false negative would be an in appropriate *failure* to reject the null hypothesis of no effect, when in fact there *was* an effect. In this case, the result would be the inaccurate conclusion that mitigation did *not* improve a biological outcome.

Previous discussions with both NPT and with NHC have indicated that it may be advantageous to pursue an iterative or adaptive approach to mitigation. In an adaptive framework, actions occur concurrent with evaluation, which informs decision-making regarding the necessity or value of additional action. For example, rather than removing a large section of rock, it may be desirable to remove a small section of rock, evaluate if and how fish passage has improved, then remove more rock only if necessary. For this approach, two things are necessary: a quantitative, explicit, and pre-defined goal of mitigation that functions as a "stopping rule," and the ability to accurately determine whether that goal has been met. Above, in the *Biological Effectiveness* section, we described two potential biological endpoints that could be evaluated as metrics of success (proportion and duration of successful passage at current barrier flows). We recommend that the goal(s) be described in terms of improvements for at least these biological endpoints. However, we also suggest that the mitigation goal in terms of biological effect should be determined by NPT staff, in cooperation with regional resource managers and other stakeholders.

For the remainder of this section, we present illustrative findings and recommendations assuming evaluation of the proportion of successful passage only. Power analyses for logistic models are more complex than for proportion comparisons, and regression designs are generally more powerful and thus likely to indicate fewer samples necessary to detect a given effect size (Cottingham et al. 2005). Thus, sample size recommendations and insights regarding statistical power presented here may be biased to be conservative, depending on analytical approach ultimately used (i.e., comparison of proportions of success among groups versus logistic modeling).

Logistics of Increasing Sample Size

If the positive biological effect of mitigation is constrained to mean an improvement in the proportion of successful fish passage, sample size is composed of the number of adult steelhead returning to the study reach that can be reasonably inferred to attempt to pass the location of the velocity barrier. Then, increasing sample size amounts to tagging and tracking more fish through the study reach. This can be accomplished through increased annual tagging efforts or by pooling samples across years, although care is required to ensure that pooling is conducted appropriately.

However, there are constraints on how many fish can be tagged in a given year, and on how many of those fish will ultimately be tracked through the study reach. According to NPT records, for the last five years adult steelhead returns to the SF Clearwater have been modest. Also, the logistics associated with capturing adult steelhead for tagging are non-trivial. This has led NPT to focus efforts to capture and tag fish at LGD, approximately 240 river kilometers (RKm) or 150 river miles (RM) downstream. Capturing and tagging fish at LGD requires positively identifying steelhead that originated above the study reach among all passing steelhead. A subset of these positively identified fish are then tagged. Because of this remote tagging location, there are substantial post-tagging losses that includes harvest, straying, and pre-spawn mortality between LGD and the study reach.

This means there are multiple factors constraining how many fish can be tagged in a given year, and only two or three opportunities within the current program for increasing the number of tagged fish. To start, adults are positively identified at LGD as a result of PIT tags that have been implanted in juvenile steelhead originating above the study reach during previous years. Since 2012, the number of juveniles available for tagging has averaged 160k (range = 118-350k; Table 3). In a given year, a small proportion of these fish are tagged (mean = 3%, range = 1.1-5.4%). After being tagged as juveniles, these steelhead must survive outmigration through the mainstem hydrosystem, successfully mature at sea during 2-3 ocean years while evading predation, then successfully evade both predation and harvest during re-entry to the freshwater system and their upriver spawning migration of more than 740 RKm (460 RM) to LGD. Since 2012, the smolt to adult return (SAR) ratio from tagging as juveniles until return to LGD has averaged 0.5% (range 0.2-0.6%). Since 2014, NPT has radio tagged approximately 50% of these positively identified adults passing LGD in a given year (range = 15-82%). Of those radio tagged fish, less than 30% typically make it back to the study reach (range = 11-40%).

Table 3. Previous tagging summary data provided by Peter Cleary (NPT).

Year	No. RT tagged Adults Returning to Reach	Proportion RT Tagged Adults Arriving at Reach	No. Adults RT Tagged at LGD	Proportion IDd Adults RT Tagged	No. Adults PIT IDd at LGD	SAR from Juvenile PIT Tagging to Adult at LGD	No. Juveniles PIT-Tagged	Proportion Juveniles PIT-tagged	No. Juveniles Available for PIT-Tagging
2012	NA	NA	NA	NA	NA	NA	9,782	2.8%	347,699
2013	NA	NA	NA	NA	NA	NA	4,144	3.1%	134,904
2014	2	40.0%	5	33.3%	15	0.2%	3,061	2.6%	118,053
2015	1	33.3%	3	15.0%	20	0.5%	4,357	3.6%	121,648
2016	4	33.3%	12	63.2%	19	0.6%	2,216	1.6%	134,353
2017	4	26.7%	15	62.5%	24	0.6%	1,771	1.1%	155,081
2018	1	11.1%	9	81.8%	11	0.5%	6,739	5.4%	123,816
2019							7,100	4.9%	145,264
Mean	***2***	***28.9%***	***9***	***51.2%***	***18***	***0.5%***	***4,896***	***3.1%***	***160,102***

To explore how increases in juvenile PIT tagging efforts and adult radio tagging efforts would be predicted to translate into increased numbers of adults that could be tracked using radio telemetry at the study site (what we will refer to hereafter as "sample size"), we constructed a simple model based on the data provided by NPT. This heuristic approach started with presumptions and assumptions based on the above described results provided by NPT, which included the following:

- The number of juveniles originating upstream of study reach that are available for PIT tagging averages ~160,000 annually.
- The proportion of these available juveniles that are PIT tagged averages ~3% annually.
- SAR from PIT tagging as juveniles to adult return to LGD averages 0.5% annually.
- The proportion of PIT identified adults at LGD that are radio tagged averages ~50% annually
- The proportion of radio tagged adults that successfully return to the study reach averages ~30% annually.
- Increases in both juvenile PIT tagging and adult radio tagging are feasible.
- NPT will track adults through the study reach using RT only.

Based on these data and assumptions, we then explored 33 hypothetical increased tagging scenarios. These included combinations of increases in juvenile PIT tagging and adult radio tagging efforts intended to extend across the range of feasible efforts beginning with the *status quo* and extending to include "massive effort" scenarios at the upper end.

Besides the *status quo* effort, ten iterations of increased juvenile PIT tagging were explored in increments of 1,000 additional tags. This led to a total of 11 scenarios, ranging from the previous mean of approximately 5,000 juveniles up to an increased effort amounting to 10,000 additional juveniles being PIT tagged, for approximately 15,000 total juveniles PIT tagged. Although costly and logistically challenging, it bears noting that this tagging effort would still amount to less than 10% of available juveniles in an average year.

Next, across this hypothetical range of increased juvenile PIT tagging efforts, we layered additional adult radio tagging efforts. Here, we chose three iterations, beginning with the *status quo* effort of approximately 50% of adults identified at LGD being radio tagged and increasing by 25% increments through 100% radio tagging of adults that were PIT identified at LGD.

The results from this exercise are presented below (Table 4) and suggest at least two important insights. First, when relying on the current approach of tagging adults at LGD that have been PIT identified as originating from above the study reach, a substantial increase in both juvenile PIT tagging and adult radio tagging is unlikely to yield a sample size greater than 20 adults in a given year. Second, the more efficient means of increasing sample size appears to be simply radio tagging a greater proportion of those adults that are positively identified at LGD. To explore this second point further, we next estimated material costs (for tags only) associated with increased tagging efforts, then quantified these costs as a function of the benefit, which we expressed in terms of predicted increases in sample size.

Table 4. Heuristic model exploring the effect of different combinations of increased juvenile PIT tagging and adult radio tagging on the predicted number of trackable adults returning to the study reach. Color ramps summarized as follows: Increasing efforts range blue (lowest) to red (highest) across multiple parameters. Gray cells show means from NPT data. Increasing sample size ranges yellow (lowest) to green (highest).

No. Radio Tagged Adults Returning to Reach	Proportion Radio Tagged Adults Returning to Reach	No. Adults Radio Tagged at LGD	Proportion PIT IDd Adults Radio Tagged	No. Adults PIT IDd at LGD	SAR From Juvenile PIT Tagging To Adult at LGD	No. Juveniles PIT-Tagged	Proportion Juveniles PIT-tagged	No. Juveniles Available for PIT-Tagging
				PIT Tag Average Number of Previous Juveniles				
3	28.9%	11	50.0%	23	0.5%	4,896	3.1%	160,102
5	28.9%	17	75.0%	23	0.5%	4,896	3.1%	160,102
7	28.9%	23	100.0%	23	0.5%	4,896	3.1%	160,102
				PIT Tag 1,000 Additional Juveniles				
4	28.9%	14	50.0%	27	0.5%	5,896	3.7%	160,102
6	28.9%	20	75.0%	27	0.5%	5,896	3.7%	160,102
8	28.9%	27	100.0%	27	0.5%	5,896	3.7%	160,102
				PIT Tag 2,000 Additional Juveniles				
5	28.9%	16	50.0%	32	0.5%	6,896	4.3%	160,102
7	28.9%	24	75.0%	32	0.5%	6,896	4.3%	160,102
9	28.9%	32	100.0%	32	0.5%	6,896	4.3%	160,102
				PIT Tag 3,000 Additional Juveniles				
5	28.9%	18	50.0%	36	0.5%	7,896	4.9%	160,102
8	28.9%	27	75.0%	36	0.5%	7,896	4.9%	160,102
11	28.9%	36	100.0%	36	0.5%	7,896	4.9%	160,102
				PIT Tag 4,000 Additional Juveniles				
6	28.9%	20	50.0%	41	0.5%	8,896	5.6%	160,102
9	28.9%	31	75.0%	41	0.5%	8,896	5.6%	160,102
12	28.9%	41	100.0%	41	0.5%	8,896	5.6%	160,102
				PIT Tag 5,000 Additional Juveniles				
7	28.9%	23	50.0%	46	0.5%	9,896	6.2%	160,102
10	28.9%	34	75.0%	46	0.5%	9,896	6.2%	160,102
13	28.9%	46	100.0%	46	0.5%	9,896	6.2%	160,102
				PIT Tag 6,000 Additional Juveniles				
7	28.9%	25	50.0%	50	0.5%	10,896	6.8%	160,102
11	28.9%	38	75.0%	50	0.5%	10,896	6.8%	160,102
15	28.9%	50	100.0%	50	0.5%	10,896	6.8%	160,102
				PIT Tag 7,000 Additional Juveniles				
8	28.9%	27	50.0%	55	0.5%	11,896	7.4%	160,102
12	28.9%	41	75.0%	55	0.5%	11,896	7.4%	160,102
16	28.9%	55	100.0%	55	0.5%	11,896	7.4%	160,102
				PIT Tag 8,000 Additional Juveniles				
9	28.9%	30	50.0%	59	0.5%	12,896	8.1%	160,102
13	28.9%	45	75.0%	59	0.5%	12,896	8.1%	160,102
17	28.9%	59	100.0%	59	0.5%	12,896	8.1%	160,102
				PIT Tag 9,000 Additional Juveniles				
9	28.9%	32	50.0%	64	0.5%	13,896	8.7%	160,102
14	28.9%	48	75.0%	64	0.5%	13,896	8.7%	160,102
18	28.9%	64	100.0%	64	0.5%	13,896	8.7%	160,102
				PIT Tag 10,000 Additional Juveniles				
10	28.9%	34	50.0%	69	0.5%	14,896	9.3%	160,102
15	28.9%	51	75.0%	69	0.5%	14,896	9.3%	160,102
20	28.9%	69	100.0%	69	0.5%	14,896	9.3%	160,102

First, we determined approximate costs for PIT and radio tags, based on quotes from technology equipment providers (Table 5). Clearly, PIT tags are less expensive, with a single PIT tag costing nearly 100x less than a single radio tag. However, if relying on PIT tags solely as a means for identifying adults to radio tag (i.e., rather than using PIT technology for directly tracking adults through the study reach), the optimal choice for maximizing sample size was not immediately apparent.

Table 5. Approximate costs for Radio tags (Lotek) and PIT tags (BioMark), as of 2019.

Tag Costs	
Radio Tag	*PIT Tag*
$150.00	$1.85

To address this optimization question, we next compared the financial investments associated with increased efforts representing either a greater proportion of PIT identified adults that are radio tagged or a greater proportion of juveniles that are PIT tagged (Table 6). The results of this effort may appear obvious retrospectively, but hopefully they help illustrate an important point: Compared to PIT tagging more juveniles and then radio tagging a consistent proportion of PIT identified adults at LGD, radio tagging a greater proportion of those PIT identified adults is the simpler and less expensive means for increasing sample size. Again, this may appear obvious when presented this way. However, the real utility of this knowledge becomes apparent when evaluating the economics of the different combinations of increased efforts for juvenile PIT tagging and adult radio tagging presented above in Table 4.

Table 6. Predicted increase in samples (adults that can be tracked through the study reach using radio telemetry) as a function of increases in juvenile PIT tagging and/or adult radio tagging, relative to tag costs, expressed as cost (USD) per sample.

Additional Samples	***Additional Radio Tags***	***Radio Tag Cost***	***Additional PIT tags***	***PIT Tag Cost***	***Cost per Sample***
RT Tagging Additional ID'd Fish, No Additional PIT Tags					
0	1	$150.00	0	$0.00	$519.23
3	10	$1,500.00	0	$0.00	$519.23
6	20	$3,000.00	0	$0.00	$519.23
PIT Tagging Additional Fish, Maintaining Current Plan of RT Tagging Fixed Proportion of ID'd Fish					
1	5	$765.01	1,000	$1,850.00	$1,774.87
4	15	$2,179.48	5,000	$9,250.00	$2,722.91
8	26	$3,947.56	10,000	$18,500.00	$2,952.57

To assist in deciding among combinations of increased juvenile PIT tagging and adult radio tagging, we next expanded upon our simple heuristic model that predicts sample size as a function of PIT tagging and radio tagging efforts. First, we included a parameter describing the *increase* in sample size, which can be thought of as the net benefit of a given scenario. Second, we estimated costs for additional PIT tags and radio tags, using the information in Table 5, above. Tabulated results from this effort are presented in Table 7.

Table 7. Cost per additional sample associated with combinations of increased juvenile PIT tagging and adult radio tagging efforts. Color ramps are similar to those used above, in Table 4: Increasing costs range blue (lowest) to red (highest). Increasing sample sizes range yellow (lowest) to green (highest).

Additional Samples	Additional Radio Tags	Radio Tag Cost	Additional PIT tags	PIT Tag Cost	Total Tag Cost	Cost per Additional Sample
1	2	$372	0	$0	$372	$433
2	8	$1,218	0	$0	$1,218	$490
4	14	$2,064	0	$0	$2,064	$501
2	5	$718	1,000	$1,850	$2,568	$1,684
3	12	$1,736	1,000	$1,850	$3,586	$1,029
5	18	$2,755	1,000	$1,850	$4,605	$845
2	7	$1,063	2,000	$3,700	$4,763	$2,175
4	15	$2,255	2,000	$3,700	$5,955	$1,328
7	23	$3,446	2,000	$3,700	$7,146	$1,054
3	9	$1,409	3,000	$5,550	$6,959	$2,437
5	18	$2,773	3,000	$5,550	$8,323	$1,518
8	28	$4,138	3,000	$5,550	$9,688	$1,194
4	12	$1,754	4,000	$7,400	$9,154	$2,600
6	22	$3,292	4,000	$7,400	$10,692	$1,650
9	32	$4,829	4,000	$7,400	$12,229	$1,295
4	14	$2,100	5,000	$9,250	$11,350	$2,711
7	25	$3,810	5,000	$9,250	$13,060	$1,746
11	37	$5,520	5,000	$9,250	$14,770	$1,371
5	16	$2,446	6,000	$11,100	$13,546	$2,792
8	29	$4,328	6,000	$11,100	$15,428	$1,820
12	41	$6,211	6,000	$11,100	$17,311	$1,430
6	19	$2,791	7,000	$12,950	$15,741	$2,853
9	32	$4,847	7,000	$12,950	$17,797	$1,878
13	46	$6,902	7,000	$12,950	$19,852	$1,478
6	21	$3,137	8,000	$14,800	$17,937	$2,901
10	36	$5,365	8,000	$14,800	$20,165	$1,925
15	51	$7,594	8,000	$14,800	$22,394	$1,516
7	23	$3,482	9,000	$16,650	$20,132	$2,939
11	39	$5,884	9,000	$16,650	$22,534	$1,964
16	55	$8,285	9,000	$16,650	$24,935	$1,549
8	26	$3,828	10,000	$18,500	$22,328	$2,971
12	43	$6,402	10,000	$18,500	$24,902	$1,997
17	60	$8,976	10,000	$18,500	$27,476	$1,576

When summarized and presented this way, a hierarchical strategy for increasing sample size became apparent. First, as many of the PIT identified adults returning to LGD should be radio tagged as possible. This realization of radio tagging potential can be thought of as recouping efforts and costs associated with PIT tagging juveniles. This cost recovery can be seen in Table 7 as a tendency of the cost per sample to decrease within each PIT tagging increment, as a function of increasing radio tagging effort.

Thus, while costs may go up by radio tagging more adults, value goes up at a higher rate, reducing the cost per sample associated with a given juvenile PIT tagging effort. However, this additional adult radio tagging alone is unlikely to provide sufficient sample size to enable analyses generating confident and conclusive results. It is predicted that a robust sample size can only be obtained by pairing an increase in juvenile PIT tagging efforts with this increased adult radio tagging effort. To quantify relative effect size and confidence associated with various increases in effort, we next explored statistical power as a function of sample size.

Statistical Power as a Function of Sample Size

Background

The purpose of this section is to translate the results presented above for estimating how increased tagging efforts are predicted to yield increases in sample size, into insights regarding statistical power associated with a range of sample sizes. As described above, statistical power represents the ability to detect an effect of a given size, and power is generally interpreted as the probability of appropriately rejecting a null hypothesis when there truly was an effect. For this section, we are assuming that the effect being evaluated is biological success, quantified as proportion of successful passage. Moreover, as described above, these analyses are based on a simple one-sided proportions test comparing passage success before and after mitigation, analyzing for an increase in passage success. Power curves are an effective way to visualize the relationship between sample size and power, for fixed values of acceptable α (risk of a false positive) and effect size.

Statistical convention prescribes a maximum acceptable $\alpha = 0.05$, corresponding to a 1/20 chance of falsely concluding that there was an effect, when there truly was no effect, and instead the results reflect random variation within otherwise equal groups. For the purposes of these evaluations, we hew to the convention of setting maximum acceptable $\alpha = 0.05$.

Effect size can be expressed as raw differences between treatment groups (e.g., "a 20 percentage point increase in successful passage"), proportional differences relative to a control or standard group (e.g. "a 100% increase (doubling) of successful passage"), or as a transformed and unitless value that ranges from 0 to 1 (e.g., "an effect size of $h = 0.5$").

Of these approaches, the first and second both appear at first glance to be more intuitive than the third. However, when effect size is expressed on a unit scale, interpretation and comparisons with other studies may be hampered by the dependence of effect size upon statistical parameters of the control group. While it suffers from lack of an immediately intuitive interpretation, the primary advantages of a unitless effect size such as Cohen's h (Cohen 1988), are twofold. First, unitless effect sizes are comparable across otherwise disparate studies. This eases comparison and contextualization within the literature. Second, unitless effect sizes are independent of control group parameters, and thus do not depend upon starting conditions.

For these reasons, we present results below in terms of Cohen's unitless h parameter. However, we do also explore absolute effect sizes, as a heuristic illustration of what can be expected given a range of sample sizes.

Power Analyses

Cohen (1988) initially suggested that studies should aim for power of at least 0.80, which has become the widely accepted convention within statistics. Further, Cohen recommends the following values for *h* when evaluating differences in proportions:

- Small effect $h = 0.2$
  - Equivalent to a raw difference in proportions of 0.05 – 0.10 (5-10 percentage points), depending on the group 1 proportion
- Medium effect $h = 0.5$
  - Equivalent to a raw difference in proportions of 0.20 – 0.25 (10-25 percentage points), depending on the group 1 proportion
- Large effect: $h = 0.8$
  - Equivalent to a raw difference in proportions of 0.35 – 0.39 (35-39 percentage points), depending on the group 1 proportion

As a first step in understanding the value associated with increasing the numbers of trackable adults through the study reach, we plotted the relationship between sample size and the ability to reliably detect an effect (statistical power) for the three above described levels of effect size, $h = 0.2$, $h = 0.5$, and $h = 0.8$ (Figure 11). For each of these relationships between sample size and power, we then determined the minimum sample size required to achieve each of three potential powers, 0.7, 0.8, and 0.9. These values of 0.7 and 0.9 were designed to bracket the generally accepted desirable power of 0.8, providing a slightly less and slightly more powerful option, respectively.

When considered within the context of the heuristic model predictions presented in the *Logistics of Increasing Sample Size* section, above, which indicated that even if 10,000 additional juveniles were PIT tagged, and 69 adults were then radio tagged, sample size within the study reach is not predicted to exceed 20 fish, inspection of the plots paneled within Figure 11 reveals some important insights. First, statistical power is limited for a proportions test based on logistically feasible sample sizes. It does not appear likely that a small effect size will be detectible: even at power = 0.7, detecting a 5-10 percentage point difference in proportion ($h = 0.2$) would require $n > 200$.

This is not to say that a study monitoring effectiveness of mitigation on improving fish passage is worthless. Even if power to detect a small effect is low, it does seem possible to achieve a sample size sufficient to detect a large effect ($h = 0.8$)—equivalent to an increase in approximately 35-39 percentage points—with reasonable power. A sample size $n = 15$ would provide power of 0.7 for a large effect; $n = 20$ would provide power of 0.8 for a large effect.

In addition, there are numerous study design and analysis strategies that could be adopted to overcome this power limitation, including lumping fish across years within a single group, leveraging previous data in the form of Bayesian prior belief distributions, or adopting an alternative method of capturing and tagging fish, closer to the study reach. In this way, it may be possible to arrive at sample sizes that would enable detection of a medium effect ($h = 0.5$)—equivalent to an increase in approximately 20-25 percentage points—with reasonable power. A sample size of 38 fish would provide a power of 0.7 for a medium effect size, while 50 fish would provide a power of 0.8 for a medium effect size.

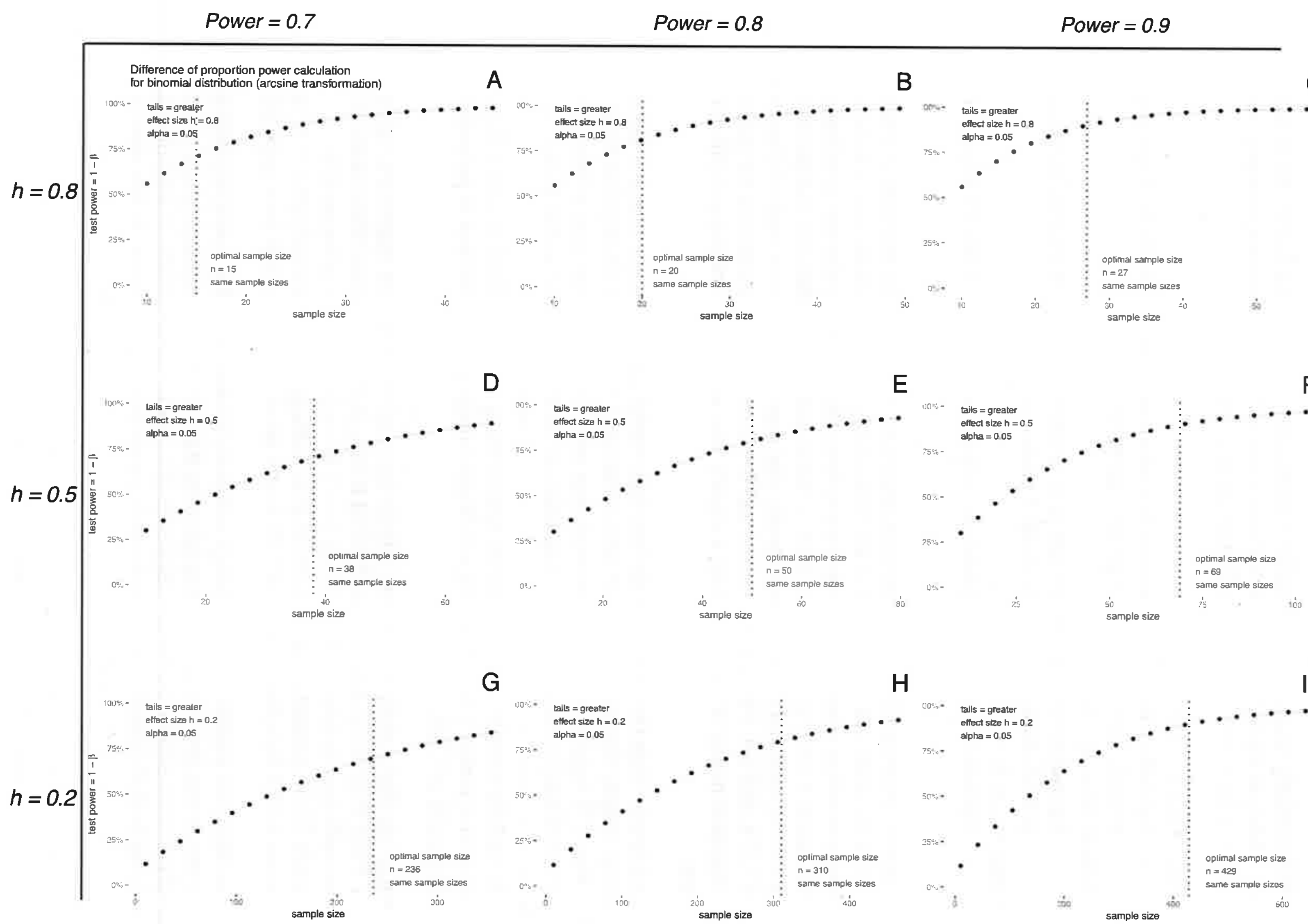

Figure 11. Plots depicting relationship between sample size (x-axis of each plot) and power (y-axis of each plot) across three effect sizes (rows). Columns indicate power used to determine optimal sample size. Figures within groups A-C, D-F, and G-I are identical, just rescaled along x-axis.

Detectible Effect Size as a Function of Sample Size

Next, the value of increasing sample sizes was contextualized and evaluated further, by plotting the relationship between sample size and detectible effect size, for the three above described power levels, at $\alpha = 0.05$ (Figure 12). This plot depicts the initially rapid, then increasingly diminishing returns associated with increasing sample size. This can be visualized as the flatting of each curve to the right of inflection point optima, beyond which increasingly more samples are required to reliably detect smaller effects.

From this plot, it appears that the value of additional samples beyond 40 rapidly decreases, and a relatively massive (and unrealistic) sample size is necessary to detect small effects. However, this plot also reinforces the notion, discernable from Figure 11, that large and medium effects can realistically be detected.

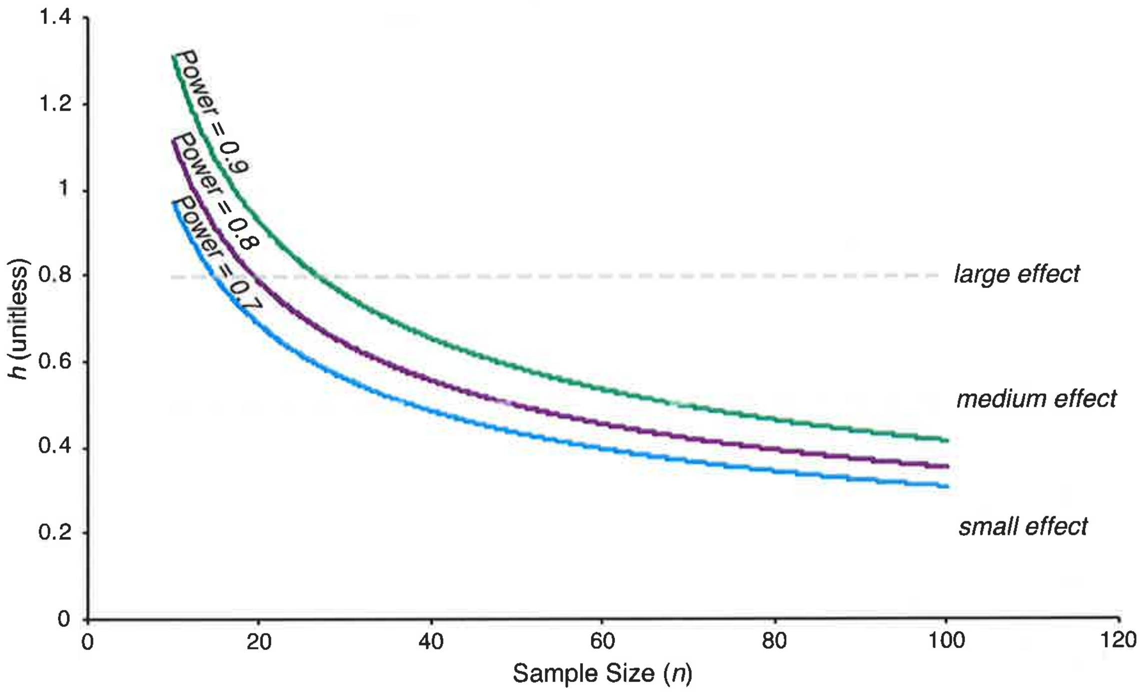

Figure 12. A large effect ($h = 0.8$) can be detected with $n = 15$ samples at power = 0.7, $n = 20$ samples at power = 0.8, or $n = 27$ samples at power = 0.9. A medium effect ($h = 0.5$) can be detected with $n = 38$ samples at power = 0.7, $n = 50$ samples at power = 0.8, or $n = 69$ samples at power = 0.9. A small effect ($h = 0.2$) can be detected with $n = 236$ samples at power = 0.7, $n = 310$ samples at power = 0.8, or $n = 429$ samples at power = 0.9.

Finally, we further contextualized the results of our power analysis by grounding these findings in terms of hypothetical scenarios for pre-mitigation passage success. We chose three realistic proportions of successful passage for the pre-mitigation group, then evaluated how increasing sample sizes would equate to an ability to detect ever smaller relative increases in success, at power = 0.8. Here, we term the pre-mitigation proportion of success Φ_1 and the post-mitigation proportion of success Φ_2. The difference between Φ_1 and Φ_1 then represents the raw increase in success, in percentage points. We then relativized this difference by dividing this difference by the starting success rate (Φ_1) to arrive at a relative proportional increase. Scaled thus, these values are easily interpretable as what most people think of (or mean) when they say, "percent change," i.e., a 100% change amounts to a doubling of passage success. We then plotted this detectible percent change against sample size at power = 0.8 for three different Φ_1 values (Figure 13).

Presented in this way, the above rationale for selecting a unitless effect size becomes more easily understood. If Φ_1 is relatively low, e.g., 0.4, and 20 samples are utilized, then the effect size must be a nearly 100% increase to be detectible at power = 0.8. However, if Φ_1 is 0.8, then 20 samples would enable detection of a much smaller absolute effect size (a 25% increase) at power = 0.8.

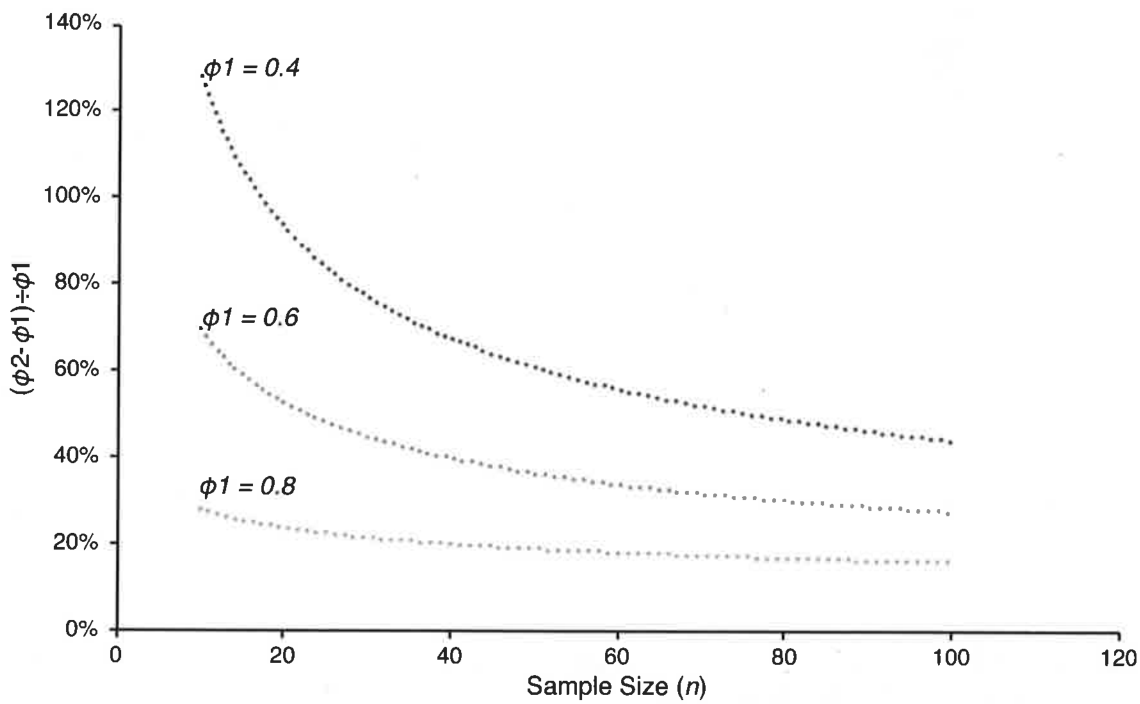

Figure 13. Percent change that is reliably detectible at power = 0.8 for three different pre-mitigation proportions of success (Φ_1). As sample size increases, ever smaller effect sizes can be detected. However, these proportional effect sizes differ depending on the value of Φ_1. When $\Phi_1 = 0.8$, 20 samples are sufficient to detect an approximately 25% increase in passage success. When $\Phi_1 = 0.4$, a 20-sample design is much weaker, and analyses using 20 samples can detect only effect sizes greater than or equal to a 94% increase. Note that all three curves markedly flatten above $n = 40$.

Additional Design Considerations

Finally, in order to optimally design and execute a study meeting the above described goals, two additional considerations warrant mention: how many years should be sampled before and after mitigation work, and how best to tag fish so as to ensure adequate representation of passage attempts through the study reach over the duration of the run?

The first is determining how many years to sample pre- and post-mitigation. The optimal decision of how many years to sample would account for both adequately sampling passage across a representative range of environmental conditions, and for collecting a sufficient sample size to enable detection of a reasonably sized effect. Sampling across a range of environmental conditions is important because, as discussed above in the section on defining success in biological terms, successful passage is the culmination of many factors beyond physical channel dimensions within the study reach. Especially if evaluating passage as a simple comparison of the proportion of success among pre- and post-mitigation groups of fish, it will be important to ensure that confounding differences in environmental conditions beyond those affected by mitigation work do not obscure—or create the false impression of—a real biological effect. One of the most important of these uncontrollable factors is regional hydrologic conditions (i.e., wet versus dry water years), which directly influence study reach hydraulics. This concern would be reduced though not eliminated by adopting a modeling approach and including model terms describing ancillary covariates such as hydrologic water year category or cumulative snowfall. Still, we recommend sampling for at least two years before and after mitigation work, and ideally for three years before and after, to ensure that passage success is evaluated across a range of realistic environmental conditions.

The second is determining how best to tag fish to provide adequate representation across the run. Similar to the above described inter-annual variation in environmental conditions and associated passage success, there is also substantial intra-annual variation in flow and success of passage through the study reach. It is not possible to guarantee that the group of fish tagged in each year arrive at and then attempt passage across the entire migration window. Still, tagging as many fish as possible at LGD (or closer to the study reach), across the widest possible time window will improve the chances of adequate representation of passage attempts through the study reach across the migration window. Previous tagging efforts have provided reasonable coverage of the spawning migration, and there is considerable uncertainty regarding this factor, so this concern is more of a caveat than a recommendation.

CONCLUSIONS AND RECOMMENDATIONS

The exercises, simulations, and analyses presented above lead us to draw the following conclusions, regarding the design, execution, and subsequent analysis of a study to monitor the effectiveness of proposed work to mitigate the velocity barrier(s) that currently emerge within the study reach at moderate to high flows.

- We recommend evaluating project success by focusing on improvements to fish passage.
- We recommend analyzing study results by constructing a series of logistic regression models, then comparing model parameters among model runs for pre- and post-mitigation fish.
- Bayesian approaches may prove useful (or necessary) for combining and optimally leveraging multiple years of data to maximize statistical power from limited sample sizes.
- It appears feasible to design a fish passage study through the study reach that would rely primarily on radio telemetry (RT), leveraging existing equipment, experience, and infrastructure that NPT has in place.
- If desired, it may be possible to bolster sample sizes by using PIT technology; however, installation, operation, and maintenance of PIT arrays would represent a relatively large capital investment.
- We recommend installing at least four RT receiver sites:
  - Two sites at the previously established "Velocity Barrier Bottom" and "Velocity Barrier Top" locations
  - Two new sites as close as possible to the location of proposed mitigation work, bracketing the work project site.
- If PIT arrays are desired, we have located potential sites that appear to be logistically appropriate; however, sourcing power to these locations may prove challenging.
- From our power analysis and sample size exercises, it appears that statistical power will be substantially limited without creative alternatives to bolster sample size and analyze data.
- In the year prior to study initiation, we recommend increasing radio tagging efforts at LGD so that as many returning adults that can be identified as originating from above the barrier are tagged as is feasible.
- In addition, to bolster sample size, we suggest PIT tagging more juveniles above the barrier as soon as feasible so that more adults can subsequently be identified at LGD.
- Without tracking at least 20 fish per year through the study reach during potential barrier flows, reliably detecting even medium effect sizes may prove challenging.
- We recommend conducting a study for at least two—and ideally three—years before and after mitigation work, that includes tagging fish across as wide a time window as possible.

REFERENCES

Brownscombe, J. W., and coauthors. 2017. Ecology of exercise in wild fish: Integrating concepts of individual physiological capacity, behavior, and fitness through diverse case studies. Integrative and Comparative Biology 57(2):281-292.

Caldwell, L. K., A. L. Pierce, and J. J. Nagler. 2013. Metabolic endocrine factors involved in spawning recovery and rematuration of iteroparous female rainbow trout (*Oncorhynchus mykiss*). General and Comparative Endocrinology 194:124-132.

Carey, J. M., and M. J. Keough. 2002. The variability of estimates of variance, and its effect on power analysis in monitoring design. Environmental Monitoring and Assessment 74(3):225-241.

Caudill, C. C., and coauthors. 2007. Slow dam passage in adult Columbia River salmonids associated with unsuccessful migration: Delayed negative effects of passage obstacles or condition-dependent mortality? Canadian Journal of Fisheries and Aquatic Sciences 64(7):979-995.

Cohen, J. 1988. Statistical Power Analysis for the Behavioral Sciences, Second edition. Lawrence Erlbaum Associates, New York, NY.

Cottingham, K. L., J. T. Lennon, and B. L. Brown. 2005. Knowing when to draw the line: Designing more informative ecological experiments. Frontiers in Ecology and the Environment 3(3):145-152.

Crozier, L. 2014. Impacts of Climate Change on Salmon of the Pacific Northwest: A review of the scientific literature published in 2014. Fish Ecology Division, Northwest Fisheries Science Center, National Marine Fisheries Service, NOAA, Seattle, WA.

Durlak, J. A. 2009. How to select, calculate, and interpret effect sizes. Journal of Pediatric Psychology 34(9):917-928.

Gotelli, N. J., and A. M. Ellison. 2012. A Primer of Ecological Statistics, Second Edition edition. Sinauer Associates, Sunderland, MA.

Jenkins, L. E., and coauthors. 2019. Elevated plasma triglycerides and growth rate are early indicators of reproductive status in post-spawning female steelhead trout (*Oncorhynchus mykiss*). Conservation Physiology 7(1).

Keefer, M. L., C. A. Peery, T. C. Bjornn, M. A. Jepson, and L. C. Stuehrenberg. 2004. Hydrosystem, dam, and reservoir passage rates of adult Chinook salmon and steelhead in the Columbia and Snake Rivers. Transactions of the American Fisheries Society 133(6):1413-1439.

Kruschke, J. K., and T. M. Liddell. 2018. The Bayesian new statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. Psychonomic Bulletin & Review 25(1):178-206.

Naughton, G. P., and coauthors. 2005. Late-season mortality during migration of radio-tagged adult sockeye salmon (*Oncorhynchus nerka*) in the Columbia River. Canadian Journal of Fisheries and Aquatic Sciences 62(1):30-47.

PacifiCorp, and Cowlitz County PUD No. 1. 2010. Aquatic Monitoring and Evaluation Plan for the Lewis River – First Revision. *Prepared by*: PacifiCorp and Public Utility District No. 1 of Cowlitz County.

PacifiCorp, and Cowlitz County PUD No. 1. 2016. Lewis River DRAFT Aquatic Monitoring and Evaluation Plan – First Revision - December 19, 2016. PacifiCorp Energy and Public Utility District No. 1 of Cowlitz County.

Quinn, G. P., and M. J. Keough. 2002. Experimental Design and Data Analysis for Biologists. Cambridge University Press, New York, NY.

Schneider, J. E. 2004. Energy balance and reproduction. Physiology & Behavior 81(2):289-317.

Timm, R., and coauthors. 2019. Drones, hydraulics, and climate change: Inferring barriers to steelhead spawning migration. WIREs Water ACCEPTED MANUSCRIPT(ONLINE ONLY).

Timm, R., and coauthors. 2017. South Fork Clearwater River MP 28 Hypothesized Velocity Barrier: Final Report. *Prepared by*: Cramer Fish Sciences & Northwest Hydraulic Consultants. *Prepared for*: Mark Johnson, Nez Perce Tribe.

Timm, R., and coauthors. 2018. South Fork Clearwater River MP 28 Hypothesized Velocity Barrier – Phase II. *Prepared by*: Cramer Fish Sciences (Portland, OR) & Northwest Hydraulic Consultants (Bellingham, WA). *Prepared for*: Mark Johnson, Nez Perce Tribe (Lapwai, ID).