



**Supplemental Figure 1.** *Problems of sampling a dataset. A classic normal distribution (blue), sampled randomly with a large (red) and small (black) number of data points. Note the significant bias to the positive side for the small population size, whereas any bias for the large population appears to be smaller. This is similar to the bias that can be seen when training machine-learning methods on datasets that are too small.*

**Probability(x) x values**