Supplementary materials for optimal diagnostic test allocation strategy during the COVID-19 pandemic and beyond

Jiacong Du, Lauren J Beesley, Seunggeun Lee, Xiang Zhou, Walter Dempsey and Bhramar Mukherjee

Figure S1: Test positive rate for different underlying true number of cases. The surveillance test allocation strategy is given under the assumption that the true number of cases is 10,000 and the probability of being asymptomatic for an infected case is 0.55 ($\bar{t_a} = 0.55$). The red horizontal line is the threshold for disease outbreak.

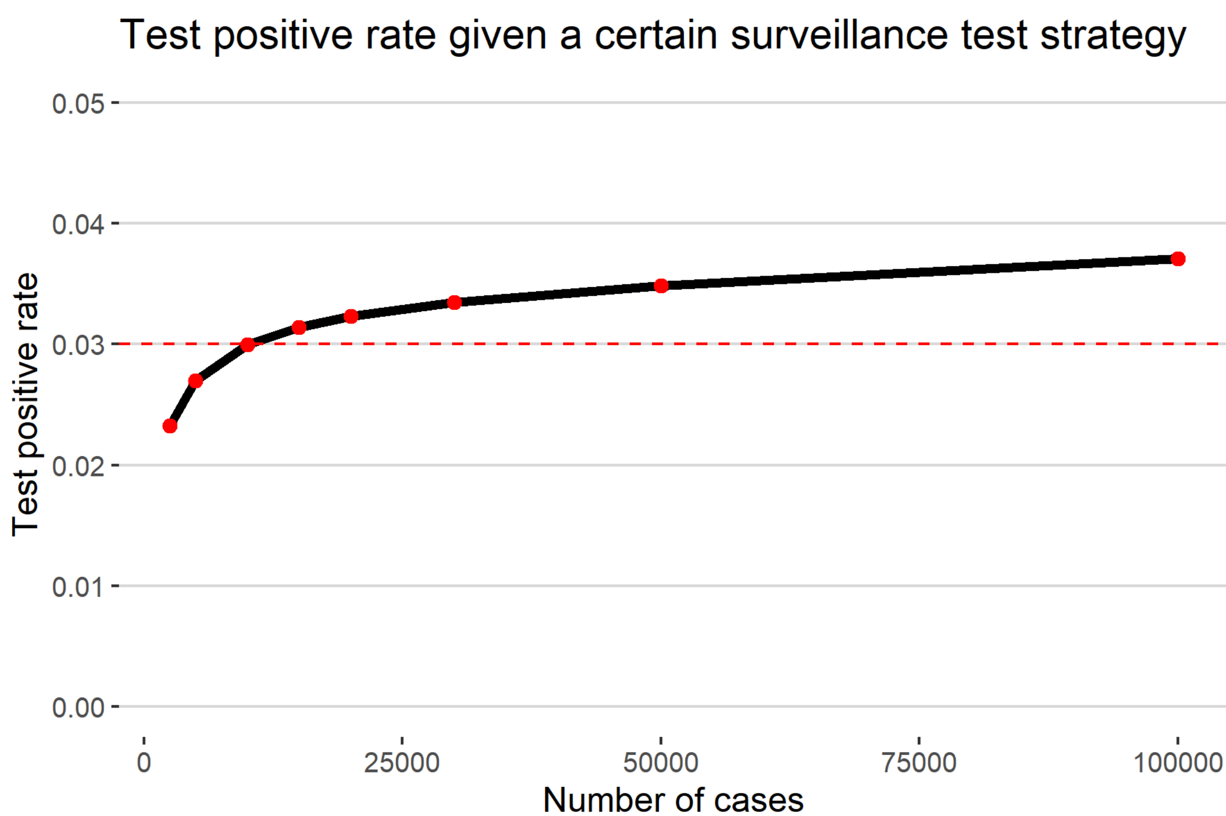


Table S1: Estimated probabilities of having severe, mild and no symptoms for the infected individual when the marginal probability of being asymptomatic after infected is 0.55 ($\bar{t_a} = 0.55$), and the probability of developing severe symptoms is 1/3 of developing mild symptoms ($\bar{t_s} = 1/3\bar{t_m}$). $t_j(z)$ is the probability of developing symptom j conditional on a person being infected in age group z. Estimates are obtained using the distribution of age in the hospitalized patients and overall cases in New York City. $f_j(z)$ is the probability of developing symptom j in age group z for the uninfected person. Estimates are obtained by using the number of flu hospitalizations and cases from New York State Department of Health 2019-2020 Flu Monitoring Archives.

age	$t_s(z)$	$t_m(z)$	$t_a(z)$	$f_s(z)$	$f_m(z)$	$f_a(z)$
0-17	5.816E-02	5.186E-03	9.367E-01	9.619E-06	2.821E-04	9.997E-01
18-44	3.571E-01	4.307E-02	5.998E-01	9.642E-06	1.287E-04	9.999E-01
45-64	5.786E-01	1.783E-01	2.431E-01	3.024E-05	1.178E-04	9.999E-01
65+	4.041E-01	4.608E-01	4.608E-01	7.907E-05	1.017E-04	9.998E-01

Table S2: Tests allocated to each symptom and age group from the proposed optimal strategy and other competing methods near the peak of the pandemic under the detecting mode, assuming that the number of true infected cases is 175,000 at generation g. The population size is 8 million. $\bar{t_a}$ is the probability for a person of being asymptomatic after being infected.

Symptom	Severe				Mild				Asymptomatic			
Age	0-17	18-49	50-64	65+	0-17	18-49	50-64	65+	0-17	18-49	50-64	65+
Limited number of tests as 50,000; $\bar{t_a}$=0.55												
Population	236	3280	6605	9738	2974	27307	21456	8577	1916788	3409411	1651937	941684
Detecting mode	236	3280	6605	9738	0	104	21456	8577	0	0	0	0
Risk-based	236	3280	6605	9738	0	104	21456	8577	0	0	0	0
Symptom-based	236	3280	6605	9738	1486	13644	10721	4285	0	0	0	0
Severe-only	236	3280	6605	9738	11	103	81	32	7239	12876	6238	3556
Universal random	1	20	41	60	18	170	134	53	11979	21308	10324	5885
Sufficient number of tests as 200,000; $\bar{t_a}$=0.55												
Population	236	3280	6605	9738	2974	27307	21456	8577	1916788	3409411	1651937	941684
Detecting mode	236	3280	6605	9738	2974	27307	21456	8577	119836	0	0	0
Risk-based	236	3280	6605	9738	2974	27307	21456	8577	0	0	0	119822
Symptom-based	236	3280	6605	9738	2974	27307	21456	8577	28999	51582	24992	14247
Severe-only	236	3280	6605	9738	67	616	484	193	43268	76961	37289	21256
Universal random	5	82	165	243	74	682	536	214	47919	85235	41298	23542

Table S2 (continued): Tests allocated to each symptom and age group from the proposed optimal strategy and other competing methods near the peak of the pandemic under the detecting mode, assuming that the number of true infected cases is 175,000 at generation g. The population size is 8 million. $\bar{t_a}$ is the probability for a person of being asymptomatic after being infected.

Symptom	Severe				Mild				Asymptomatic			
Age	0-17	18-49	50-64	65+	0-17	18-49	50-64	65+	0-17	18-49	50-64	65+
Limited number of tests as 50,000; $\bar{t_a}$=0.90												
Population	66	754	1506	2221	1072	6405	4918	1980	1918860	3432839	1673574	955797
Detecting mode	66	754	1506	2221	1072	6405	4918	1980	31093	0	0	0
Risk-based	66	754	1506	2221	1072	6405	4918	1980	0	0	0	31072
Symptom-based	66	754	1506	2221	1072	6405	4918	1980	7470	13364	6515	3721
Severe-only	66	754	1506	2221	6	36	27	11	10907	19514	9513	5433
Universal random	0	4	9	13	6	40	30	12	11992	21455	10459	5973
Sufficient number of tests as 200,000; $\bar{t_a}$=0.90												
Population	66	754	1506	2221	1072	6405	4918	1980	1918860	3432839	1673574	955797
Detecting mode	66	754	1506	2221	1072	6405	4918	1980	181089	0	0	0
Risk-based	66	754	1506	2221	1072	6405	4918	1980	0	0	0	181072
Symptom-based	66	754	1506	2221	1072	6405	4918	1980	43534	77883	37969	21684
Severe-only	66	754	1506	2221	26	156	120	48	46906	83916	40910	23364
Universal random	1	18	37	55	26	160	122	49	47971	85820	41839	23894

Table S3: Tests allocated to each symptom and age group from the proposed optimal strategy and other competing methods after the peak of the pandemic under the surveillance mode, assuming that the number of true infected cases is 10,000 at generation g. The population size is 8 million. $\bar{t_a} = 0.55$ is the probability for a person of being asymptomatic after being infected. The total number of available tests is 200,000. The actually number of used tests may not be added up to the exact number of available tests due to rounding of numbers.

Symptom	Severe				Mild				Asymptomatic				Used tests
Age	0-17	18-49	50-64	65+	0-17	18-49	50-64	65+	0-17	18-49	50-64	65+	
Population	30	218	425	628	680	1978	1412	582	1919288	3437802	1678161	958789	
Surveillance mode	30	218	425	628	342	1088	763	300	19327	2521	24479	52956	103077
Risk-based	30	218	425	628	680	1978	1412	582	0	0	0	194042	199995
Symptom-based	30	218	425	628	680	1978	1412	582	46587	83446	40734	23273	199993
Severe-only	30	218	425	628	16	49	35	14	47677	85398	41687	23817	199994
Universal random	0	5	10	15	17	49	35	14	47982	85945	41954	23969	199995
Detecting mode	30	218	425	628	680	1978	1412	582	194049	0	0	0	200002

Figure S2: Tests allocated to each symptom and age group near the peak of the pandemic under the detecting mode, assuming either 50,000 or 200,000 tests are available and that the number of true infected cases is 175,000 in a region of 8 million people. Test allocation strategy is obtained with and without the false positive tests. The false positive rate is 0.01.

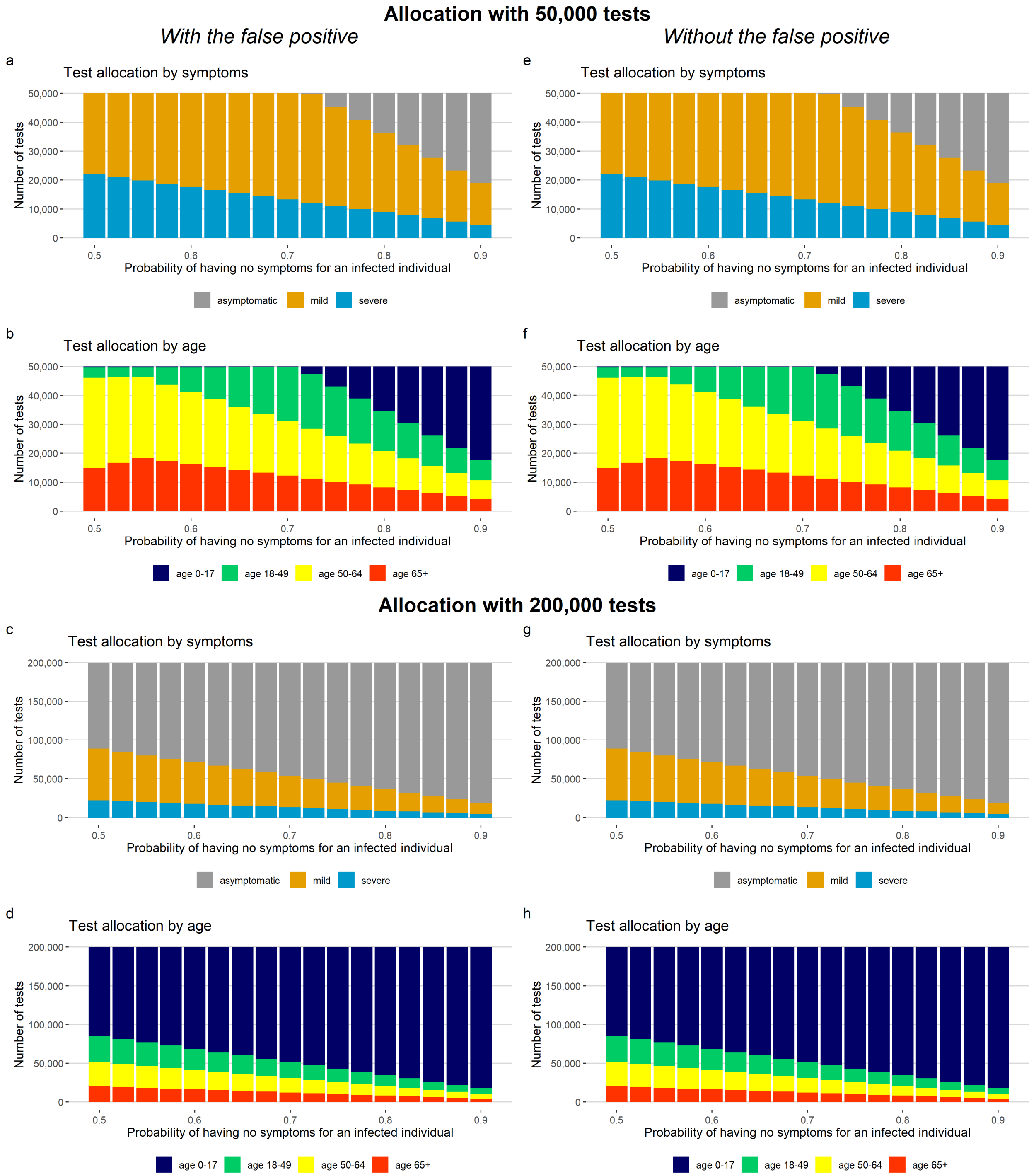


Figure S3: Tests allocated to each symptom and age group for surveillance past the peak of the pandemic, assuming 200,000 tests are available and the number of true infected cases is low (10,000) in a region of 8 million people. The test positive rate for the disease outbreak is 0.03. Test allocation strategy is obtained with and without the false positive tests. The false positive rate is 0.01.

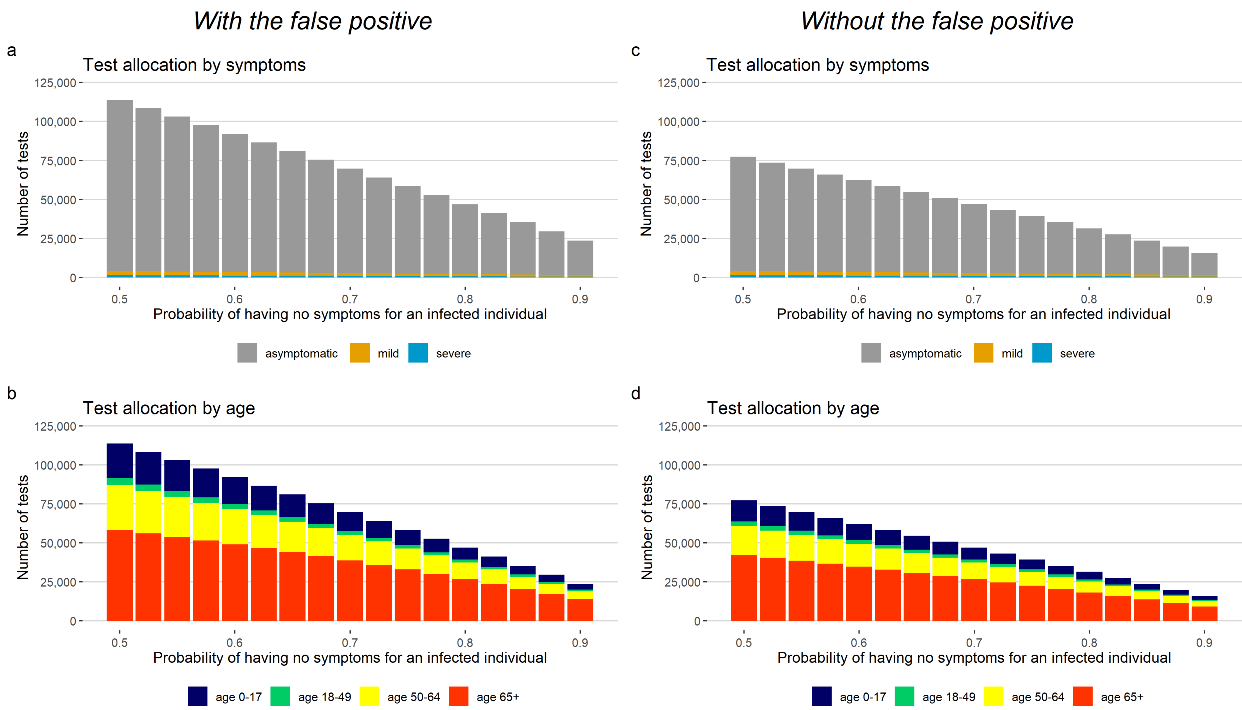
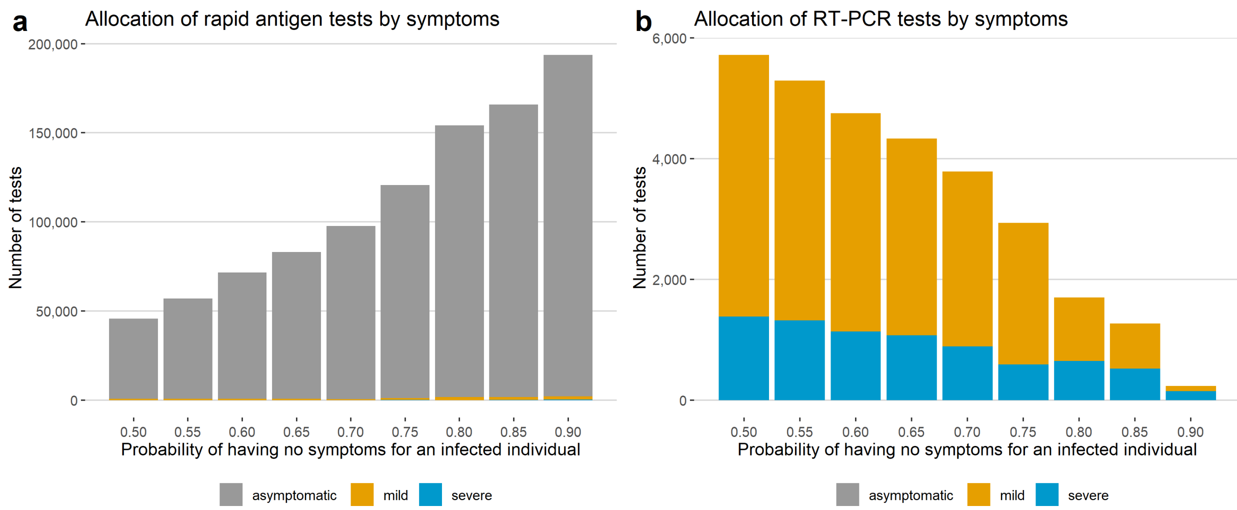


Table S4: The performance of surveillance test allocation strategies considering with and without false positive tests. The *True+false strategy* (with false positive tests) and the *True-only strategy* (without false positive tests) are obtained assuming that 200,000 tests are available in a region of 8 million people, the true infected cases is 10,000 and the probability for a person of being asymptomatic after being infected ($\bar{t_a}$) is 0.5. The test positive rate for the disease outbreak is 0.03 (c). Two strategies are then applied to the population to obtain the number of true positive tests, false positive tests, total number of tests, and test positive rate. The test false positive rate (α) is 0.01 and the test false negative rate (β) is 0.3. The mean and standard deviation are calculated based on 100 repetitions.

	True positive Mean (sd)	False positive Mean (sd)	Total # of positive Mean (sd)	Test positive rate Mean (sd)
True+false strategy	2306 (47.5)	1106 (31.6)	3411 (56.2)	0.03 (5e-04)
True-only strategy	2316 (47.0)	742 (25.4)	3058 (51.4)	0.04 (7e-04)

Figure S4: The number of each type of test allocated to each symptomatic group. The total budget for testing is 1 million. The price of a single antigen test and RT-PCR test are \$ 5 and \$ 135, respectively. The number of infected cases is assumed to be $10,000$ in a region of 8 million people.



Relationship between test positive rate and disease prevalence

Let r be the ratio of selection probability in the infected population versus the uninfected population; that is, $r = \frac{P(S_i^{(g)}=1|D_i^{(g)}=1)}{P(S_i^{(g)}=1|D_i^{(g)}=0)}$. Let β, α, c be the test false negative rate, false positive rate and the outbreak threshold test positive rate. If $r > 1$, $1-\beta > c$ and $1-\beta > \alpha$, then the disease prevalence $P(D_i^{(g)} = 1) < \frac{c-\alpha}{(1-\beta-c)r+(c-\alpha)}$.

Proof:

$$\begin{aligned}
P(D_i^{\star(g)} = 1|S_i^{(g)} = 1) &= \frac{P(D_i^{\star(g)} = 1, S_i^{(g)} = 1)}{P(S_i^{(g)} = 1)} \\
&= \frac{\sum_{d=0,1} P(D_i^{\star(g)} = 1, S_i^{(g)} = 1, D_i^{(g)} = d)}{\sum_{d=0,1} P(S_i^{(g)} = 1, D_i^{(g)} = d)} \\
&= \frac{\sum_{d=0,1} P(D_i^{(g)} = d)P(S_i^{(g)} = 1|D_i^{(g)} = d)P(D_i^{\star(g)} = 1|S_i^{(g)} = 1, D_i^{(g)} = d)}{\sum_{d=0,1} P(D_i^{(g)} = d)P(S_i^{(g)} = 1|D_i^{(g)} = d)}
\end{aligned}$$

Under assumption $A.3$, we have $P(D_i^{\star(g)} = 1|S_i^{(g)} = 1, D_i^{(g)} = 1) = 1-\beta$ and $P(D_i^{\star(g)} = 1|S_i^{(g)} = 1, D_i^{(g)} = 0) = \alpha$. Let $r = \frac{P(S_i^{(g)}=1|D_i^{(g)}=1)}{P(S_i^{(g)}=1|D_i^{(g)}=0)}$, we have:

$$\begin{aligned}
P(D_i^{\star(g)} = 1|S_i^{(g)} = 1) &= \frac{(1-\beta)rP(D_i^{(g)} = 1) + \alpha P(D_i^{(g)} = 0)}{rP(D_i^{(g)} = 1) + P(D_i^{(g)} = 0)} \\
&= \frac{(1-\beta)rP(D_i^{(g)} = 1) + \alpha\big(1 - P(D_i^{(g)} = 1)\big)}{rP(D_i^{(g)} = 1) + \big(1 - P(D_i^{(g)} = 1)\big)}
\end{aligned}$$

Note that $P(D_i^{\star(g)} = 1|S_i^{(g)} = 1)$ is a function of the disease prevalence $P(D_i^{(g)} = 1)$. When $P(D_i^{\star(g)} = 1|S_i^{(g)} = 1) < c$, we invert the inequality and solve for the boundary for $P(D_i^{(g)} = 1)$. When $r > 1$, the denominator is bigger than 0, and

$$\begin{aligned}
&(1-\beta)rP(D_i^{(g)} = 1) + \alpha\big(1 - P(D_i^{(g)} = 1) < c\Big(rP(D_i^{(g)} = 1) + \big(1 - P(D_i^{(g)} = 1)\big)\Big) \\
&\big((1-\beta-c)r + (c-\alpha)\big)P(D_i^{(g)} = 1) < c - \alpha
\end{aligned}$$

Since $(1-\beta-c)r + (c-\alpha) > 1-\beta-\alpha > 0$, we have:

$$P(D_i^{(g)} = 1) < \frac{c-\alpha}{(1-\beta-c)r + (c-\alpha)}$$