pHOS

To analyze the number of strays, Josephson et al. 2021 used proportion of hatchery-origin spawners, or pHOS, as their metric of comparison between streams, years, management units, etc. While my study primarily uses the average number of strays (total strays / number of surveys) for modeling and comparison purposes, I wanted to include a brief results section with some pHOS values as well.

Josephson et al. 2021 calculated a weighted pHOS for each stream-year observation, which was as follows for stream *i*

where

And

So, the weighted pHOS for stream *i* in a single year is the pHOS determined from each visit, multiplied by a weighting factor, the proportion of the total dead count counted on that visit, summed for all surveys.

I do not have dead counts available to me for 2008-2011 survey years (confirmed by phone call to Andy Piston), so I cannot replicate the Josephson et al. 2021 weighting method. At least, I can’t for 2008-2011 surveys, which is about 40% of my data. At first, I thought that perhaps I could substitute wiv dead counts for the count of samples, but this just leaves me with the raw (unweighted) pHOS. Here is a quick demonstration of how/why, because it took me a bit to think through and figure out initially:

Admiralty Creek was sampled 2 times in 2009. Once on 8/12/09, and once on 8/17/09.

|  |  |  |  |
| --- | --- | --- | --- |
| Survey Date | Number of H fish | # of fish sampled on date (H + W) | Total fish sampled in Admiralty Creek in 2009 (96 + 21) |
| 8/12/09 | 39 | 96 | 117 |
| 8/17/09 | 9 | 21 | 117 |

The weighted pHOS equation above summing wiv \* qiv for both visits, where wiv gives the sampled counts instead of dead counts, would be

Because you are cross-multiplying, the 96 and the 21 cancel out, which leaves you with

And 48/117 is just the raw pHOS for Admiralty as a whole in 2009. In the above table, there were 48 total hatchery fish across all visits and 117 total fish sampled across all visits. Hence, you don’t need to (and really actually can’t) do any weighting of pHOS for your full study if you are including the years of 2008-2011 because you don’t have dead counts. And dead counts are not the same as sample counts.

For those who may ask about why I did not calculate a dead count-weighted pHOS for use in my study, I can tell them that 1) I couldn’t because I don’t have dead counts available for all years (missing in 2008-2011) and 2) the unweighted pHOS estimates for each stream-year observation aren’t significantly different from the weighted pHOS’s. See Figure S1 in my paper, *pHOSdiff.tiff* in “Wt\_vs\_Avg\_pHOS” folder, and *Compare\_pHOS.R* for the figure and t.test showing this. The way that I accounted for variation in sampling pressure was not by weighting with dead counts, but by dividing the number of strays detected by the number of surveys, thereby avoiding pHOS altogether.