**Third Pass: Quantitative aspects of the study**

* **Response variables / traits of interest**Repeat for each trait/parameter of interest

|  |  |
| --- | --- |
| Have you chosen your treatment levels to reflect your best guess of expected response norms to the main drivers? (see Fig 2 of GCB manuscript).  Does your design allow you to determine responses to both individual and multiple drivers? |  |
| What sre the measurement units?  How do you standardise your response data?  Size, body weight, surface area, sex, time, …? |  |
| What biological variation do you expect in the responses/traits?  e.g. within samples, across time, across places: different populations /genotypes / phenotypes, states of acclimatisation …  Do you expect spatial and temporal dynamics in the responses/traits  Latitudinal, diurnal, seasonal, …?  How do you accommodate such temporal and spatial variability in your measurements?  Replication, time of measurement, sample origin, analytical detection limit, etc … |  |

* **Explanatory variables / drivers of interest:**

|  |  |
| --- | --- |
| Are your drivers constant or variable?  Do you expect temporal dynamics or spatial variation in the drivers? e.g. tidal, diurnal, seasonal, weather dependent, random fluctuations, etc…?  How do you accommodate the variation in your measurements?  e.g. for how long and over what spatial extent will you need to measure? What is the best sampling frequency, and spatial resolution? |  |
| Are you interested in changes in means / variances/ maxima or minima?  What are the measurement units? |  |
| For experiments: what is an environmentally relevant duration of exposure?  Is such duration feasible? What are the constraints imposed by using shorter durations? |  |
| Which of your drivers are naturally confounded and can be combined / collapsed?  e.g. salinity & alkalinity, temperature and aragonite saturation state |  |
| How will you deal with interactions among drivers?  Physical-chemical e.g. salinity – pH – temperature; metals – pH…  See also “Confounding” |  |

* **Experimental design and statistical analyses:**

|  |  |
| --- | --- |
| What is your sampling design:  How many treatment levels per driver?  Do you have enough treatment levels for a regression analysis? Do you expect the response to be non-monotonic?  How many replicates per treatment level and control?  How variable are your replicates? Do you expect to see obvious or subtle effects? Do you have enough replicates for each level of a categorical analysis (ANOVA-style design)? What other drivers could introduce variability? Are there any covariates that could explain variability?  What is your experimental sampling unit / replicate?  This is typically the container or area to which you apply your experimental treatment. Is this the level you *want* to replicate? Are you *able* to apply each treatment independently at this level?  Note: your sampling unit may be within each experimental unit (e.g. multiple fish in a tank), but if the treatment is applied to the tank, then the tank becomes the replicate. |  |
| Do you understand the relationship between the number of treatment levels, the number of replicates per level, and the sample units?  Have you got preliminary data you could use for an *a priori* Power Analysis?  Answering this question may be the most important component of your whole study!  Run simulations and look at changes in the performance of your study. You can use existing code to assess this, e.g., a link to the free R code:  <https://edild.github.io/lc50_bias_sim/>  or a free downloadable simulator such as G\*Power:  <http://www.gpower.hhu.de/en.html>  (NB neither of these methods can be used for *post-hoc* Power Analysis. See GCB Stats Note for details). |  |
| What is the most relevant and valid control?  Is the concept of a “control” relevant? Would a gradient study be better?  (e.g. are you measuring gradients of change)?  If using controls, are they locally/regionally relevant in all aspects (including co-limitations)?  Have you considered using pre-industrial conditions as main (or additional) ‘control’?  Do you also need a control from Time zero or a ‘baseline conditions’ control? If you collected those data, what would you do with them? Do you need the same or a greater level of effort to collect this data point? |  |
| What type of data will your response and driver constitute?  e.g. continuous variable, time-series measurements, point measurements, presence/ absence, count data, categorical data, compositional data…  e.g. univariate or multivariate responses |  |
| What do you expect to be your greatest source of variation?  Is this variation biologically important, or is it due to experimental error? (e.g. measurement error).  How will you deal with this variability?  Will you average measurements within a replicate (e.g. across individuals within tanks / site)?  Are the levels of variation similar in all your treatments, or are the variances heterogeneous? If the latter, how will you deal with this when you analyse the data? |  |
| What type of statistical analysis or modelling approach will you use?  What will be the specific statistical model you will apply to your data?  Do you know how to implement this model and understand its results?  e.g. does your model consider both individual main effects and interactions?  Can it deal with both fixed and random factors (if present)?  What is the power of your experimental design? Test your design with a dummy data set. |  |
| Validation: Can you validate your results in any way?  e.g., compare lab against field data? use more than one measurement technique?  Quality assurance of sample analyses via cross-lab comparisons/calibrations?  Quality control of data?  What criteria will you use for rejecting spurious data?  In theory, what could go wrong during the experiment, and how might this affect your data (e.g. can you separate biological contamination of a header tank from a treatment effect or are they confounded)? Is this a plausible outcome, and can you redesign your study to address this? If you cannot modify your design for a plausible risk (e.g. due to physical constraints, biohazard risk, etc.), what can you do to help convince yourself and others that this outcome has or has not occurred? |  |

* **What resources do you need, and do you have access to all of them?**

|  |  |
| --- | --- |
| Field study. Access to:   * Transport * Personnel * Sampling gear * Measurement instruments & calibration kits * How many days will you need to go into the field, and how many contingency days will you need to account for bad weather? * What infrastructure do you need to get your samples back to the lab (if relevant)? |  |
| Lab study. Access to:   * Lab / aquarium / seawater of appropriate quality * Measurement instruments & calibration kits * Biota * Personnel for regular maintenance and measurement, as well as for final analysis and take-down * How long do you need to run the experiment for? |  |
| Sample processing:  Access to   * Sample preservatives/storage * Laboratory instruments * Reagents |  |
| Statistical analysis   * Expert guidance, software * Are you aware that statisticians or other experts are much more likely to provide useful help if you meet with them during the design stage rather than after data have been collected? * Similarly, are you aware that statistical modelling cannot rescue a poorly designed study? |  |
| Time and costs:  How long will the whole study take, and much will it cost?  Try to upscale:  How long will the measurements and sample processing take? How much will it cost to process each sample? Can I afford this?  Try upscaling with a spreadsheet: List each individual step that needs to be done to execute the work; add a column with your best guess of how long each step will take per sample, another one of the cost per sample, and one for the number of samples/items. Multiply to obtain the total amount of time and costs. Remember to add time for breaks.  Try to validate your estimates (we all are far too optimistic!!): e.g., ask a peer how many samples they have been able to process per day etc.  If the numbers are too big, can you use alternate methods? (e.g. digital photographs of samples can be quicker than measuring directly *in situ*, but may be less accurate and require more time back in the office). |  |
| What are your most important constraints?  e.g., Time? Resources? Experience and skills by self and advisor / mentor / collaborators?  How can these constraints be overcome?  e.g., can you enter into collaboration with an additional person? |  |

If any of the constraints can’t be overcome, refine your question, and run through the questionnaire again.

Finally, check your objectives again, and see whether they are all met:

|  |  |
| --- | --- |
| Does this design really answer your main question and objectives?  If not: is there scope to further fine-tune your question to make it more relevant / tangible / feasible?  [add links to virtual scientist & webinars] |  |