Estimating the risk of COVID-19 being present at events

Proportion of population now infected	0.02
Sensitivity of lateral-flow tests	0.6
Compliance with test+isolate request	0.2
Proportion of infected people who come	0.88
Probability of a random person being healthy	0.98
Probability of an attendee being healthy	0.9824

Cells highlighted in yellow are inputs

		p(N infected)								P(1<=N<=8 infected)	Sanity check
Event Size	**p(all healthy)**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**		
5	92%	8%	0%	0%	0%	0%				8%	1.000000
10	84%	15%	1%	0%	0%	0%	0%	0%	0%	16%	1.000000
20	70%	25%	4%	0%	0%	0%	0%	0%	0%	30%	1.000000
30	59%	32%	8%	1%	0%	0%	0%	0%	0%	41%	1.000000
40	49%	35%	12%	3%	0%	0%	0%	0%	0%	51%	1.000000
50	41%	37%	16%	5%	1%	0%	0%	0%	0%	59%	1.000000
60	34%	37%	20%	7%	2%	0%	0%	0%	0%	66%	0.999999
70	29%	36%	22%	9%	3%	1%	0%	0%	0%	71%	0.999996
80	24%	35%	25%	11%	4%	1%	0%	0%	0%	76%	0.999988
90	20%	33%	26%	14%	5%	2%	0%	0%	0%	80%	0.999968
100	17%	30%	27%	16%	7%	2%	1%	0%	0%	83%	0.999927
110	14%	28%	27%	18%	8%	3%	1%	0%	0%	86%	0.999848
120	12%	26%	27%	19%	10%	4%	1%	0%	0%	88%	0.999707
130	10%	23%	27%	20%	12%	5%	2%	1%	0%	90%	0.999472
140	8%	21%	26%	21%	13%	6%	3%	1%	0%	92%	0.999102
150	7%	19%	25%	22%	15%	8%	3%	1%	0%	93%	0.998544

Rationale for the risk calculations

ONS publish population-level stats for the number of people infected with COVID-19 on a given date.
Using the regional numbers and the age-demographic numbers we estimate an overall infection rate for our likely audience.
This is expressed as the proportion of the population now infected and is a number between zero and one.

Many infected people have no symptoms.
We ask people to do a lateral-flow test before coming to an event. Some proportion of them will comply with this request.
Lateral-flow tests are not perfect. Research suggests that they detect about 60% of infections.
We express both numbers as a proportion between zero and one.

The spreadsheet calculates the probability of a randomly-chosen member of the audience being healthy. Call this p(healthy)

Now, for a given audience size A, the probability of **all** being healthy is p(healthy) raised to the power A
because this is only true if each member is themselves healthy.

In most cases it is likely that more than one person has COVID-19, so we estimate the probability for each number from 1 to 8
As an example, the probability of 3 people being infected from an audience of 50 is:
p(healthy)**(50-3) * (1-p(healthy))**3 * COMBI(50,3)
In other words, the probability that exactly 47 people are healthy times the probability that the other 3 are infected, times the number of combinations of 3 people chosen from 50

The last two columns in the spreadsheet form a sanity check, by adding up the probabilities of zero to eight people being infected.
For most audience sizes of interest the sum should be very close to 1.

Andrew Findlay

NOTE: all my stats knowledge is self-taught – mostly for COVID-19 risk analysis – so please let me know if I have got this wrong!