

# THINKING PROBABILISTICALLY II

WARREN R HUGHES – August 2021

In July 2020, the WHO organized a fact-finding mission to Wuhan in China to investigate the origin of the COVID-19 virus. Then in January 2021, a team of international and Chinese experts spent two weeks investigating the outbreak and concluded that four scenarios were possible. These are summarized below in Table 1.

**TABLE 1: WHO POSSIBLE SCENARIOS AND LIKELIHOODS FOR THE ORIGIN OF COVID-19**

#	Scenario	Likelihood
1	Direct transmission from bat to human	Possible to likely
2	Transmission through an intermediate animal	Likely to very likely
3	Transmission through frozen food	Possible
4	Transmission through a laboratory incident	Extremely unlikely

Here is a very good example of THINKING PROBABILISTICALLY, but we can only imagine the arguments before decisions on “Likelihood” were arrived at! Here we re-examine this problem using the methodology of “more likely”. The routine calculations are shown below in Table 2 and are discussed further in *Structuring Probability Assessments* available at <https://doi.org/10.17265/1537-1506/2020.05.003> or on this website. Three hypotheses concerning the virus origin are re-stated and considered here ranked in order of increasing likelihood:

1. NonCh - The virus originated outside of China and is the least likely hypothesis (Non-Chinese).
2. WuMkt - The virus emerged naturally (animal to human) from a Wuhan wet-market.
3. WuLab -The virus initially escaped from a Chinese laboratory in Wuhan (Wuhan Lab Escape) and is considered here to be the most likely of the three hypotheses.

If the virus originated in China, hypotheses 2 and 3 above are the only possibilities we consider although other Chinese origins are possible with the Wuhan wet-market an “amplifying phenomenon”. The Wuhan wet-market is commonly understood as being the most likely origin of the virus, but recent investigations have increased the likelihood of a lab escape. Pairwise values and resulting probabilities could vary depending on the decision-maker’s (DM’s) judgments. Many people believe that the virus emerged almost certainly in China. We examine possible judgments below starting with a Chinese origin as “10 times” more likely than a Non-Chinese alternative. The two China origin hypotheses are initially judged to be equally likely (Table 2) but subsequently WuLab is seen as “more likely” in a sensitivity analysis in Table 3.

**TABLE 2: PROBABILITIES OF COVID-19 VIRUS ORIGIN**

Virus Origin	Pairwise Value	Compound Likelihood	Probability	Percent Probability <sup>#</sup>
NonChinese	1.00 (base value)	1.00	1/21 = 0.048	5
WuMkt	10.00 (10 x more likely)	1.00 x 10 .0 = 10.00	10/21 = 0.476	48
WuLab	1.00 (equally likely)	10.00 x 1.0 = 10.00	10/21 = 0.476	48
		<b>21.00</b>	<b>1.000</b>	<b>101</b>

<sup>#</sup>Percentage probabilities may not add to 100 due to rounding

Note that these initial calculations can be considered to show a 95% Chinese origin possibility. The sensitivity analysis of Table 3 below shows how the “10 times” more likely judgment affects the calculations. Then probabilities are calculated assuming the WuLab source is a little more likely (25% “more likely”) than a naturally occurring origin from animals to humans at the Wuhan wet-market (WuMkt). The Table 2 calculations are reproduced in the first line of Table 3 below.

**TABLE 3: PROBABILITIES ON VIRUS ORIGIN FOR VARYING PAIRWISE JUDGMENTS**

Scenario	Pairwise Values	Probabilities <sup>#</sup>			Percent Probabilities <sup>#</sup>		
		NonCh	WuMkt	WuLab	NonCh	WuMkt	WuLab
China Origin 10 times ML*	1, 10, 1	.048	.476	.476	5	48	48
China Origin 5 times ML	1, 5, 1	.091	.455	.455	9	46	46
China Origin 7.5 times ML	1, 7.5, 1	.063	.469	.469	6	47	47
WuLab ML than WuMkt	1, 5, 1.25	.082	.408	.510	8	41	51
As above but 7.5 times ML	1, 7.5, 1.25	.056	.420	.524	6	42	52

\*ML = more likely <sup>#</sup>Probabilities may not add to 1.0 or 100 due to rounding

In summary we can conclude that a Non-Chinese origin probability is less than 10%, a Wuhan wet-market origin will be in the 40% range, and a Wuhan Lab Escape in a high 40% to low 50% range. These values define the “ballpark” for a final probability distribution assessment by the DM which could be the last line in Table 3 above. Alternatively, the DM may decide the probabilities of the respective hypotheses to be 5%, 45% and 50%, avoiding any charges of spurious accuracy.

Another point to note is that whether we take a large “more likely” difference to be 5 or 10 times “more likely” does not materially alter the resulting probabilities, apart from the lowest probability event which this (large) pairwise value is intended to minimize. This is reassuring since it would be hard to accurately estimate exactly how many times one event is “more likely” than another when the difference is large, say in the 3+ range. This means an average value such as 7.5 times (i.e., between 5 and 10 times) “more likely” will get the DM into the right “ballpark” prior to a final probability assessment by him or her. Alternatively, the DM could input this pairwise value last into a spreadsheet and raise (or lower) it progressively to a level consistent with his/her assessment of the resulting chance of the least likely event. For example, 5% in the virus origin case. Using a spreadsheet to calculate the probabilities makes this procedure very simple. Of course, the resulting “more likely” values necessary to achieve the desired probabilities must accord with the DM’s judgments although typically these will be tentative or rudimentary.

The ordering of the scenarios in terms of likelihood as done initially means estimates of the “more likely” values between adjacent events would typically be minimal values in the 1+ to 2 range (e.g., 1.25 in Table 3). Unlike the 5 or 10 value above, these judgments should be less demanding of the DM, especially if the relative likelihoods for the events are close. Of course, some comparisons of events later in the ordering may require pairwise values of 2 or more on occasion. The routine calculations as shown in Table 2 above are very simple allowing a sensitivity analysis on likelihoods for differing pairwise values (as in Table 3) to be easily undertaken. A good review of the lab leak hypothesis is given in the journal *Nature* at doi: <https://doi.org/10.1038/d41586-021-01529-3>.