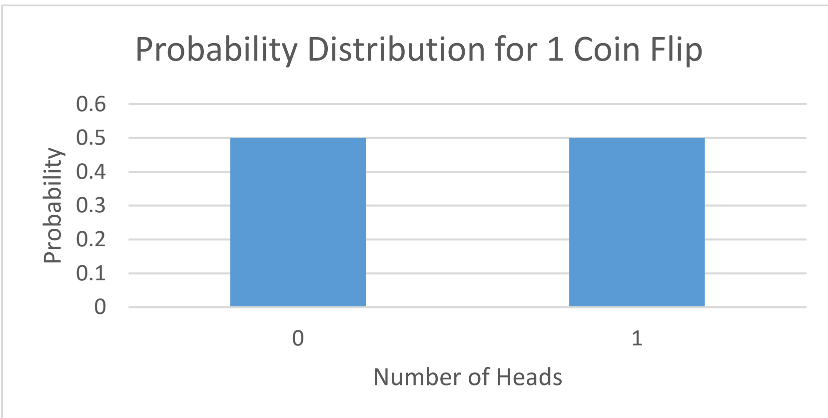
PROBABILITY DISTRIBUTION

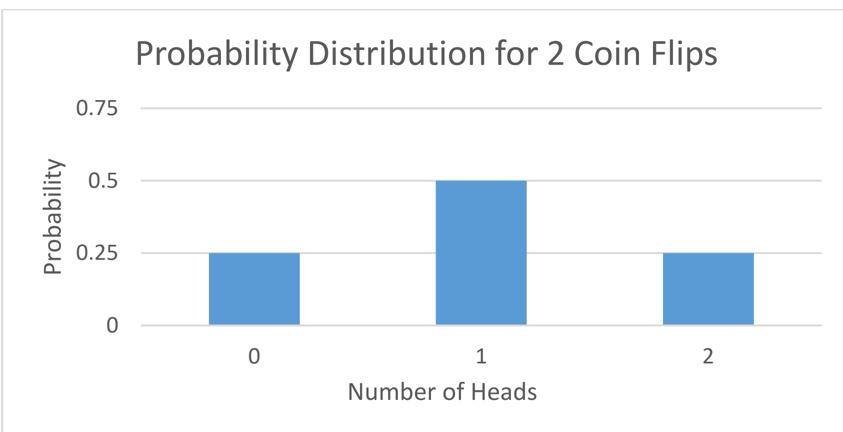
A probability distribution provides the likelihood of occurrence for possible outcomes. The probability for some events, such as flipping a coin or rolling a pair of dice, is well known, while the probability for other events, such as the amount of rainfall in a hurricane, must be estimated.

Coin Flip Example

From probability, we know that if a coin is flipped one time, there is a 0.5 probability of the outcome containing 0 heads and a 0.5 probability of there being 1 heads. If the coin is flipped two times, there is a 0.25 probability of there being 0 heads, a 0.5 probability of there being 1 heads, and a 0.25 probability of there being 2 heads. These probabilities can be displayed in either a chart or a histogram.



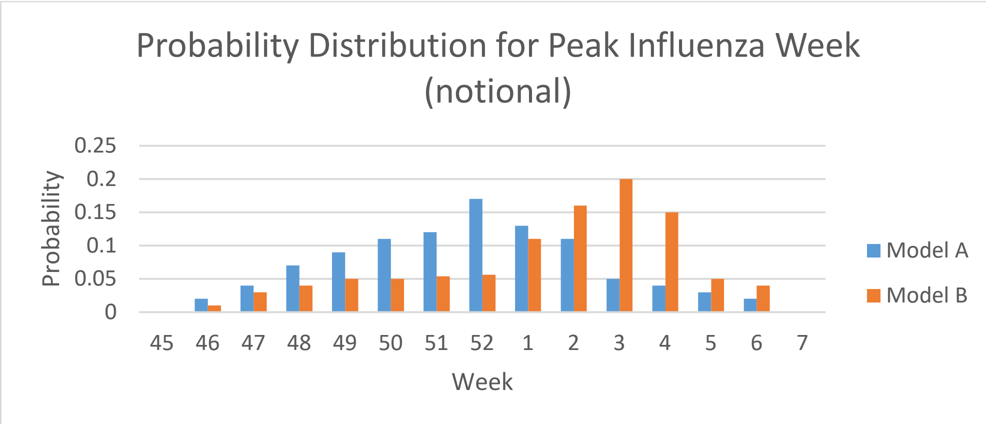
Heads	Probability
0	0.5
1	0.5



Heads	Probability
0	0.25
1	0.5
2	0.25

Peak Influenza Week Example

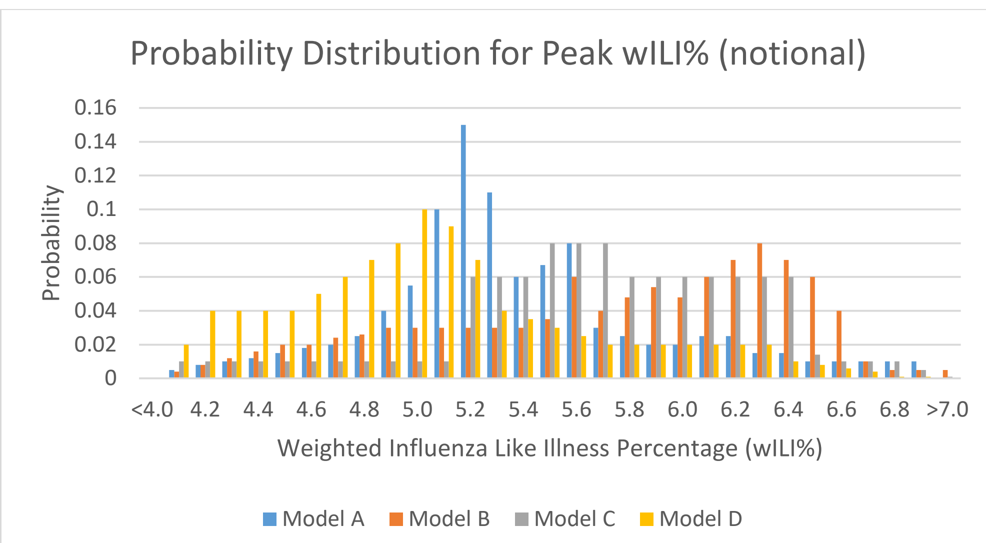
The probability distribution for the peak week of the current influenza season is currently (as of October 2018) unknown. Modelers are asked to provide probability distributions for when the peak week occurs. Different models are likely to yield different probability distributions.



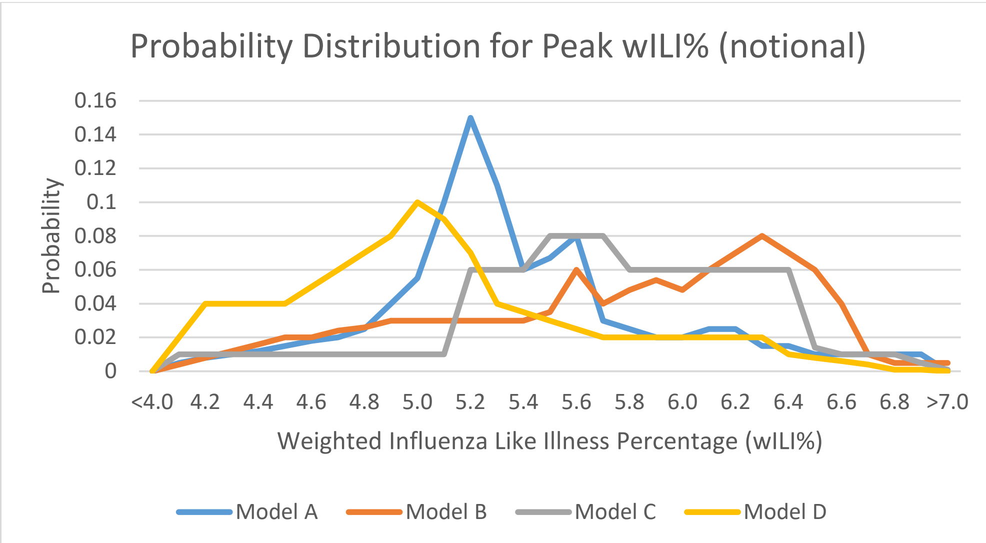
Peak Weighted Influenza Like Illness percentage (wILI%) Example

The CDC Influenza Challenge uses wILI% as a measure of influenza intensity. Basically, ILI% represents the percentage of patients who visit participating health care facilities with symptoms that suggest influenza or an influenza-like illness. It is “weighted” based on state population. For example, the ILI% reported by California is given more weight than the ILI% reported by Idaho when the states are combined into the National wILI%.

Similarly, the true probability of the wILI% during the peak week of the current influenza season is unknown. Modeling teams will provide probability distributions for their expectations.



With four models, the chart becomes busy and a little challenging to get a clear picture of the probability distribution predicted by each model. I will typically present 9 different models and the x-axis may go from 0 to 13. To improve readability, I depict the probabilities as line charts rather than histograms.



For this project, willing participants are asked to select which model they believe is likely to score the most points (essentially, the model that is most accurate). There is no prescribed method for participants to use to select the model. As an example, Table 1 shows how I would select based on how they compare to my expectations. Similar to the "Area Under the Curve" taught in calculus, I am mentally comparing the area under each curve for the range of outcomes I consider most likely. Of course, each participant is likely to come up with their own approach for selecting models.

Table 1. Examples of how my expectations for wILI% would lead to model selection

What I think will happen	Which model I would select
Peak wILI% to be around 5.0, but more likely to be lower than higher	Model D
Peak wILI% to be around 5.0, but more likely to be higher than lower	Model A
Peak wILI% to be around 6.0, but more likely to be lower than higher	Model C
Peak wILI% to be around 6.0, but more likely to be higher than lower	Model B
Peak wILI% to be between 4.2 and 4.8	Model D
Peak wILI% to be between 5.0 and 5.6	Model A
Peak wILI% to be between 5.4 and 6.4	Model C
Peak wILI% to be between 6.4 and 7.4	Model B