REVIEW OF TRANCHE 2 MATERIAL SUBMITTED BY THE VANCOUVER-FRASER PORT AUTHORITY IN RESPECT OF ROBERTS BANK TERMINAL 2

The Roberts Bank Terminal 2 Independent Scientific Body (ISB)

April 2025

INDEPENDENT SCIENTIFIC BODY MEMBERS

Chair: Mona Nemer, Chief Science Advisor of Canada

David M. Paterson, Scottish Oceans Institute, School of Biology, University of St Andrews, United Kingdom.

Hannah S. Wauchope, School of GeoSciences, University of Edinburgh, Edinburgh, United Kingdom.

Margaret Rubega, Department of Ecology and Evolutionary Biology, University of Connecticut, U.S.A.

Kelly Munkittrick, Department of Biological Sciences, University of Calgary, Canada.

Jean-Michel Weber, Professor Emeritus, Department of Biology, University of Ottawa, Canada

OFFICE OF THE CHIEF SCIENCE ADVISOR STAFF

C. Scott Findlay, Researcher in Residence

1. **BACKGROUND**

The Vancouver Fraser Port Authority (VFPA) has proposed the construction and operation of a new three-berth marine container terminal located at Roberts Bank in Delta, British Columbia (the Roberts Bank Terminal 2 (RBT2) project). Among its findings, the Review Panel struck for the RBT2 project was not able to conclude with certainty whether the project would result in an adverse effect on polyunsaturated fatty acid production in biofilm, a potentially critical nutritional component of biofilm that is consumed by western sandpipers and other shorebirds during their migration stopovers at Roberts Bank. Consequently, the Panel was unable to conclude with reasonable confidence that the Project would or would not have a residual adverse effect on western sandpiper.

The RBT2 project was given conditional Ministerial approval on April 20, 2023. Pursuant to RBT2 Decision Statement Condition 10.4.1, an Independent Scientific Body (ISB) was established by the Chief Science Advisor of Canada to review the proponent's follow-up monitoring plans for salinity, biofilm and Western sandpiper.

The first tranche of material submitted by the VFPA to the ISB comprised:

- Western Sandpiper (WESA) Follow Up Program - monitoring program design
- Appendix A: salinity study design
- Appendix B: biofilm availability study design
- Appendix C: WESA diet study design

The ISB's review (T1R) of the Tranche 1 (T1) material was submitted to the Impact Assessment Agency on October 18th, 2024. It included:

- An explicit description of the criteria used by the ISB to evaluate the proposed studies.
- An explicit description of the constraints on follow-up programs generally and the implications of these limitations to inferences about RBT2 project effects.
- A set of recommendations that, if implemented, would – in the ISB's view – improve one or more of the proposed follow-up studies described in T1.

This report provides a comprehensive review of the second tranche (T2) of material submitted by the VFPA, including:

- Appendix D: WESA Energy Intake Study Component Design
- Appendix E: WESA Foraging Distribution and Intensity Study Component Design
- Appendix F: WESA Abundance Study Component Design

The ISB will also review the WESA Adaptive Management Approach (AMA) that will be developed following the finalized design of the monitoring program and implementation of pre–construction data collection. The AMA will focus on characterizing the thresholds beyond which a potential adverse effect on biofilm or WESA is likely to occur because of changes in salinity caused by the RBT2 project; a description of the adaptive management process that will be implemented if thresholds are exceeded; and potential adaptive management measures that might be implemented to mitigate any threshold exceedances.

Like T1, T2 does not include any information on adaptive management thresholds. Hence, this review focuses on the proponent's proposed set of monitoring parameters and methods for WESA energy intake, foraging and abundance pursuant to Decision Statement Condition 10.4.1.

In preparing the review, the ISB has also taken note of:

- The proponent's covering letter accompanying the Tranche 2 material (PS2)
- The proponent's written response to the ISB's review of the Tranche 1 material (PR1)
- Email communication between the proponent and IAAC in response to several questions raised by the ISB during its review of T2 (PR2)

2. GENERAL FINDINGS AND RECOMMENDATIONS

Here we describe findings that are relevant to several of the proposed T2 studies. Findings specific to individual studies are given in sections 3 -5 below.

2.1. Follow up, WESA population viability and implications for the RBT2 Adaptive Management Approach

Condition 10.4 of the Decision Statement requires the VFPA to develop a follow-up program to verify predicted changes caused by the project on salinity, biofilm and western sandpiper. As noted in T1 (p. 14), of ultimate concern is the impact of RBT2 on *population viability*[1] of Western sandpiper through the hypothesized salinity-biofilm-WESA pathway. Yet any RBT2 follow-up study can characterize only *local* effects (Fig. 1). It follows that implicit to any follow-up study is the assumption that local effects are predictive (at least to some degree) of population-level effects. Or put another way, from (local) follow-up results, inferences must be made about population-level effects. But these inferences are not made explicit in either T1 or T2.

[1] This point is made explicitly by Environment and Climate Change Canada in its review of the proponent's response to additional information requests, available at: https://iaac-aeic.gc.ca/050/evaluations/proj/80054/contributions/id/56952 (accessed March 18, 2025).

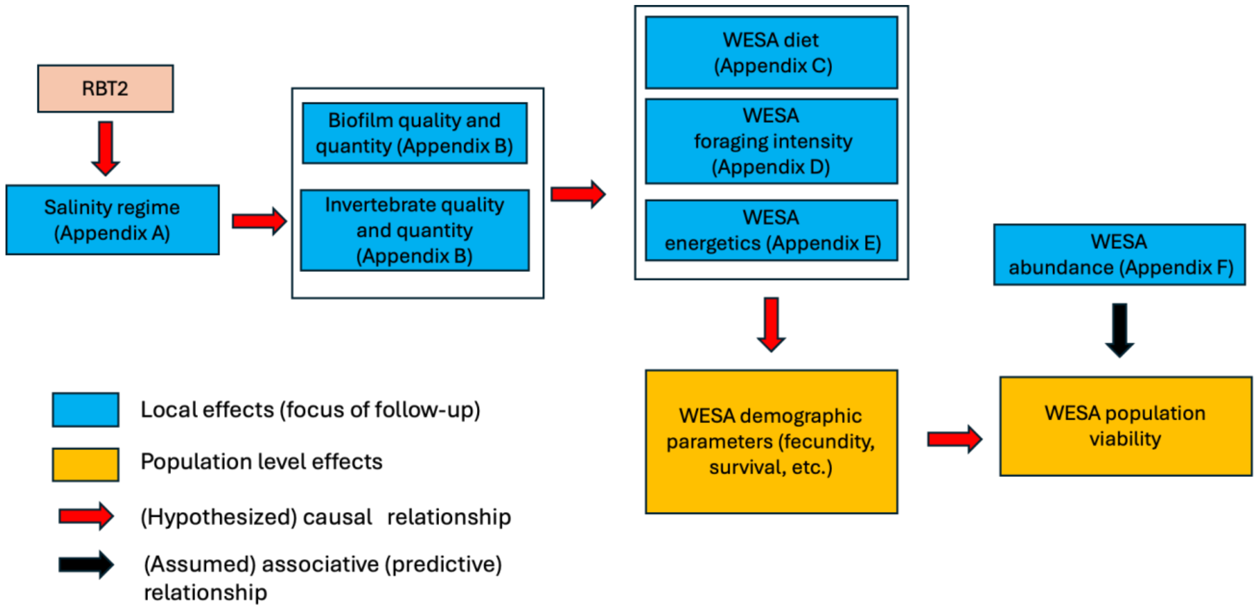

Fig. 1. The hypothesized salinity-biofilm-WESA effects pathway.

Once a follow-up study design is implemented, the set of all *possible* results can be characterized before any *realized* results are known.[2] The set of all possible results can then be classified into discrete subsets according to what conclusion(s) would be drawn about corresponding population-level effects (Fig. 2).

> **Recommendation 1**. Once the set of follow-up studies are finalized, and *before data collection begins*, the proponent should provide an explicit and unambiguous classification of all possible follow-up results into the following categories:

- Results leading to the conclusion of no adverse population-level effects on WESA
- Results leading to the conclusion of adverse population-level effects on WESA, but of a magnitude not requiring additional mitigation.
- Results leading to the conclusion of adverse population-level effect on WESA of a magnitude requiring additional mitigation measures M_1.
- Results leading to the conclusion of adverse population-level effects on WESA of a magnitude requiring additional mitigation measures M_1 and M_2.
- … (as needed, e.g. results leading to conclusion of adverse population-level effects on WESA of a magnitude requiring additional mitigation measures M_1, M_2 and M_3, etc.))

---

[2] For example, in the proposed WESA foraging study (Appendix E), the set of possible results are the possible average metabolite concentrations in WESA before (B) and after (A) project construction in the proposed control (C) and impact (I) areas. So, each set of four average metabolite levels (and associated precision) defines a possible monitoring result.

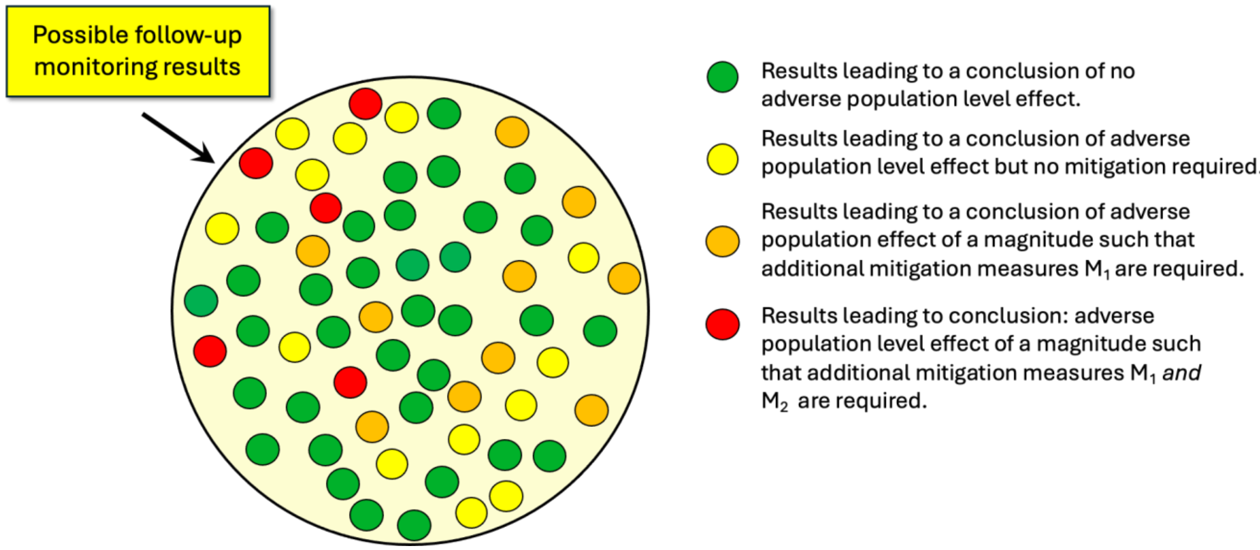

Fig. 2: Classification of possible follow-up results. In this figure, circles represent different *possible* monitoring results, with the colour of the circle indicating the conclusion that would be drawn about population-level effects if the results were obtained. For example, if any monitoring result shown in green were obtained, the conclusion would be that there is no adverse population-level effect on WESA. By contrast, if any result shown in orange were obtained, the conclusion would be that there has been an adverse population-level effect of a magnitude such that mitigation measures M_1 are required.

We note that the boundaries between the categories of possible follow-up results shown in Fig. 2 *define thresholds for adaptive management*. For example, the "boundary" between the set of results (shown in yellow in Fig. 2) leading to the conclusion of an adverse effect for which no additional mitigation measures are required versus those (shown in orange) leading to the conclusion that additional mitigation measures M_1 are required *represents the first adaptive management threshold*. In a similar fashion, the boundary between the set of results shown in orange and red in Fig. 2 represents *the second adaptive management threshold.* The proponent states that the forthcoming WESA Adaptive Management Approach will identify these monitoring thresholds and associated mitigation measures based on a "weight of evidence approach" (PR1, p. 5).

> **Recommendation 2**. Adaptive management thresholds for monitoring parameters should correspond to the boundaries of the categories described in Recommendation 1. Insofar as these thresholds are derived based on a weight of evidence approach, the characterization should make explicit (1) what lines of evidence were used; (2) what evidentiary strength was assigned to a given line, and how this strength was assessed; and (3) what weights were assigned to each line, in characterizing monitoring thresholds, and on what basis these weights were determined.

2.2. Selection of control and impact areas

All three T2 studies propose using one impact area (Brunswick) and two control areas (Westham and Boundary Bay). In T1R, concerns were raised about the appropriateness of both (see e.g. T1R Recommendations 3.1.1 and 3.1.2). In response, the proponent states that it:

> "… intends to implement the ISB's recommendations regarding identifying and comparing factors affecting monitoring parameters within the impact area and control areas to evaluate the suitability of control areas. In addition, as part of the monitoring program, the suitability of control areas will be re-evaluated based on the pre-construction data collection results." (PS2, p. 2)

Unfortunately, no information is provided in T2 about the design, field or laboratory methods, statistical analysis, etc. of these proposed suitability studies. Nor is any information provided on the critical issue of how *given* study results: (1) suitability of Westham and/or Boundary Bay *as control sites* for specific monitoring parameters will be inferred, and (2) the implications of this suitability evaluation to inferences about project effects based on the proposed Before-After/Control-Impact (BACI) design. This despite the recommendation that:

> "…the appropriateness of selected control areas for each study component should be comprehensively evaluated, as should the implications of this evaluation to (a) the choice of experimental design and (b) inferences about project effects (or lack thereof)." (T1R, Recommendation 1, p. 7)

In T1R we registered particular concern about the appropriateness of Westham as a control area for investigating project effects on WESA in part because of the potential for WESA to move between the control area and Westham. Movement between them, and the resulting lack of independence means that a fundamental assumption of BACI designs – indeed, also of Control-Impact (CI) designs – may well be invalid. Lack of independence dramatically reduces the chances of detecting project effects.

In T2, the proponent addresses the possibility of birds moving among control and impact areas:

> "To minimize the potential for double counting shorebird flocks that may move between areas, which can inflate estimates of usage, surveys across the impact and control areas will be conducted concurrently, to the extent possible." (Appendix D, p. 9)

They also state that: "Recent telemetry data indicate that WESA movement among areas of the FRE [Fraser River Estuary] during northward migration remains relatively low and WESA tend to remain within one area of the FRE (e.g., the impact area, Westham Island, Boundary Bay) to forage during a semi-diurnal tidal cycle." (Appendix D, p. 7). However, none of these telemetry results are included in T2 nor in PR1.

Recommendation 3. The proponent should provide strong evidence of the suitability of all proposed control sites for the studies described in T1 and T2. For biofilm, this evidence comes from suitability studies such as those suggested in T1R.[3] Study design, field and analytic methods, sampling procedure, etc. of these studies should be independently reviewed, as should the results thereof. The description of the proposed suitability studies should also provide detailed information on how, *given* field data results: (1) suitability of Westham and/or Boundary Bay *as control sites* for *each* selected monitoring parameter will be inferred, and (2) the implications of this suitability evaluation to *inferences about project effects*. For the WESA studies proposed in T2, such evidence of independence would derive from a study specifically designed to estimate movement rates of WESA among the proposed control and impact areas during the spring migration period.

Recommendation 4.1. Unless it can be convincingly shown that for a *given* monitoring parameter, at least one of the proposed control areas (Westham or Boundary Bay) has high suitability and sufficient independence from the impact area, the BACI design should be abandoned and replaced by a Before-After (BA) design. If a BA design is employed, adaptive management thresholds as described in Recommendation 1 should take into account the revised study design.

Recommendation 4.2. If, for a given monitoring parameter, the decision is made to replace the proposed BACI design with a BA design, the number of "before" years should be extended from two to three to increase the likelihood of detecting project effects (see s. 2.3(1) below).[4]

Although the proponent notes – correctly - that, as a rule, BACI designs are better than BA or CI designs, this conclusion assumes that the selected control areas are indeed suitable *as controls*. With *bona fide* controls, BACI designs improve the strength of the inference that detected changes over time (i.e. between "before" and "after") at impact sites *are due to the project* compared to the strength of this inference under BA designs. That is, BACI designs attempt to resolve the "attribution" problem by testing whether the change between "before" and "after" at impact sites is different than at control sites. But the greater the lack of independence, the more similar the control and impact trends over time will be, and hence, the smaller the probability of detecting a real project effect.

---

[3] An important criterion is that of 'parallel trends' – that is, the temporal dynamics of a given monitoring parameter should be similar between the impact area and the candidate control site before the initial construction phase (see, e.g. Wauchope, H.S. et al. (2021). Evaluating impact using time-series data. Trends in Ecology and Evolution 36:196-205 (DOI: 10.1016/j.tree.2020.11.001)

[4] Indeed, for the monitoring parameters described in Appendix F, three "after" years may be insufficient. If, for example, RBT2 resulted primarily in increases in WESA yearling mortality, population decline would result from the lost reproduction in what would have been their second, third, etc. years, and as such, would be unlikely to be detected for at least 3 years.

2.3. Data analysis

In T1R, it was recommended that the proponent be more explicit and specific about what models will be fitted to the monitoring data, how they will be fitted, and how inferences about project effects will be inferred from fitted models. T2 provides more detail, and it appears that additional information on statistical modelling will be forthcoming (PR 1, pp. 3, 9, etc.)

The additional information provided in T2 reinforces our initial suggestion that something other than a conventional frequentist hypothesis-testing approach to inference about project effects should be adopted (see also T1, Recommendation 2.3.3.3):

(1) *Issues with change detection.* For all follow-up monitoring parameters, assessment of project effects will be based on two years of "Before" and 2-3 years of "After" data. It is well known that the ability of BACI and BA designs to detect project effects decreases when only a small number of years are sampled, especially for systems experiencing substantial natural interannual variability. For both BACI and BA designs of WESA abundance, the number of years represents the number of independent replicates for change detection: repeat days within years (Appendix F), or blocks within areas (Appendix E) do not. Hence, for all proposed follow-up studies, there is a substantial risk of Type 2 error if a (frequentist) hypothesis-testing approach to inference is adopted.
(2) *Issues with sample size*: both T1 and T2 include information on required sample size. As was pointed out in T1R, under hypothesis-testing approaches to inference, estimates of minimal sample size require that the *desired* type 1 and type 2 error rates, along with the *desired* minimal detectable effect size (MDES) be specified: change any of these 3 parameters, and one changes the minimum sample size required to detect an effect of a size specified by the MDES. The corollary is that effects *smaller* than the MDES *will not be detected* under a (frequentist) hypothesis-testing approach, potentially resulting in erroneous conclusions.

The proponent notes that "… sample sizes within WESA FUP study components have been informed by the use of before-after control-impact (BACI) designs and associated power analyses with the ability to detect a maximum [*sic]* effect size of 50% with 80% power and an alpha of 0.05 (when multisite and multiyear data are available)" (PR1, p. 4).

An MDES of 0.50 means that anything smaller than a 50% change in a monitoring parameter over time interval spanning two before years and 2-3 after years (i.e. 4-5 in total) years would not be detected. Yet changes considerably smaller than 50% may be biologically significant, especially if they occurred over such a short time interval.

On the other hand, there is substantial "natural" inter-annual variability in many of the selected monitoring parameters. Hence, lower/smaller MDES increases the risk of detecting year-over-year changes that reflect natural variability rather than project effects. While having a large MDES reduces the chances of misattributing an observed effect to the project when it simply reflects natural variability, it also reduces the chances of detecting smaller project effects that over longer periods may result in large cumulative adverse effects.

These considerations strongly suggest that irrespective of the study designs eventually adopted, a conventional (frequentist) hypothesis-testing approach to inference about project effects is not recommended. Rather, the probability of a project effect of a specific size should be estimated based on a set of plausible fitted models using Bayesian approaches to parameter estimation.

> **Recommendation 5**. For all monitoring parameters, inferences about project effects should be based on (a) specification of set of *plausible* empirical models; (b) model selection using Bayesian approaches; and (c) Bayesian estimation of project effects and associated credible intervals based on selected models. Where multiple predictor variables could potentially be included, the rationale for the chosen set of plausible models as well the basis and procedure(s) for model selection should be clearly described. In cases where multiple models provide comparable fits, estimates of project effects should be based on model averaging. We strongly recommend that Bayesian methods be used to directly estimate the probability of project effects of a specific magnitude based on selected models.[5]

In PR1, the proponent states that: "...however, a hypothesis-testing BACI model is proposed as it is expected to directly and efficiently address the key question of project effects (i.e., is there a project-related differential change in the mean response over time between impact and control areas)." (p. 15)

However, the key question is not whether there any project effects, but rather: *based on the follow-up monitoring data, what is the probability that the true project effect is of a certain magnitude?* Bayesian approaches to model selection and model parameter estimation allow for explicit and direct estimation of this probability, whereas conventional hypothesis-testing approaches do not.

Follow-up is intended to provide an unbiased - and hopefully reasonably precise - estimate of the *probability of (realized) adverse effects of a given size*. Clearly, *what* data are collected, *how* they are collected, and *where* they are collected will affect the accuracy and precision of such estimates. But so too will *the data analysis itself*. For example, a recent study[6] presents compelling evidence of substantial variation in the results of statistical analysis – and inferences

---

[5] See, for example: Conner, M.M., Saunders, W.C., Bouwes, N. *et al.* Evaluating impacts using a BACI design, ratios, and a Bayesian approach with a focus on restoration. *Environ Monit Assess* **188**, 555 (2016). https://doi.org/10.1007/s10661-016-5526-6

[6] Gould, E., Fraser, H.S., Parker, T.H. *et al.* Same data, different analysts: variation in effect sizes due to analytical decisions in ecology and evolutionary biology. *BMC Biol* **23**, 35 (2025). https://doi.org/10.1186/s12915-024-02101-x

therefrom - of the *same dataset* depending on the choice of statistical analysis – that is, *what* models were fitted, *how* they were fitted, and *on what basis* inference about effects were drawn. The authors conclude that:

> "The existence of substantial variability among analysis outcomes raises important questions about how ecologists and evolutionary biologists should interpret published results, and how they should conduct analyses in the future." (p. 3)

This underscores the importance of establishing the data analysis methods and procedures that will be deployed as well as how the resulting outcomes will be used to infer project effect *prior to collection of the monitoring data themselves*. This reduces the risk of *post hoc* selection of methods and procedures to yield desired outcomes and associated inferences about project effects (see T1R, Recommendation 4)

> Recommendation 6. For all proposed follow-up studies, statistical methods and procedures should be described comprehensively and with sufficient detail *prior to the collection of monitoring data*. This should include a clear and detailed description of: (a) each dependent and independent variable, including variable type (nominal, ordinal, interval, or ratio) associated units of measurement, and (in the case of independent variables) the rationale/justification for inclusion; (2) the set of candidate models to which the data will be fit, and the basis on which the set was characterized; (3) the method(s) employed to fit models, and the accompanying rationale; (4) how the final fitted model (or models) will be selected; (5) what results of the model fitting and selection exercise will be reported; and (6) on what basis project effects (or lack thereof) will be inferred from final fitted models. If a Bayesian approach to model fitting and selection is adopted (see Recommendation 5 above), reporting of model results should conform to current best practices for Bayesian analysis.[7,8]

In both T2 and PR1, the proponent has taken significant steps to improve the description of the modelling of follow-up monitoring results as per Recommendation 4 in T1. Nonetheless, some important components as enumerated above are still lacking.

---

[7] See, for example, Kruschke, J.K. Bayesian Analysis Reporting Guidelines. Nat. Hum. Behav. 5, 1282–1291 (2021). https://doi.org/10.1038/s41562-021-01177-7.

[8] Recommendation 5 also has implications to the RBT2 Adaptive Management Approach. If a Bayesian approach is adopted to estimate the probability of a project effect of a specific size given the monitoring data, the corresponding cumulative distribution function can be used to define a probability threshold associated with a set of monitoring parameter threshold values. For example, the trigger for implementation of additional mitigation measures M_1 might be when Bayesian estimation based on follow-up results indicates that the probability that the true value of the monitoring parameter is greater than T_1 is greater than P_1, whereas the trigger for implementation of additional measures M_2 might be when the estimated probability that the true value of the monitoring parameter is greater than T_2 is greater than P_2, etc. Thus, under this scheme, each adaptive management threshold is characterized by a threshold probability (P_1, P_2, etc.) *and* an associated threshold magnitude of the monitoring parameter (T_1, T_2, etc.).

Recommendation 5 has important implications to the estimation of statistical power. In Bayesian estimation, the goal of power analysis is most often to achieve a certain precision for an estimated parameter[9], e.g. an estimate of the "true" project effect on a specific monitoring parameter.

> Recommendation 7. If Recommendation 5 is adopted, then sample size calculations (which hitherto have been based on a frequentist hypothesis-testing approach to inference) for all proposed studies should be based on desired credible intervals for estimates of project effects.

In response to T1 Recommendation 5, the proponent states:

"To clarify, sample sizes within WESA FUP study components have been informed by the use of before-after control-impact (BACI) designs and associated power analyses with the ability to detect a maximum effect size of 50% with 80% power and an alpha of 0.05 (when multisite and multiyear data are available)." (PR1, p. 4).

80% power corresponds to a threshold Type 2 error rate of 0.2, 4 times larger than the threshold type 1 error (0.05). Although this is a usual threshold type 2 error rate, this is not sufficient justification for using it in the follow-up context: particularly for follow-up of highly variable systems, the odds of missing a true project effect are higher than falsely detecting an effect when there isn't one.[10]

> Recommendation 8. If a hypothesis-testing approach to inference about project effects continues to be employed (Recommendation 5 notwithstanding), the proponent should provide a clear and compelling justification for the choice of threshold Type 1 and Type 2 error rates. Specifically, there should be a clear and compelling justification for different threshold Type 1 and Type 2 error rates.

2.3. Additional monitoring components

All three T2 studies propose a set of additional monitoring components (see Table 3-2 in Appendices D-F). These additional components will be treated as covariates in fitted models, potentially removing sources of variation in monitoring parameters that would otherwise be

---

[9] See, for example: Kruschke, J.K., Liddell, T.M. The Bayesian New Statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. Psychon. Bull Rev 25, 178–206 (2018). https://doi.org/10.3758/s13423-016-1221-4.

[10] In particular, it is unclear why we should be four times more forgiving of errors in accepting a false null hypothesis (i.e. in concluding there is no project effect, when in fact there is one) compared to rejecting a true null hypothesis (i.e. concluding there is a project effect when there isn't one).

considered random error. The presumption is that this reduced error variance would result in greater ability to detect project effects.

But this would be so only if there is no covariance between modelled project effects and the proposed additional monitoring components. If covariances are present, wildly different models (e.g., those with a project effect but no additional explanatory variables, versus those with explanatory variables but no modelled project effect) may have similar fits. As inferences about probability of project effects of a given magnitude are derived from fitted models, *which* set of models one chooses will then determine conclusions about project effects. Such a possibility underscores the importance of Recommendations 5 and 6 and prompts:

> Recommendation 9. The selection of additional monitoring components for a specific study should be based on 3 criteria: the component is (1) unlikely to itself be affected by the project; (2) likely to vary between control and impact areas; and (3) one for which there is independent evidence of an association between it and one or more selected monitoring parameters for the study in question.

3. **WESA ENERGETICS STUDY (APPENDIX D)**

3.1. Study design

The proponent proposes to use plasma triglyceride and β-hydroxybutyrate levels in WESA captured in a specific area (either control or impact) as an index of mass changes, rate of energy gain and *habitat conditions in that area*. In the RBT2 follow-up context, this last assumption amounts to the assumption that under the proposed BACI design, plasma metabolites caught in, say, Brunswick, reflect habitat conditions in Brunswick and not in, say, Westham.

The studies in cited in Appendix D, as well as other studies not cited, do indeed provide substantial evidence that plasma metabolite levels can be used as indicators of energy intake/mass change in the field. But these studies do not provide sufficient evidence to justify the *critical* assumption that plasma metabolite levels in captured WESA reflect foraging habitat quality in the immediate vicinity of where they were captured (Appendix 1).

> Recommendation 10. Absent compelling evidence of limited movement between the three areas over the time scales that matter for plasma metabolites (i.e., < ~ 1hr), the proponent should abandon the idea of control areas for the WESA energetics follow-up, and adopt a Before-After design (see also Recommendation 4 above).

During the peak migration period, 30 WESA will be sampled per area per year. In PR2, the proponent states that:

> "In the WESA Energy Intake Study Component, sample sizes were informed by the practical feasibility of capturing WESA during specific tidal periods. Previous capture

> programs indicated that, given the challenges of capturing WESA on low tides, a sample size of 30 WESA/area would be the largest we could consistently achieve across monitoring years. This informed a conservative decision to maintain a sample size of 30 WESA/area for preliminary power analysis."

Here the estimated required sample size is *not* based on a specified minimal detectable effect size (MDES) and type 1 and type 2 error rate as it is for abundance (Appendix F) and foraging intensity (Appendix E). Rather, it is the sample size, and desired type 1 and type 2 error rates that are specified to generate estimates of (relative) MDES of 29% (see Appendix D, Fig. 5, p. 19).

> Recommendation 11(a). The proponent should state explicitly that unlike the sample size calculation for other studies, in this case the calculation concerns the MDES.

Little information is provided about when during the spring migration period WESA will be sampled. Moreover, as noted in Appendix D, plasma metabolite levels may vary substantially throughout the diurnal cycle. The proponent states that:

> " Efforts will be made to capture WESA during peak blood plasma triglyceride levels (i.e., after 8AM) (as per Seaman et al. 2006) but will also target low-light periods of day (e.g., dawn, dusk) and nighttime captures as needed to increase capture efficiency by reducing net visibility to WESA."

Capturing birds from dawn to dusk (and beyond) will only serve to increase variation in plasma levels and decrease the precision of annual estimates.

> Recommendation 11(b). Every effort should be made to increase the precision of annual average estimates of plasma metabolite levels by capturing birds during a fixed time window of several hours following mud flat emergence (to ensure that birds have been actively feeding for a relatively constant period of time) and concentrating captures during the peak migration period.

3.2. Monitoring parameters

The proposed study uses plasma triglyceride and β-hydroxybutyrate concentrations as monitoring parameters to infer changes in energy intake. However, plasma triglyceride and β-hydroxybutyrate do not indicate other, potentially major, sources of energy for WESA migration. There is, for example, some evidence that phospholipids transport twice as much energy as triglycerides in WESA during spring migration.[11]

[11] See, for example, Fig. 1 of Guglielmo et al. (2002) Plasma and muscle phospholipids are involved in the metabolic response to long-distance migration in a shorebird. J. Comp. Physiol. B 172: 409-417.

Recommendation 12. Phospholipid concentrations should be added as a monitoring parameter for the WESA energetics study (Appendix D, Table 3-1)

4. WESA FORAGING STUDY (APPENDIX E)

4.1. Study design

The WESA foraging study proposes the same control and impact areas as the energetics study (Appendix D) and WESA abundance (Appendix F) studies and is therefore subject to the same concerns about independence and suitability of the proposed control and impact areas.

However, the risk of non-independence among control and impact areas is considerably higher for WESA abundance (Appendix F) and energy intake (Appendix D) than it is for foraging intensity (Appendix E) because of the rapid WESA defecation rate (every 2.2 min on average). ***If*** this is approximately the defecation rate for WESA at Roberts Bank, droppings will generally reflect foraging activity within a comparatively small spatial neighbourhood. The implication is that droppings counted in, say, Brunswick, are much more likely to reflect foraging in Brunswick versus foraging in Westham (and *vice versa*), such that the problem of lack of independence (as noted above for the study described in Appendix D) is reduced.

As noted above, independence of control and impact areas is a critical assumption underpinning the validity of the proposed BACI design. A study of WESA movement as suggested in Recommendation 3 would provide estimates of the distribution of WESA "foraging residence times" in small spatial neighbourhoods, which could then be used to estimate the likelihood of a bird moving between neighbourhoods in short time intervals.

4.2. Monitoring parameters and components

Foraging intensity is simply the product of dropping density and a constant. Since it is assumed to be constant over time, over area – indeed, over everything - results based on (calculated) foraging intensity will be identical to those based on (raw) dropping densities. So why use it?

Recommendation 13. The monitoring parameter should simply be the number of size class 1 droppings per m^2.

Statistical analysis

In assessing foraging intensity, it is proposed that each area will be subdivided into 10 blocks. There is concern about the lack of independence not just among proposed impact and control areas, but among sites or survey blocks within areas. Based on the model descriptions provided, this has not been accounted for. Accounting for such lack of independence requires the fitting of models that account for spatial autocorrelation.

In T1, concern was expressed about the spatial autocorrelation issue in the context of the proposed salinity study (Appendix A). In PR1, the proponent responded that:

> "Spatial and temporal autocorrelation could be important when considering the salinity measurements collected at the 11 monitoring stations; however, given the care that was taken in choosing the monitoring locations, spatial autocorrelation is not expected. An empirical variogram will be plotted as a first estimate of the variogram model to assess potential autocorrelation and to assess if an adjustment to the study design is required." (p. 10).

Spatial autocorrelation may influence estimates of the magnitude of project effects (in the Bayesian case) or the likelihood of detecting a project effect (under a frequentist hypothesis-testing approach). Based on the salinity modelling (Appendix A), the largest project effect is predicted to be in blocks BR4,5,6 and 7 (see Appendix A, Fig. 3.1), with weaker effects in blocks BR1, 2, 9 and 10. As the latter are directly adjacent to the Westham control blocks W15 and 16 it is not unreasonable to expect that there might still be weakened impacts occurring there. This spatial gradient in effects within the control and impact areas may reduce the likelihood of detecting a project effect on foraging intensity.

> Recommendation 14. The set of models fit to the foraging intensity data should include models incorporating spatial autocorrelation effects.

5. WESA ABUNDANCE STUDY (APPENDIX F)

5.1. Study design

For the studies described in Appendices A-C in T1, and D and E in T2 all selected monitoring parameters are indicators of (hypothesized) *local* effects of RBT2. As noted in Recommendation 1, estimated project effects on these monitoring parameters should be associated with explicit conclusions about population-level effects (see Fig. 2)

But local WESA *abundance* is qualitatively different. If there were suitable control areas, then a BACI design could be used to estimate RBT2 effects on the spatial distribution of migrating WESA at Roberts Bank by comparing temporal trends in local abundance at control and impact areas, as proposed in Appendix F. But inferring population-level effects from changes in the spatial distribution of WESA at Roberts Bank is problematic: as explained in Appendix 2, this inference assumes extreme year-over-year site fidelity of migrating WESA at Roberts Bank, and in Appendix F, there no evidence is presented that justifies this critical assumption.

> Recommendation 15. Absent strong evidence of migratory site fidelity at the appropriate spatial scale within Roberts Bank, the proposed BACI design Appendix F should be abandoned. Rather, effort should be made to generate annual estimates of local WESA abundance at Roberts Bank to allow for inferences about changes in WESA population size over time using a BA design where surveys are conducted in all 3 areas.

All of the recommendations that follow below are made assuming that annual monitoring of WESA abundance at Roberts Bank will continue.

5.2. Monitoring parameters and components

In Appendix F, it is proposed that each area (Brunswick, Westham etc.) be divided into two spatially disjunct blocks, with one surveyor per block per survey day. This means that there will be only a single total count per area (presumably based on summing counts for each surveyor), and consequently, no way of estimating measurement error.

> Recommendation 16(a). On each survey day, three independent surveyors should survey each block in each area. For each area, for each day, this will yield three independent counts for each block, one per surveyor, thereby permitting an estimate of measurement error as well as observer effects. For each survey day in each area, the 3 surveyors should be randomly assigned to a starting block and conduct their surveyors finishing in the other block.

While bird surveys often use two independent observers, the substantial increase in the precision of daily counts associated with N = 3 versus N = 2 replicates per block per area will increase the ability to detect changes in abundance over time.

According to Appendix F, based on the proposed 10 daily total counts per area, an estimate of the total number of birds using the site over the migration period is obtained by:

(1) For each survey day and area, estimating the proportion of birds that are WESA and multiplying the total bird count by this proportion.
(2) Using daily counts and proportions to estimate WESA counts on days when no survey was conducted; then
(3) Summing estimated WESA counts over all days in the migration period and dividing by the average length of stay (LOS) to correct for the possibility that surveys on successive days may be counting the same birds.

Notwithstanding the fact that this approach has been used previously by Drever et al.:

(a) Because there is only a single WESA count per area per survey day, the method does not allow for an estimate of measurement error on daily counts.
(b) The estimated proportion of birds (in each survey area, on a given day) that are WESA clearly has an associated error, which appears to be assumed to be zero in the Drever et al. method.
(c) In generating estimates of WESA counts during days with no surveys, Drever et al. used model estimates, even though, based on their data, the fitted models have considerable associated uncertainty.
(d) Estimated daily counts of WESA and estimated population totals are the sums, products or ratios of random variables, for which associated sampling variances should be computed.[12] Based on the published description, Drever et al. appear not to have done so.

(a)-(d) mean that the uncertainty (imprecision) associated with any estimate of the number of WESA using an area is likely to be underestimated by the Drever et al. method.

> Recommendation 16(b). The proportion of birds that are WESA should be considered a monitoring parameter (i.e. added to Table 3-1) rather than an additional monitoring component (i.e. deleted from Table 3-2).

> Recommendation 16(c). Any estimate of the total WESA population ("Total WESA bird days") at Roberts Bank should ensure that all relevant associated uncertainties are captured by ensuring that: (1) daily surveys of WESA abundance provide estimates of measurement error (see Recommendation 16(a); (2) surveys to estimate the proportion of WESA in daily shorebird counts provide estimates of associated precision;[13] (3) model uncertainty associated with estimating abundance on non-survey days is incorporated; and (4) estimates involving products, quotients or ratios of random variables should employ well-established methods for estimating associated sampling variances.

As noted above, the Drever et al. method requires an estimate of the length of stay (LOS). In Appendix F, the LOS is assumed constant, not only over time, but among control and impact areas. However, if RBT2 is adversely affecting local foraging habitat quality, it may also affect LOS by, for example, requiring WESA to stopover longer at Roberts Bank to build up the energy stores required for the northward migration. Indeed, increases in LOS over time may result in increases in WESA daily counts, not because the WESA population has increased, but because birds stopping over at Roberts Bank are staying longer.

---

[12] On common way of doing so it to employ (minimally) first order Taylor expansions (see, for example http://www.senns.uk/Stats_Notes/Variance_of_a_ratio.pdf).

[13] One possible method is at K randomly selected points in an area, to use a non-overlapping field of view photograph, with each photograph providing a count of the total number of shorebirds (*N)* and the number of WESA (*n)* in the field of view. The fitted empirical relationship between *n* and *N* can then be used to estimate the number of WESA for any shorebird count, as well as the associated precision of the estimate.

Recommendation 16(d). Length of stay (LOS) should be considered a monitoring parameter (i.e. added to Table 3-1), and a follow-up study to assess potential changes in LOS should be designed and implemented. Such a study would necessarily involve marking WESA and estimating residence times based on resighting decay rates. If carefully designed and implemented, the same study could also be used to assess movement rates among Westham, Brunswick and Boundary Bay (see Recommendation 3 above). Empirical estimates of LOS (along with associated imprecision) should then be used to estimate annual total bird days.

In Table 3.2, the cumulative number of avian predators is considered an additional monitoring component. While we understand that the presence of avian predators may affect shorebird counts, it is unclear what "cumulative" refers to (Over species of predator? Over time?).

Recommendation 16(e). For each survey day, in each area, the number of avian predators observed during the survey period should be counted by each surveyor. The result would be three (one per surveyor – see Recommendation 16(a)) independent estimates of the number of avian predators per survey day per area.

APPENDIX 1

The proponent states:

> "Prior studies conducted in the early to mid-1990s have documented high site fidelity by WESA during northward migration stopovers, with little movement from the locations where WESA are initially detected (Butler et al. 2002)." (Appendix D, p. 7)

But Butler et al. (2002) provides little evidence of site fidelity *at a spatial scale relevant to the situation at Roberts Bank*. Their study involved marking WESA on Sidney Island, a small island 30 km from the Fraser River Delta, and recording how many of them were resighted one, two, …. days after marking. Information on *where* on the island birds were marked, and *where t*hey were resighted is not provided in the paper but based on the description of the field methods and accompanying map, it appears birds could have been resighted up to 3 or 4 km from where they were marked. Moreover, Sidney Island (being an island) is an area of comparatively limited and isolated WESA foraging habitat. Indeed, it was precisely because of this habitat isolation that it was chosen as the study site, the authors noting that:

> "The probability of resighting a marked sandpiper on the vast Fraser River delta was low so we chose Sidney Island about 30 km southwest of the Fraser River delta." (p. 104)

In other words, if marked birds were still around on Sidney, they would have to be resighted in a comparatively small and isolated area of habitat close to where they were marked. By contrast, the implication is that in the Fraser River Delta, they could be anywhere.

The proponent also states:

> "Regardless of the potential for WESA movement among control and impact areas (see Section 2.1), measures of energy intake are expected to reflect habitat quality at the capture location due to the rapid turnover rate of blood metabolite levels (see Section 2.2). Thus, even if some WESA move between the impact and control areas, measures of blood metabolites are expected to provide site-specific measures of energy intake. For example, all birds captured and sampled at Westham can reasonably be expected to have been feeding within the Westham control area during the previous 10–20 minutes, the time period for which blood metabolite levels reflect energy intake rates (Zajac et al. 2006)." (Appendix D, p. 7)

However, in birds - as in homeotherms generally - *whole body* metabolic rate scales as power function of body size, with an estimated exponent of 0.66 - 0.75. Consequently, *mass-specific* metabolic rate scales as a power function with exponent [-0.25, -0.34]. Zajac et al.'s work is with warblers that weigh about 8 g; adult WESA weigh 22-35 g. As WESA have 2-4X the body mass of Wilson warblers, a back-of-the-envelope calculation using an exponent of range of 0.61-0.75 and a body size range [22-35g] suggests that mass specific metabolic rate for WESA ([17-38 Kj/day/kg] is 2-5X *slower* than Wilson's Warblers. Hence, plasma metabolite levels in WESA may well reflect intake not within 10-20 minutes, but within *20 minutes to 1.5 hours*. With flight speeds of 45-65 km/hr, this would be sufficient time for WESA to move back and forth between the Westham control and Brunswick impact areas dozens of times while feeding, and multiple times between Brunswick and Boundary Bay. The result would be that plasma metabolites of birds captured in, say, Brunswick could well reflect feeding in 2 - if not all 3 - areas.

Guglielmo et al. (2002), Zajac et al. (2006) and Smith et al. (2021) are cited as evidence that plasma triglyceride and β-hydroxybutyrate levels are reliable indices of mass changes, rate of energy gain and habitat conditions at the capture area. For example:

> "Fieldwork will be tide dependent with the initiation of mist netting occurring at least one hour after mudflat emergence to allow WESA to forage and blood plasma metabolite levels to adjust and be representative of habitat conditions of the capture area (Guglielmo et al. 2002, Zajac et al. 2006)." (Appendix D, p. 8)

Although the cited studies do provide substantial evidence that plasma metabolite levels can be used in the field as useful indicators of energy intake/mass change, they provide little evidence that plasma metabolites have significant predictive value *with respect to habitat quality* in the field *at the spatial resolution required for RBT2 follow-up using a BACI design:*

- Guglielmo et al. (2002) compared plasma metabolite levels of migrating WESA at Sidney Island and Boundary Bay, which are 30 km apart. The former is (as the name suggests) an island, with no other foraging habitat in the immediate vicinity. Brunswick and Westham are immediately adjacent to one another. Metabolite profiles that have some predictive value with respect to habitat quality at sites located 30 km apart need not have the same predictive value with respect to sites located ~3 km (or less) apart, which is the case for the mid-points of the Westham and (immediately adjacent) Brunswick areas.
- Smith et al. (2021) studied the relationship between plasma metabolite concentrations and changes in body mass of lesser scaup under controlled laboratory conditions with two treatment groups: (1) fasting (no food for duration of stay in captivity); and (2) feeding (manually fed with a slurry of food and water by way of an esophageal tube). Strong correlations between plasma metabolite levels and changes in body mass (loss or gain) were indeed documented. But in this study, the extreme dietary variation under laboratory conditions was specifically designed to maximize mass loss or gain *in order to study the correlation with plasma metabolite levels*. There is no evidence from this study that plasma metabolite levels are predictive of foraging habitat quality in the wild, even in scaup, much less in WESA.
- As noted above, Zajac et al. (2006) provide compelling evidence that plasma metabolite levels respond rapidly (within 10-20 minutes) to dietary intake in 8 g Wilson warblers. But based on the well-described mass-specific metabolic rate – body mass scaling relationship in homeotherms, there is good reason to believe that response times for WESA may be considerably longer.

- Even at large spatial scales, the predictive value of plasma metabolite levels with respect to *local* foraging habitat quality is unclear. For example, Williams et al. (2007)[14] could detect no such relationship for spring migrating WESA over flyway sites that varied over more than a 10-fold range in prey biomass density (see Fig. 5 of the Williams et al. study).

Given the above, there is insufficient evidence to justify the *critical* assumption that plasma metabolite levels in captured WESA reflect foraging habitat quality in the immediate vicinity of where they were captured.

[14] Williams, T. D., Warnock, N., Takekawa, J.Y and M.A. Bishop (2007). Flyway-Scale Variation in Plasma Triglyceride Levels as an Index of Refueling Rate in Spring-Migrating Western Sandpipers. The Auk Vol. 124: 886-897.

APPENDIX 2

At issue is the extent to which *local* changes in the spatial distribution of migrating WESA at Roberts Bank based on the proposed BACI design permit valid inferences about changes in WESA *population* abundance. Of particular concern is that even if adverse local effects are detected on other monitoring parameters (e.g. salinity, biofilm, WESA foraging intensity, etc.) the lack of a detected effect on the (local) spatial distribution of WESA between control and impact areas might be used to justify the conclusion that despite RBT2 having local effects on some monitoring parameters, it has no effect on the outcome of ultimate concern, *viz.* WESA population viability (see Fig. 1). Such a conclusion would have potentially profound implications to, for example, the need for additional mitigation measures.

Is a detected local effect on spatial distribution of WESA at Roberts Bank predictive of *population-level* effects on WESA abundance?

(1) Suppose RBT2 *does* have an adverse effect on the WESA population. Unless adult WESA have extreme year-over-year migratory site fidelity at Roberts Bank (i.e. in successive years, survivors return to the same area (e.g. Brunswick), and don't go to another area (e.g. Boundary Bay or Westham)), even a dramatic decline in population size overall would not be reflected in differences in local abundance trends in the proposed control versus impact areas. Hence, there could be an adverse population-level effects of RBT2, yet the inference from the proposed (BACI) WESA local abundance study be that there are none.
(2) Suppose that a high level of migratory site fidelity does not exist, a BACI design was implemented, control areas were *bona fide* controls, and local WESA abundance declined over time more rapidly in the impact area compared to control areas. The inference would then be that the project has adversely affected the WESA population. But this need not be the case: the difference may simply reflect the fact that over time, the impact area has become less attractive to migrating WESA, who now prefer the control areas more highly (perhaps due to RBT2-induced reduction in foraging habitat quality, or perhaps to a change in habitat quality that is unrelated to RBT2) but are still able to obtain sufficient food resources to successfully migrate and breed. So as with (1), local changes detected using the proposed BACI design could easily result in an incorrect inference about adverse population-level effects.

The inferential errors associated with scenarios (1) and (2) above are *not mitigated* by replacing BACI designs with BA designs as suggested in Recommendation 4 if, in such a BA design, *only the impact area is surveyed annually*. They are mitigated to a substantial degree if annual surveys encompass all three areas, since all that is then required is year-over-year fidelity to Roberts Bank *generally*, not to specific subareas (Brunswick, Westham, etc.).

Extreme migratory site fidelity at Roberts bank is a necessary - but not sufficient - condition for the validity of inferences from *local* changes in WESA abundance at Roberts Bank to *population level* changes in abundance where local changes are based on (a) the proposed BACI design, or

(b) a BA design focused only on the Brunswick impact area. Absent evidence of such fidelity, inferring population-level effects on WESA abundance from changes in the (local) spatial distribution of migrating WESA at Roberts Bank (under a BACI design) or from local changes over time (under a BA design) is unwarranted. By contrast, if much of the WESA population uses the Roberts Bank area during migration, then from changes over time in local WESA abundance as *determined from surveys in Brunswick, Westham and Boundary Bay collectively,* inferences to population-level effects would seem to be justified.