Supplemental Results and discussion

The health of fish at sites can be estimated in several ways by using different numerical approaches. The analyses can be spatial, temporal, or spatio-temporal, but can also vary by which technique is used to compare the data. For example, traditional EEM analyses of fish health use log-linear regression models to test for differences in slope and intercept of two sets of data between two or more sites [11]. Alternatively, a generalized linear approach can also be used to examine these data, but regression diagnostics can also be used [53].

Upstream reference vs downstream exposed

The fish monitoring program to examine the impacts of industrial development in the OSR was based on the EEM design using upstream reference sites and downstream exposed sites [7,8]. Statistical differences in mean age, length, and weight in females (SI Table 4) and males (SI Table 5) were observed in lake chub at the Lower Ells compared to the Upper Ells, but were not routinely found. Among these comparisons, when statistical differences were found, females at the Lower site were younger in 2015, longer in 2014 and shorter in 2018, and lighter in 2014 (SI Table 4). Among males, fish from the Lower site were younger in 2015, but older in 2018, and longer and heavier in 2014 (SI Table 5). Differences in absolute gonad weight and liver weight were also observed in females and males. Gonad weight of females at the Lower site was higher than gonad weight of females at the Upper site in 2014, but was lower in 2015 and 2018 (SI Table 4). Similarly, absolute liver weight of females was higher at the Lower site in 2014, but was lower at the Lower site in 2015 (SI Table 5). Among males, absolute gonad weight was higher at the Lower site in 2014, but was lower in 2018. Finally, the absolute liver weight of males captured at the Lower site in 2014 was higher than among males from the Upper site (SI Table 5).

Among female and male lake chub, differences in relative gonad weight, relative liver weight, and relative body weight were observed between sites using both the OLS and GLMs (SI Table 6). However, larger livers and larger gonads (estimated as a statistical difference in y-intercepts) were observed in female lake chub in 2015 using both approaches (SI Table 6). Similarly, many potential statistical differences in slope were also observed, but no p-values were lower than 0.01 suggesting the tests of intercept were robust to these differences. Among these spatial analyses, only two sequential differences were observed: the intercept of gonad weight in female lake chub was statistically different between sites in 2014 and 2015 and statistical differences in slope (using α=0.05) were observed in relative liver weight of males in 2014 and 2015 (Supplemental Table 6). While the statistical difference in slope of male liver weight in 2014 and 2015 may not be relevant, the difference in gonad weight of females in 2014 is, however, challenging to interpret because of the size differences in fish (Supplemental Table 4). However, if the difference is real, then a similar pattern also occurred in 2015, suggesting an extent and magnitude study was required [70]. However, the difference was not apparent in 2018, suggesting the importance of some annually variable effect.

Although there were many similarities between GLM and OLS, there were also discrepancies among the statistical hypothesis tests. For example, the slope of gonad weight of female lake chub captured in 2014 compared to body weight was statistically different in the GLM, but not in the OLS (SI Table 6). In contrast, the OLS suggested a statistical difference in the y-intercept (SI Table 6). This difference in the results may be related to two factors. First, the female lake chub captured at the Upper location were shorter and lighter than females captured at the lower location violating an assumption of ANCOVA (SI Figure 8). Secondly, ln-transformation (or $\log_{10}$) of fish data prior to analysis may not normalize these data (SI Figure 10). Additional analyses using models of fish from the Upper location to predict fish measurements at the Lower site suggest more differences relative to either the OLS or GLM analyses (SI Figures 6 and 7). While there were no indications of consistent differences in OLS and GLMs, GW of male fish was consistently lower than predicted from the Upper models (SI Figure 7).

While there were statistical differences in the health metrics of fish captured at the Lower site compared to fish from the Upper site, the analyses provide little causal information and identifying an industrial signal with these comparisons remains a challenge [19]. While these differences between sites are supported by the indications of exposure to PAC related compounds in water and sediment resulting in increased whole-body PACs [14] and elevated EROD levels in the livers of downstream fish, as observed elsewhere [18] differences between sites can not be used to separate the contributions of industrial and natural stressors [13-16,24]. While PAH analysis indicates exposure in water and sediment, the differences in fish tissue are very minor suggesting little uptake or efficient elimination and little toxicological exposure irrespective of the source [14] which may also include effects of forest fires [13] The conclusions from that previous work are consistent with the results presented here including SPMDs from the Ells River [14].

Site-specific analyses using the EEM statistical approach

The challenge of overlapping stressor gradients in streams in the OSR is a primary motivator for examining sites over time. To better understand the variability across sampling years on the Ells River, we also examined the indicators of fish health within each site (Upper and Lower) across the sampling years (2013, 2014, 2015 and 2018). Year-wise comparisons (e.g., 2013 vs 2014, 2013 vs. 2015, 2013 vs. 2018) within each location were done with both OLS and GLM. Among the OLS and GLM comparisons, similar to the spatial analyses, many statistical differences were observed in both males and females (SI Tables 7-9; SI Figures 12 and 13). Ignoring differences in slope associated with p-values greater than 0.01, the most notable difference among these comparisons was observed in the relative gonad weight of females captured at the lower site. Among these comparisons, in both OLS and GLMs, the relative gonad weight of female fish captured in 2013 was higher than all other sampling years (SI Figure 12). However, there were no differences in y-intercepts among others, although the slope for 2018 was not statistically similar to 2014 or 2015. Similarly, among females captured at the Upper site the relative liver weight in 2015 was higher compared to 2013, 2014, and 2018 when compared using the GLM (SI Table 7; SI Figure 12).

Although there are statistical differences over time, by themselves these also do not allow the identification of industrial and natural sources. For example, relative liver weight among female fish captured at the Upper location in 2015 was higher than compared to all other sampling years, but the 2015 sampling event occurred between October 5-9, whereas the other years were sampled between September 19 and 26. These analyses suggest that in a study area such as the OSR where exposure environments are uncertain and contrast strongly with the environments where EEM designs are typically employed [21] suggest additional data and information are required to account for both natural and anthropogenic stressors [53]. These findings prompted additional statistical analyses of these data, including variable selection done using both environmental and industrial covariates and the Elastic Net regularized regression. Previous analyses have suggested the value of the EN for identifying the potential sources of variation in Canada's Oil Sands Region [32,37,51] and its likely utility for other environmental indicators. However, recent analyses also suggest the potential influence of facilities outside the Ells River basin on benthos residing in the mainstem [37].

Among the GLM analyses, an additional comparison was also done and was analogous to the regression diagnostic technique also used to identify differences between groups of fish described in the main text. Among these results, the statistical differences correspond exactly in all cases but one (gonad weight of males at the lower site; SI Table 8) with the comparisons using residual variation and the percentiles estimated from the double bootstrap (Figures 2 and 3). However, the p-value for this comparison was 0.043 suggesting the effect was not large. The p-value is also larger than has been used in some analyses to account for robustness of tests of intercept to minor violations of the assumption of parallel slopes for ANCOVA and its analogs. Importantly, this analysis shows the correspondence of the inferential statistics based on p-values and those based on confidence intervals, but, similar to the supplemental results described already, also suggest that additional work is required to account for potential annual differences in environmental conditions. These particular differences at each site in 2018 compared to 2013-2015 also prompted the additional analyses described in more detail in the main text.