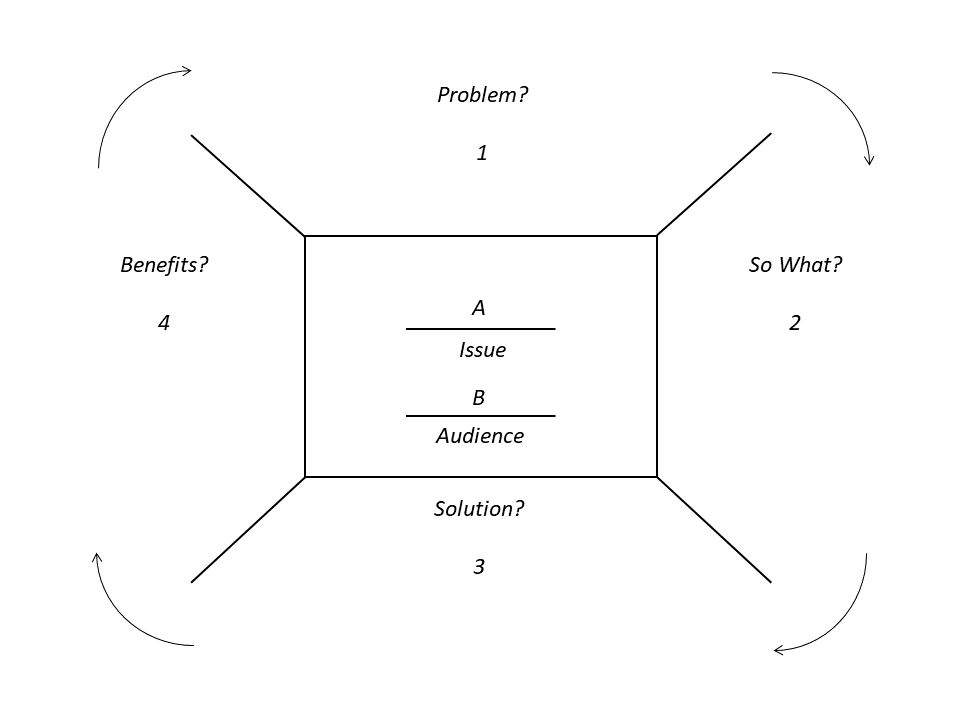
**Brian’s vision for paper:**

<https://mcolvin.github.io/2019-04-03-aldridge-blog-post/>



A, issue: Taking advantage of as much data as possible

B, audience: Scientists working on contaminant research

1. Problem(s)?

* Opportunistic sampling and budgets limit the amount of samples available to estimate/understand contaminant concentrations at the waterbody level.
* If we want to use these data for contaminant monitoring, we need to find a way to standardize across waterbodies. If there is no strong contaminant-body size relationship (i.e., bioaccumulation), an average concentration across all fish caught might make sense. If there is a strong contaminant-body size relationship, then it makes sense to standardize the concentration to a common body size.
* Past research has focussed on individual lake-level regressions and predictions to 1000 g fish for example. However, if the distribution of fish sizes has not be sampled “properly” (i.e., sample availability), then the predictions may need to be extrapolated for that waterbody. Or, if there are too few samples (e.g., <4), then the lake may need to be discarded altogether.
* Some data are more sparse than others.

1. So what?

* Discarding lakes where significant effort was put into sampling fish.
* For contaminant monitoring and across-waterbody studies, we might not catch size classes of interest leading to likely over/under estimating concentrations when extrapolating when not “borrowing strength”

1. Solution?

* Mixed model approaches borrow strength from all observations to generate predictions while still allowing for individual lake intercepts and slopes.
* Estimate importance of contaminant-body size relationship, between-waterbody incercept/slope, and residual
* ML - fast model output but predictions do not account for uncertainty without bootstrapping. Depending on model size, this could take a long time.
* AB - fast model output and incorporates uncertainty of random effects
* B - slow model output and incorporates uncertainty of random effects

1. Benefits?

* Use all available data and borrow strength for estimates in sparse data locations (i.e., “better extrapolating”?)
* Even if there is no strong contaminant-body size relationship, the predicted concentration will be approximately the average concentration in that waterbody.
* Use information in mixed model to inform sampling
* For other contaminants with limited sampling, mixed models could be a useful tool.