

Sequential informed pooling approach to detect SARS-CoV2 infection

Renato Millionsi, University of Padova, Italy

Cinzia Mortarino, Department of Statistical Sciences, University of Padova, Italy^(*)

Abstract

The alarming spread of the pandemic coronavirus disease 2019 (COVID-19) caused by the SARS-CoV-2 virus is requiring several measures to reduce the risk of contagion. Every successful strategy in controlling SARS-CoV2 infection depends on the timely viral diagnosis which should include asymptomatic carriers. Consequently, strategies to increase the throughput for clinical laboratories to conduct large-scale diagnostic testing are urgently needed. Here we support the hypothesis that standard diagnostic protocol for SARS-CoV-2 virus could be conveniently applied to pooled samples obtained from different subjects. We suggest that a two-step sequential pooling procedure could identify positive subjects, ensuring at the same time significant benefits of costs and time. Simulation data are used to assess the efficiency, in terms of number of required tests, both for random assignment of the subjects to the pools and for situations when epidemiological and clinical data are used to create an "informed" version of the pooling. Different scenarios are examined in the simulations to measure the effect of different pool sizes and different values for the virus frequency. Our results allow to customize the pooling strategy according to the specific characteristics of the cohort to be tested.

Introduction

The pandemic coronavirus disease 2019 (COVID-19) is caused by the SARS-CoV-2 virus. The infection is predominantly transmitted through large droplets and by contact with infected surfaces or fomites. The alarming spread of the infection and the severe clinical disease that it may cause have led to several measures to reduce the risk of contagion. Active case detection, rapid case isolation and contact quarantine, as well as rigorous application of infection control practices are successful strategies in controlling SARS-CoV2 infection outbreaks. The first step on which these strategies are set is the viral diagnosis. The overloading to which the laboratories are currently subjected causes a cascade delay of all virus containment procedures with potential dramatic results for the prevention of the infection.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

^(*)Corresponding author: cinzia.mortarino@unipd.it

In most countries, testing for COVID-19 is mainly restricted to people with symptoms. However, a large percentage of asymptomatic subjects is estimated to exist. Asymptomatic spread has likely driven the silent growth of SARS-CoV2 epidemic which emerged only when the health system began to collapse. Asymptomatic cases play a role in the transmission and thus pose a significant infection control challenge. How much asymptomatic individuals affect the virus diffusion is actually considered a crucial task to evaluate [1]. Tracing contacts of known positive cases, travel bans and social distancing are the main strategies to reduce the risk of contagion due to asymptomatic subjects. A widespread testing strategy to screen asymptomatic subjects could be useful in reducing transmission of SARS-CoV2, but this approach is highly challenging taking into account of the amount of work, time and costs that it would entail.

For this reason, we propose here a pre-screening strategy which should increase the capacity for clinical laboratories to conduct large-scale diagnostic testing, enough to screen a significant portion of the asymptomatic population.

SARS-CoV2 is an enveloped viruses containing a single strand of positive-sense RNA and its diagnostic protocol is a RT-PCR assay, as previously described in details [2, 3]. Briefly, SARS-CoV2 have been detected from a variety of upper and lower respiratory sources including throat, nasal nasopharyngeal (NP), sputum, and bronchial fluid [4, 5]. Oropharyngeal (OP) and NP swabs are the most frequently used samples. The sampling is carried out using two distinct swabs which can be inserted in the same test tube containing the viral transport medium to increase the yield for the RT-PCR analysis [6]. Total RNA is extracted and SARS-CoV2 target genes are simultaneously amplified and tested during the quantitative RT-PCR assay.

Recently, Hogan et al (2020) [7] performed a retrospective study on SARS-CoV-2 based on sample pooling. “Pooling” means that swab samples taken from different subjects can be combined before the RNA extraction phase. These Authors used 2888 samples from nasopharyngeal and bronchoalveolar lavages that were collected between January 1, 2020, and February 26, 2020, from subjects who had not been tested for SARS-CoV-2. Nine or ten samples were pooled, and screening was performed by RT-PCR. A total of 292 pools were screened and the confirmed positivity rate for SARS-CoV-2 was 0.07% (2/2888). The aim of pooling is to reduce the number of test kits used, significantly shortening the time and costs of analysis. As reported by on-line media (<https://www.tabletmag.com/sections/science/articles/pooling-covid-19-israel>), a pooling approach has also recently being successfully used in Israel but, to the best of our knowledge, no studies have been published yet. Scientists from the Technion Israel Institute of Technology stated that when 64 samples are mixed together, the analysis on the pool test gives a positive result even if just one of the swabs came from an infected subject. The pooled sample can give laboratory results of equal quality of those with individual testing because there are no changes to the RT-PCR diagnostic protocol.

The basic concepts for understanding the pooling strategy are simple: 1) only a pool made up of all negative samples will give a negative result for the pool analysis; 2) a single positive sample within a pool makes the

result of the pool analysis positive. If the pool is positive, it is necessary to proceed to individual testing, to identify the true positives (TP) and the false positives (FP: i.e., a negative subject whose swab has been mixed with at least one positive swab). As all individual samples in a negative pool are considered as true negative (TN), the pooling approach significantly reduces time and cost when a large proportion of pools tests negative. However, it is clear that the effectiveness of pooling is inversely proportional to the frequency of the virus in the selected cohort and, as we will demonstrate more precisely in the results section, this approach can be inefficient or even counter-productive if the presence of the virus is high.

The aim of this paper is to i) propose a two-step sequential pooling strategy, ii) identify the variables for which the pooling method can be more or less effective and to iii) develop strategies to further improve this approach. For this purpose, we began by identifying the main variables which should be included in our model. The first and perhaps most important variable, as already mentioned, is the frequency of the virus. Unfortunately, this information is not known *a priori* but it can be estimated. The second variable is the effectiveness of the clinical and epidemiological criteria that can be adopted to create the pools, compared to an analysis in which these pools are created randomly. The third variable is the size of the pool. We have taken into consideration a wide range of scenarios to adjust the variables to give the best result with fewer tests.

As pooling strategies that can improve the pooling approach, we compared alternative methods to create pools and evaluate their different performances according to the different conditions of the variables previously described. Our data suggest that pre-screening strategy based on the use of a sequential informed pooling approach ensures that, in the most favourable conditions with low virus frequency, the number of tests can drop to 20% of the number of test required by individual testing. Higher virus frequencies still make sequential pooling efficient, provided pool size is decreased and/or reliable epidemiological and clinical data are used to create pools.

Methods

In the proposed procedure, for each patient involved in the study, three distinct swabs will be performed, following the standard protocols. The first and second swab will be used for the creation of the pools that we will define below. The third swab will be kept and eventually used as a validation test.

The swab used for the creation of a pool "H" (sample 1) is firstly placed in a single sample tube and subsequently transferred in a pooled sample tube together with the sample 1 of other subjects (Figure 1). It is important that the pooled sample could be analyzed with the same standard procedures that are applied for individual samples.

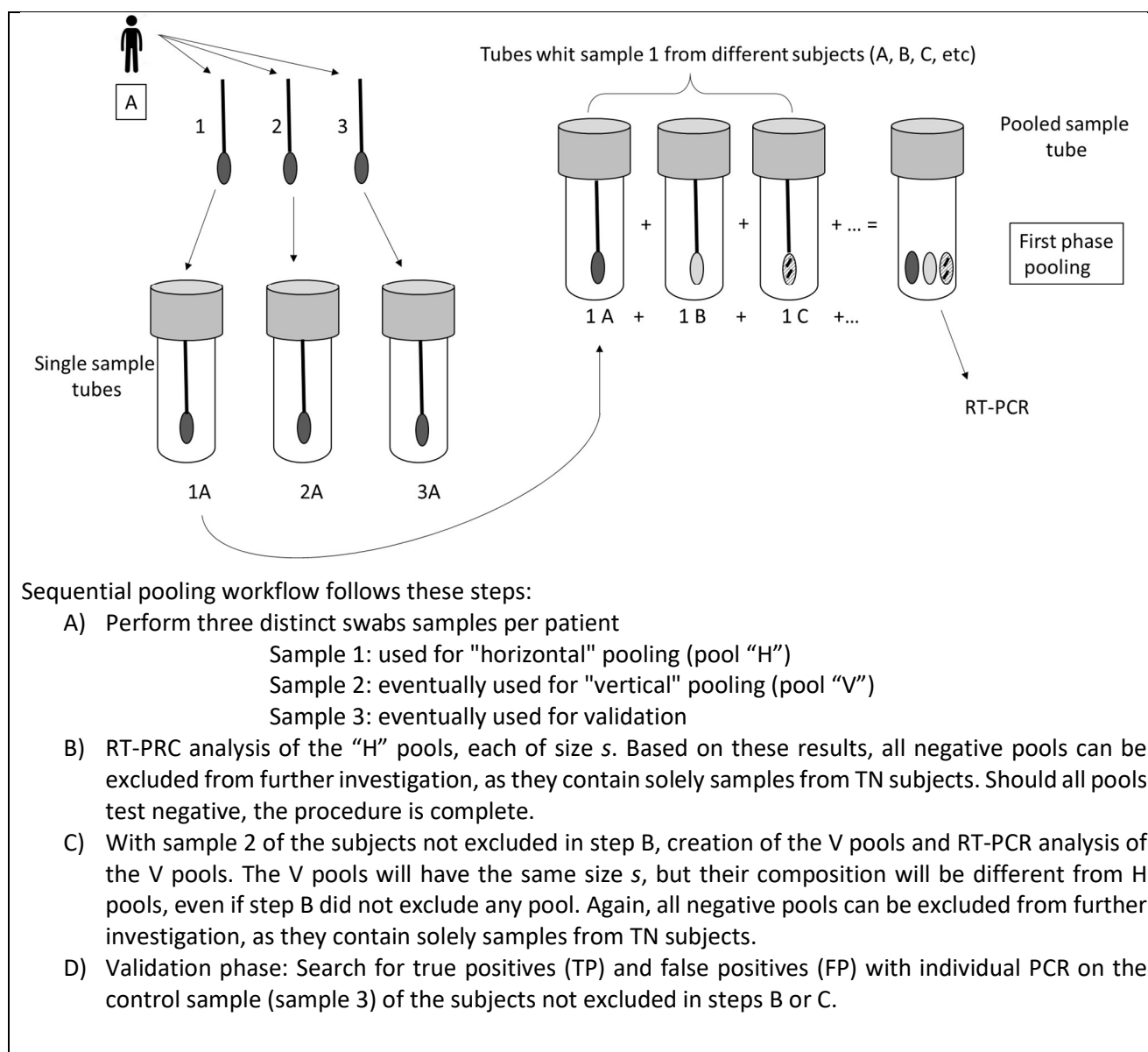


Figure 1. Sequential pooling workflow.

The informed sequential pooling follows exactly the same procedure with the only difference that a score about the probability to be infected will be associated to each subject, in order to tag the subject as "suspected positive" or "suspected negative". The aim is to include in the same pool subjects with the higher scores, avoiding their random spreading in the matrix. The correct assignment of this score should be accomplished by compiling a dedicated questionnaire. The score is calculated on the basis of clinical and epidemiological criteria that have already been associated to a higher risk of acquiring COVID-19 [8]. For instance, susceptibility seems to be strongly associated with age and biological sex [9, 10, 11] suggesting that these simple criteria may play a pivotal role for the pool assignment.

In Figures 2 and 3, we show a simple graphic representation of the sequential pooling and of the informed sequential pooling approach respectively. For these explicative images, we have chosen a cohort to test of dimension N equal to 30, just to facilitate the visual representation.

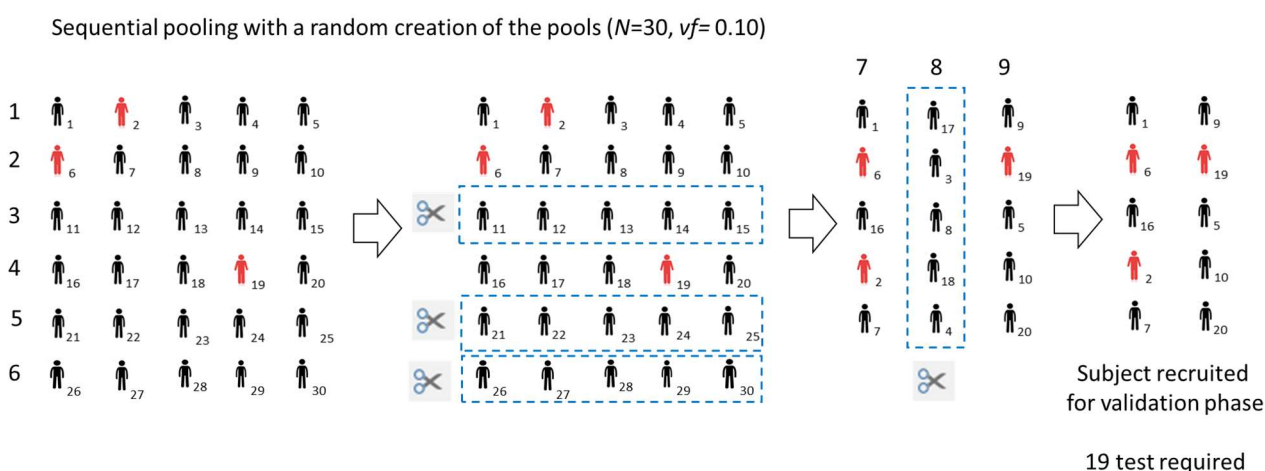


Figure 2. Graphic representation of the sequential pooling approach. The cohort dimension is $N=30$, the pool size is 5, the virus frequency, vf , is 0.10.

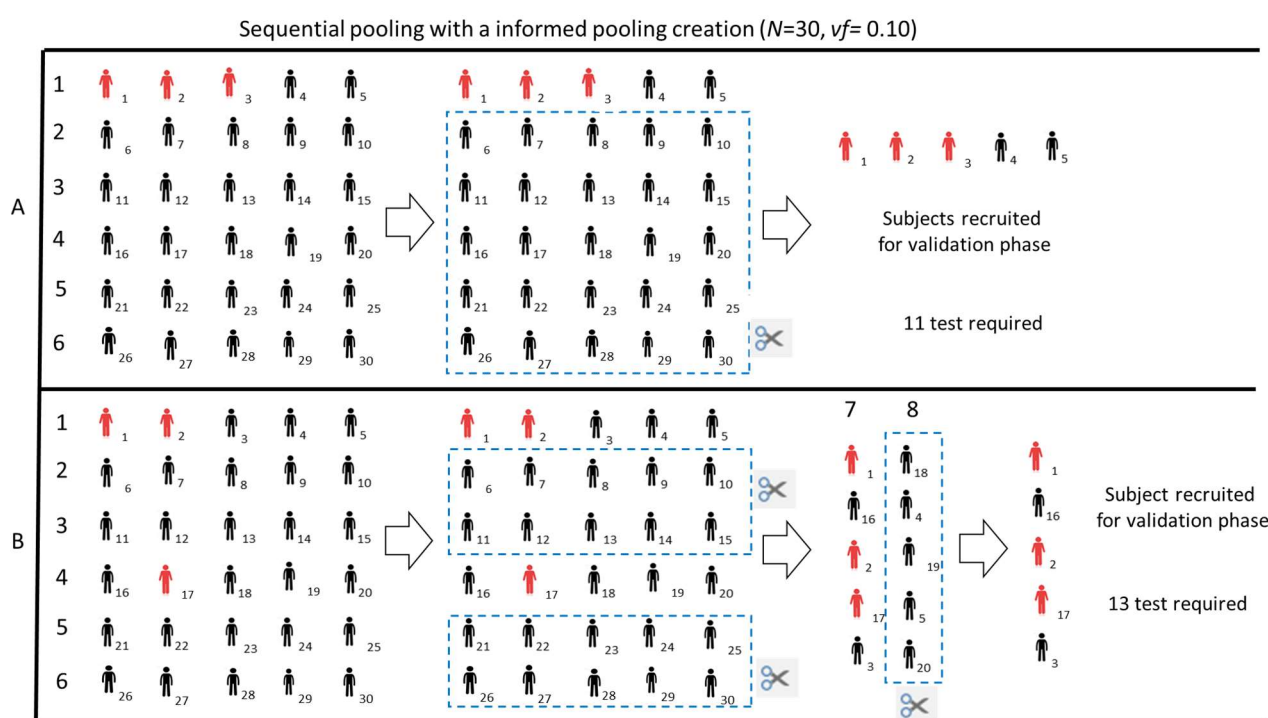


Figure 3. Graphic representation of the sequential informed pooling approach. The cohort dimension is $N=30$, the pool size is 5, the vf is 0.10. Panel A and B show two possible scenarios according to a different reliability of the information available to classify subjects as “suspect positive” or “suspect negative”.

In Figure 3, the upper panel shows a hypothetical scenario for which all positive subjects are grouped in the first pool. This result can be obtained if the information available to classify subjects as “suspect positive” or “suspect negative” are optimal. In the lower panel, we show another scenario for which clinical and epidemiological information allowed a grouping of the positives subjects which is only partially correct.

However, also in this case, the informed approach is still useful to improve the efficiency of the method as compared to a random creation of pools.

Results

In order to assess the advantage of this two-step sequential pooling strategy in comparison with a standard approach, where each subject's swab is separately tested, we performed simulations¹ under different conditions. We assumed to analyse a group of $N=600$ subjects. The size s of each pool (both H and V) was allowed to vary from 2 to 300.² We examined a virus frequency, vf , ranging from 0.01 to 0.30 (the latter situation thus corresponding to 30% of the subjects TP to the virus).

Preliminary, we examined the performance of this strategy without using prior information about the subjects, that is creating pools completely at random. To do this, after setting s and vf , we performed 5,000 simulations and we recorded the ratio between the total number of swab tests required, T , and N . For two-step sequential pooling, T includes both H and V pools required in steps B and C, but also validation tests in step D on all the swabs from subjects not previously excluded. Since, without a pooling strategy, N tests should be performed, the ratio T/N measures the efficacy of the proposed procedure. The smaller is its value, the larger is the reduction of required tests. Conversely, values close to 1 (or even above 1) would represent a useless (or a counter-productive) strategy. Table 1 and Figure 4 show the results for s equal to 5, 12 and 24 (the entire set of plots is available in the Supplementary Material section, Figure S.1). The curves plotted represent the 1st, the 25th, the 50th (median), the 75th, and the 99th percentiles of T/N obtained in the set of 5,000 simulations, for different vf values. In particular, the 1st and the 99th percentiles give an idea of the range of T/N between "lucky" or "unlucky" assignments to the pools. The spread between the 25th and the 75th, which is always very small in Figure 4, represents the central half of the simulations (after excluding the 25% more "lucky" and the 25% more "unlucky" ones). As the pool size increases, we notice that the curves are less linear and the spread between the 1st and the 99th percentile increases. For very small pools ($s=3$), with a low virus frequency, the number of tests required is about 40% with our procedure as compared to separately testing each subject. As vf moves to larger values, the number of tests grows slowly and the pooling is still efficient ($T/N < 1$) even if 25% of the subjects are positive in the group. Conversely, if we use larger pools ($s=24$), the number of tests could drop to 20% for low virus frequency. However, the number of tests would increase faster as vf grows, and the procedure would be efficient only up to about 10% of positive

¹ Results were obtained with Wolfram Mathematica 12.1.

² The list of possible sizes s to split $N=600$ subjects is equal to $\{2,3,4,5,6,8,10,12,15,20,24,25,30,40,50,60,75,100,120,150,200,300\}$.

subjects in the analysed cohort. In summary, the linear path of small pools ensures efficiency even for larger vf , but the nonlinear path of larger pools allows great efficiency for populations with a low virus presence.

T/N	vf	1st	25th	50th	75th	99th
$s=3$	0.01	0.353	0.363	0.363	0.368	0.373
	0.05	0.463	0.478	0.485	0.490	0.505
	0.10	0.597	0.618	0.627	0.635	0.657
	0.15	0.723	0.750	0.760	0.770	0.797
	0.20	0.842	0.872	0.883	0.897	0.927
	0.25	0.952	0.985	1.000	1.013	1.048
	0.30	1.057	1.090	1.105	1.118	1.157
$s=12$	0.01	0.133	0.153	0.173	0.173	0.193
	0.05	0.393	0.442	0.463	0.485	0.547
	0.10	0.640	0.718	0.743	0.772	0.848
	0.15	0.828	0.897	0.933	0.958	1.038
	0.20	0.955	1.018	1.042	1.065	1.127
	0.25	1.038	1.083	1.105	1.125	1.167
	0.30	1.082	1.125	1.145	1.167	1.167
$s=24$	0.01	0.130	0.172	0.210	0.212	0.252
	0.05	0.510	0.630	0.670	0.712	0.797
	0.10	0.800	0.918	0.960	1.000	1.083
	0.15	0.960	1.042	1.043	1.083	1.083
	0.20	1.002	1.083	1.083	1.083	1.083
	0.25	1.042	1.083	1.083	1.083	1.083
	0.30	1.083	1.083	1.083	1.083	1.083

Table 1. Random sequential pooling. Values of the 1st, the 25th, the 50th (median), the 75th, and the 99th percentiles of T/N obtained in the set of 5,000 simulations, for three values of the pool size ($s=3, 12, 24$) and some values of the virus frequency ($vf=0.01, 0.10, 0.15, 0.20, 0.25, 0.30$).

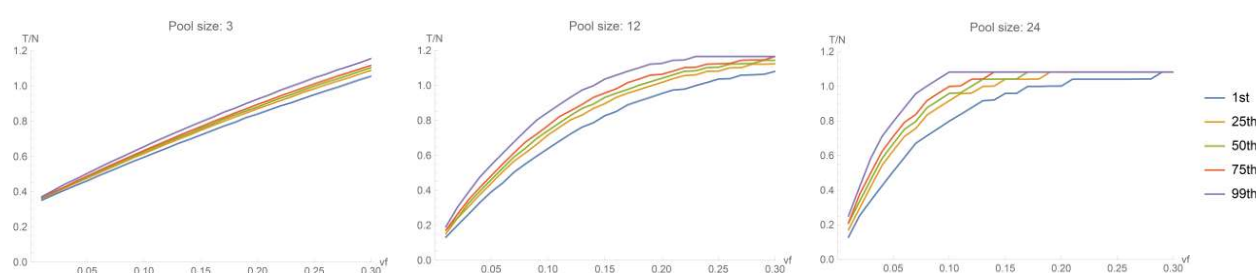


Figure 4. Random sequential pooling. The curves plotted represent the 1st, the 25th, the 50th (median), the 75th, and the 99th percentiles of T/N obtained in the set of 5,000 simulations, for three values of the pool size ($s=3, 12, 24$).

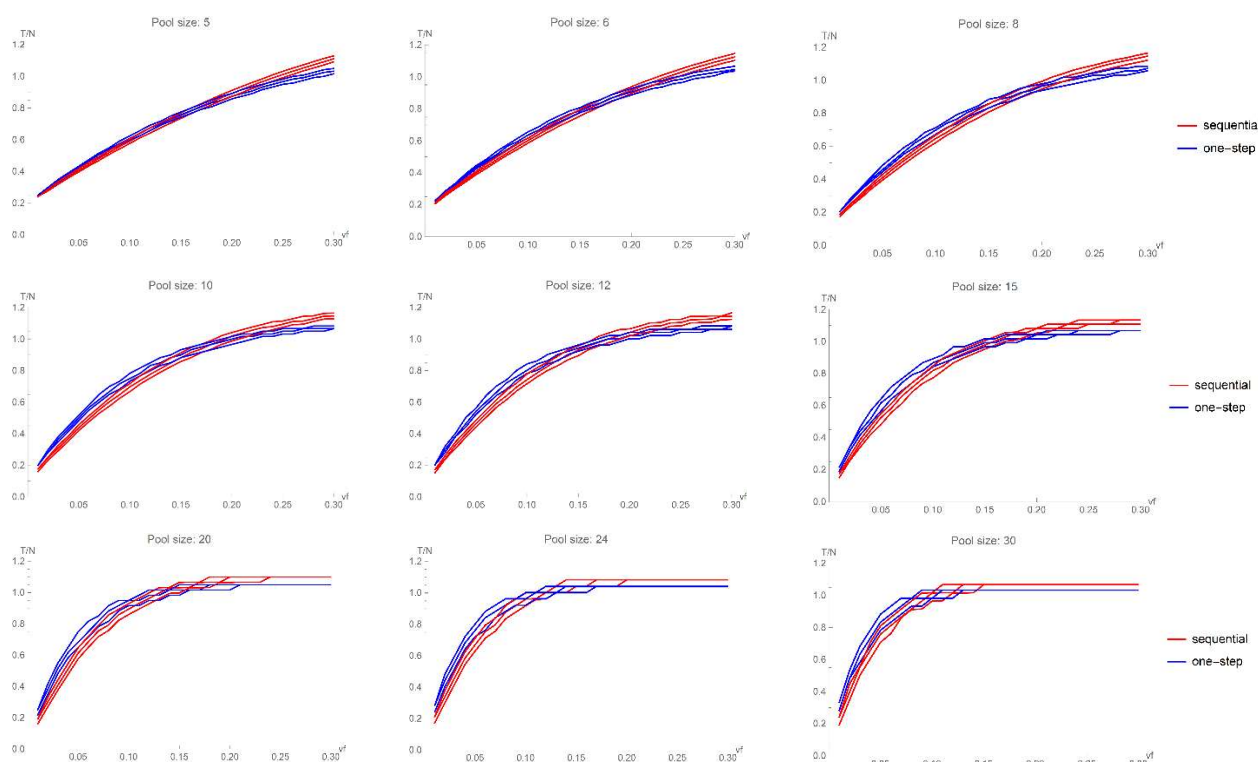


Figure 5. Random sequential pooling vs one-step pooling. The curves plotted represent the 25th, the 50th (median), and the 75th percentiles of T/N obtained in the set of 5,000 simulations, for 9 values of the pool size ($s=5, 6, 8, 10, 12, 15, 20, 24, 30$).

As mentioned in the Introduction, simple pooling has been recently proposed in Hogan et al [7]. We notice that their study does not provide efficiency results apart from their specific application, where pools of size 9 and 10 have been used and a very small v_f has been reported (their value is even smaller than the smallest virus frequency assessed in our simulations). Figure 5 shows a comparison of a simple one-step pooling strategy with our two-step sequential procedure for different v_f and s values. In this picture, the 25th, the 50th (median), and the 75th percentiles of T/N are shown. For very small pools ($s=5$), they are almost equivalent. But, as soon as s is slightly increased to sensible values (ranging from 8 to 20), the sequential two-step pooling shows a better performance up to $v_f=0.15$. For bigger pools ($s=24, 30$), we observe the same result up to v_f around 0.10. For higher frequency virus, both pooling strategies are counter-productive, as highlighted before for sequential pooling.

As explained, all the previous results have been obtained assuming a completely random assignment of subjects to the pools. Often, however, clinic and epidemiologic data about the subjects are available. If we could use these data to concentrate a portion of the positive subjects in the same horizontal pools, we would increase efficiency due to a higher number of negative pools at step B. In order to assess the savings of such an "informed pooling creation", we extended our simulations to different settings. Imagine that, prior to the test we detect a certain number of subjects, say x , that we expect to be positive (according to epidemiological criteria). We create thus x/s horizontal pools, each of size s , with those subjects. The remaining $(N-x)$ subjects

are assigned to the remaining $(N-x)/s$ horizontal pools. Should epidemiological criteria be perfect, all the x subjects turn out to be true positive and thus the first x/s pools are positive. At the same time, all the $(N-x)$ subjects without prior indication of an infection, with perfect epidemiological criteria, would be true negative and thus their $(N-x)/s$ horizontal pools would give a negative result. Of course, such an assumption is unrealistic and we expect that some of the x subjects suspected to be positive are true negative and also some of the $(N-x)$ subjects suspected to be negative are true positive. Let us denote by α the fraction of the $vf \cdot N$ true positive subjects in the population that are correctly assigned to the initial pools. The remaining $(1-\alpha)$ fraction is undetected and it is wrongly assigned to the second part of the pools. Criteria with perfect performance in prior detection of positive subjects would result in $\alpha=1$. In addition, let us denote by β the fraction of the $(1-vf) \cdot N$ true negative subjects in the population that are correctly assigned to the final pools. The remaining $(1-\beta)$ fraction is wrongly assigned to the first part of the pools. Criteria with perfect performance in prior detection of negative subjects would result in $\beta=1$. For the same settings analysed in the random creation of the pools ($N=600$, vf from 0.01 to 0.30, and s from 2 to 300), we explored the performance of the sequential procedure for different values of α and β . In particular, we allowed α and β to vary in the set $\{0.5, 0.6, 0.7, 0.8\}$. When both α and β are equal to 0.5, criteria are essentially unreliable and our situation is equivalent to the random assignment setting above discussed.

Figure 6 shows the results of the simulations obtained for three values of the pool size ($s=3,12,24$) for different combinations of α and β (the plots for the remaining s values are shown in the Supplementary Material section, Figures S.2, S.3). Since there are many situations, to improve clarity, we plotted only the 50th percentile (the values for all the 5 percentiles are displayed in Tables 2a, 2b, 2c). Our aim is to compare the results of the number of tests required when swabs are randomly assigned to the pools with the number of tests required for different α and β values. As above mentioned, we started with α and β equal to 0.5, because this is substantially equivalent to uninformative prior criteria. As α and/or β increase, we observe that the number of required tests decreases, and this decrease is larger when the virus frequency is larger. When vf is below 5%, random pooling and informed pooling are almost equivalent. With a low vf , sequential random pooling was however already very performant, substantially decreasing the number of test with respect to separate individual tests. For larger vf , the curves corresponding to random assignment and informed pooling separate more and more. This entails that reliable informed pooling increases the performance of the pooling exactly when the situation is less favourable. For example, with a pool size equal to 12, with a random assignment, the median of T/N is equal to 1 when $vf \approx 0.18$ (making random pooling application questionable). Conversely, if informed pooling is performed with $\alpha=\beta=0.8$, at the same vf , the median of T/N is equal approximately to 0.73. With $\alpha=\beta=0.8$, pooling is still efficient ($T/N < 1$) even if the virus frequency approaches 30%. In summary, reliable informed pooling makes the performance path much more linear than we observed for random pooling, even if we use larger pools. That is, larger pools, besides

providing substantial savings for low vf , ensure efficiency even for larger vf if epidemiological criteria provide reliable information.

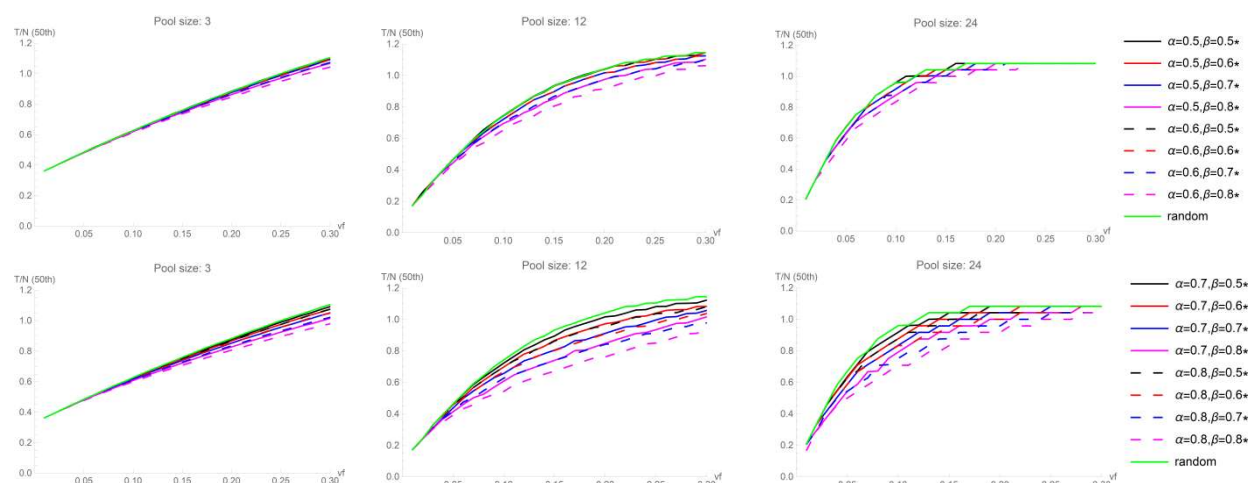


Figure 6. Informed sequential pooling. The curves plotted represent the 50th percentile (median) of T/N obtained in the set of 5,000 simulations, for three values of the pool size ($s=3,12,24$). The upper plots were obtained with $\alpha=0.5$ and $\alpha=0.6$, combined with $\beta=0.5, 0.6, 0.7, 0.8$. The lower plots were obtained with $\alpha=0.7$ and $\alpha=0.8$, combined with $\beta=0.5, 0.6, 0.7, 0.8$.

Discussion

Every successful strategy in controlling SARS-CoV2 infection depends on the viral timely diagnosis. Hence, there is an urgent need for a systematic population screening at a scale mass. Currently, around the world, there is a plethora of different scenarios, which mainly depends on the spread of infection. Even in the same country, we can find very different situations: for example, it is likely that the asymptomatic population has a lower viral frequency than the symptomatic one or than who belongs from a category at risk, such as the health personnel. Moreover, in this variegated context, there are completely different economic situations, and the pooling strategy can become truly attractive for countries with fewer resources. The study published on JAMA [7] is certainly an excellent starting point to evaluate an alternative approach to individual analysis of swab sample for the RT-PCR based diagnosis of SARS-CoV2, but some appropriate considerations are needed. Firstly, it must be highlighted that in the JAMA study, 292 pools of 9 or 10 samples were created and two positive cases in a collection of 2888 samples were founded. The one-step pooling method gave excellent results because the frequency of the virus in the analyzed samples was extremely low (0.07%). Secondly, if it was possible to roughly estimate the frequency of the virus in the collection as lower than 5%, our data suggest to increase the pool size. Using a pool size of 24, for example, the screening of the 2888 samples would needed about 120 tests instead of 292.

Our most straightforward result is that the sequential pooling approach is more efficient than one-step pooling method. In addition, the informed version of the sequential pooling can further improve its performance, in particular for larger size pools and moderate to large virus frequency. Table 3 broadly describes practical suggestions to decide the pool size, s , according to rough assumptions about the virus frequency, both for random and informed sequential pooling. Larger pools ensure very large reduction in the number of tests when vf is small. Smaller pools may be a conservative approach when dealing with cohorts with heavier exposure. Finally, indications are also given to avoid the use of pooling when virus frequency is higher and random pooling would result in a waste of resources, since too many pools are expected to give a positive result.

Random sequential pooling	Informed sequential pooling ($\alpha, \beta \geq 0.7$)
If we can guess a vf below 10%, pools with size from 10 to 15 can provide relevant savings in the number of tests. With vf below 5%, even stronger savings can be obtained with pool size increased to 20 or 25.	If we can guess a vf below 10%, very large pools, with size from 20 to 25 could substantially reduce the number of tests. Pools with size equal to 30 or 40 are a good strategy with vf below 5%.
For situations when vf may reach 10%-20% of the cohort, we can still have a moderate reduction in the number of tests, with pools of size between 5 and 8.	For situations when vf may reach 10%-20% of the cohort, we can still have a relevant reduction in the number of tests, with pools of size between 12 and 20.
For situations when there is the risk of a vf value above 20% of the cohort, pooling strategies should be avoided.	For situations when there is the risk of a vf value above 20% of the cohort, a moderate reduction can be attained with pools of size about 12, to be further reduced to 8, if the vf may exceed 25%.

Table 3. Summary of practical indications to pooling creation.

References

- 1] Al-Tawfiq JA. Asymptomatic coronavirus infection: MERS-CoV and SARS-CoV-2 (COVID-19). Travel Med Infect Dis. 2020; 27:101608. doi: 10.1016/j.tmaid.2020.101608.
- [2] Wang D, Hu B, Hu C, Zhu F, Liu X, Zhang J, et al. Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus–infected pneumonia in Wuhan, China. JAMA. 2020. doi:10.1001/jama.2020.1585
- [3] Lan L, Xu D, Ye G, Xia C, Wang S, Li Y, et al. Positive RT-PCR Test Results in Patients Recovered From COVID-19. JAMA. 2020 Feb 27. doi: 10.1001/jama.2020.2783. [Epub ahead of print]
- [4] Charlton CL, Babady E, Ginocchio CC, Hatchette TF, Jerris RC, Li Y, et al. Practical guidance for clinical microbiology laboratories: viruses causing acute respiratory tract infections. Clin Microbiol Rev. 2019;32(1). doi:00010.01128/CMR. 00042-00018.

- [5] Azzi L, Carcano G, Gianfagna F, Grossi P, Gasperina DD, Genoni A, et al. University of Insubria COVID-19 Task Force, Baj A5. SALIVA IS A RELIABLE TOOL TO DETECT SARS-CoV-2. *J Infect*. 2020: S0163-4453(20)30213-9. doi: 10.1016/j.jinf.2020.04.005. [Epub ahead of print]
- [6] Falsey AR, Formica MA, Walsh EE. Simple method for combining sputum and nasal samples for virus detection by reverse transcriptase PCR. *J Clin Microbiol*. 2012; 50(8):2835. doi:2810.1128/JCM. 01473-01412.
- [7] Hogan CA, Sahoo MK, Pinsky BA. Sample Pooling as a Strategy to Detect Community Transmission of SARS-CoV-2. *JAMA*. 2020 doi: 10.1001/jama.2020.5445
- [8] Sasmita PA, Sha M, Yu-Ju W, Yu-Ping M, Rui-Xue Y, Qing-Zhi W, et al. Epidemiology, causes, clinical manifestation and diagnosis, prevention and control of coronavirus disease (COVID-19) during the early outbreak period: a scoping review. *Infect Dis Poverty*. 2020; 9: 29. doi: 10.1186/s40249-020-00646-x)
- [9] Fehr AR, Channappanavar R, Perlman S. Middle East respiratory syndrome: emergence of a pathogenic human coronavirus. *Annu Rev Med*. 2017;68:387–399. doi: 10.1146/annurev-med-051215-031152.
- [10] Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *Lancet*. 2020;395:507–513. doi: 10.1016/S0140-6736(20)30211-7.
- [11] Backer JA, Klinkenberg D, Wallinga J. The incubation period of 2019-nCoV infections among travellers from Wuhan. *China Euro Surveill*. 2020; 10.2807/1560-7917.ES.2020.25.5.2000062.

		$\beta=0.5$					$\beta=0.6$					$\beta=0.7$					$\beta=0.8$				
		1st	25th	50th	75th	99th	1st	25th	50th	75th	99th	1st	25th	50th	75th	99th	1st	25th	50th	75th	99th
$\alpha=0.5$	$\nu f=0.01$	0.353	0.363	0.363	0.368	0.373	0.353	0.363	0.363	0.368	0.373	0.357	0.363	0.363	0.368	0.373	0.353	0.363	0.363	0.368	0.373
	$\nu f=0.05$	0.463	0.478	0.485	0.490	0.505	0.465	0.478	0.485	0.490	0.503	0.463	0.478	0.485	0.490	0.507	0.463	0.477	0.482	0.488	0.503
	$\nu f=0.10$	0.597	0.618	0.627	0.635	0.657	0.597	0.618	0.627	0.635	0.657	0.595	0.615	0.625	0.633	0.653	0.590	0.612	0.620	0.628	0.650
	$\nu f=0.15$	0.725	0.75	0.760	0.772	0.797	0.723	0.748	0.758	0.768	0.797	0.718	0.745	0.755	0.767	0.793	0.712	0.737	0.745	0.757	0.782
	$\nu f=0.20$	0.842	0.872	0.885	0.897	0.927	0.840	0.870	0.883	0.895	0.925	0.833	0.865	0.877	0.888	0.920	0.822	0.850	0.863	0.875	0.903
	$\nu f=0.25$	0.953	0.985	1.000	1.013	1.047	0.950	0.983	0.997	1.010	1.043	0.942	0.975	0.988	1.002	1.033	0.928	0.957	0.970	0.982	1.015
$\alpha=0.6$	$\nu f=0.30$	1.055	1.090	1.105	1.120	1.155	1.053	1.087	1.102	1.117	1.152	1.045	1.077	1.092	1.105	1.138	1.025	1.055	1.068	1.082	1.115
	$\nu f=0.01$	0.353	0.363	0.363	0.368	0.373	0.357	0.363	0.363	0.368	0.373	0.353	0.363	0.363	0.368	0.373	0.353	0.363	0.363	0.368	0.373
	$\nu f=0.05$	0.463	0.478	0.485	0.490	0.507	0.463	0.478	0.485	0.490	0.503	0.463	0.477	0.483	0.488	0.503	0.462	0.475	0.482	0.487	0.502
	$\nu f=0.10$	0.597	0.618	0.627	0.635	0.657	0.595	0.617	0.625	0.633	0.653	0.592	0.613	0.622	0.630	0.650	0.587	0.607	0.615	0.623	0.643
	$\nu f=0.15$	0.723	0.748	0.760	0.770	0.797	0.718	0.745	0.755	0.767	0.792	0.712	0.738	0.748	0.758	0.785	0.702	0.727	0.737	0.747	0.772
	$\nu f=0.20$	0.842	0.872	0.883	0.895	0.923	0.837	0.865	0.877	0.888	0.920	0.827	0.855	0.867	0.878	0.908	0.810	0.837	0.847	0.858	0.887
$\alpha=0.7$	$\nu f=0.25$	0.952	0.983	0.997	1.010	1.047	0.943	0.975	0.988	1.002	1.035	0.932	0.962	0.975	0.987	1.018	0.908	0.938	0.950	0.962	0.992
	$\nu f=0.30$	1.053	1.088	1.103	1.117	1.152	1.045	1.077	1.092	1.107	1.143	1.028	1.060	1.073	1.087	1.122	1.002	1.032	1.043	1.057	1.088
	$\nu f=0.01$	0.353	0.363	0.363	0.368	0.373	0.353	0.363	0.363	0.368	0.373	0.353	0.363	0.363	0.368	0.373	0.353	0.363	0.363	0.368	0.373
	$\nu f=0.05$	0.463	0.478	0.483	0.490	0.505	0.463	0.477	0.483	0.488	0.503	0.462	0.475	0.482	0.487	0.502	0.458	0.473	0.480	0.485	0.500
	$\nu f=0.10$	0.595	0.617	0.625	0.633	0.653	0.593	0.613	0.622	0.630	0.652	0.587	0.608	0.617	0.625	0.647	0.578	0.600	0.608	0.617	0.637
	$\nu f=0.15$	0.718	0.745	0.755	0.767	0.792	0.713	0.740	0.750	0.760	0.787	0.705	0.730	0.740	0.750	0.775	0.690	0.713	0.723	0.733	0.757
$\alpha=0.8$	$\nu f=0.20$	0.835	0.865	0.877	0.890	0.918	0.827	0.855	0.868	0.880	0.910	0.813	0.842	0.853	0.865	0.893	0.790	0.818	0.828	0.838	0.865
	$\nu f=0.25$	0.943	0.975	0.990	1.002	1.035	0.932	0.963	0.975	0.988	1.022	0.915	0.943	0.957	0.968	0.998	0.887	0.913	0.925	0.935	0.965
	$\nu f=0.30$	1.042	1.077	1.092	1.105	1.142	1.030	1.062	1.075	1.088	1.122	1.007	1.038	1.050	1.063	1.093	0.975	1.003	1.015	1.027	1.053
	$\nu f=0.01$	0.353	0.363	0.363	0.368	0.373	0.353	0.363	0.363	0.368	0.373	0.353	0.363	0.363	0.368	0.373	0.353	0.362	0.363	0.368	0.373
	$\nu f=0.05$	0.463	0.477	0.483	0.490	0.503	0.462	0.477	0.482	0.488	0.503	0.458	0.475	0.480	0.487	0.502	0.457	0.470	0.477	0.483	0.498
	$\nu f=0.10$	0.592	0.613	0.622	0.630	0.650	0.587	0.608	0.618	0.627	0.647	0.583	0.603	0.612	0.620	0.642	0.572	0.592	0.600	0.608	0.628
	$\nu f=0.15$	0.713	0.738	0.750	0.760	0.787	0.707	0.732	0.742	0.752	0.778	0.693	0.718	0.728	0.738	0.765	0.677	0.698	0.708	0.718	0.742
	$\nu f=0.20$	0.825	0.855	0.867	0.880	0.910	0.813	0.843	0.855	0.867	0.895	0.797	0.825	0.835	0.847	0.875	0.772	0.797	0.807	0.817	0.842
	$\nu f=0.25$	0.930	0.962	0.975	0.988	1.023	0.917	0.945	0.957	0.970	1.002	0.892	0.922	0.933	0.945	0.972	0.862	0.885	0.897	0.905	0.930
	$\nu f=0.30$	1.028	1.062	1.075	1.088	1.123	1.008	1.040	1.052	1.065	1.098	0.982	1.010	1.022	1.033	1.063	0.943	0.970	0.980	0.990	1.013

Table 2a. Informed sequential pooling, pool size $s=3$. Values of the 1st, the 25th, the 50th (median), the 75th, and the 99th percentiles of T/N obtained in the set of 5,000 simulations, for all combinations of α and β in $\{0.5, 0.6, 0.7, 0.8\}$, for some values of the virus frequency ($\nu f=0.01, 0.10, 0.15, 0.20, 0.25, 0.30$).

		β=0.5					β=0.6					β=0.7					β=0.8				
		1st	25th	50th	75th	99th	1st	25th	50th	75th	99th	1st	25th	50th	75th	99th	1st	25th	50th	75th	99th
α=0.5	νf=0.01	0.133	0.153	0.173	0.173	0.213	0.133	0.153	0.173	0.173	0.213	0.133	0.153	0.173	0.173	0.193	0.132	0.153	0.173	0.173	0.213
	νf=0.05	0.395	0.442	0.465	0.487	0.547	0.382	0.442	0.463	0.485	0.547	0.382	0.440	0.462	0.483	0.543	0.375	0.422	0.443	0.477	0.523
	νf=0.10	0.642	0.720	0.745	0.782	0.850	0.640	0.715	0.742	0.767	0.848	0.632	0.698	0.722	0.748	0.823	0.598	0.672	0.695	0.718	0.782
	νf=0.15	0.828	0.898	0.935	0.958	1.038	0.828	0.893	0.932	0.957	1.022	0.805	0.872	0.897	0.933	0.997	0.763	0.828	0.852	0.888	0.953
	νf=0.20	0.955	1.020	1.042	1.080	1.127	0.940	1.017	1.040	1.062	1.125	0.917	0.980	1.017	1.040	1.102	0.873	0.938	0.975	0.997	1.062
α=0.6	νf=0.25	1.038	1.085	1.105	1.125	1.167	1.022	1.083	1.103	1.125	1.167	0.998	1.060	1.082	1.103	1.147	0.960	1.020	1.043	1.080	1.125
	νf=0.30	1.082	1.125	1.145	1.167	1.167	1.082	1.123	1.145	1.147	1.167	1.060	1.103	1.125	1.145	1.167	1.020	1.082	1.103	1.123	1.167
	νf=0.01	0.133	0.153	0.173	0.173	0.193	0.132	0.153	0.173	0.173	0.213	0.132	0.153	0.173	0.173	0.193	0.132	0.153	0.172	0.173	0.193
	νf=0.05	0.385	0.442	0.463	0.487	0.547	0.378	0.440	0.462	0.483	0.545	0.375	0.435	0.457	0.478	0.523	0.368	0.415	0.437	0.457	0.500
	νf=0.10	0.640	0.715	0.742	0.767	0.847	0.635	0.698	0.723	0.760	0.825	0.612	0.675	0.698	0.722	0.785	0.568	0.632	0.653	0.675	0.737
α=0.7	νf=0.15	0.815	0.893	0.932	0.957	1.022	0.805	0.872	0.897	0.933	0.998	0.763	0.830	0.867	0.890	0.953	0.718	0.782	0.805	0.827	0.890
	νf=0.20	0.940	1.017	1.040	1.062	1.123	0.917	0.980	1.017	1.040	1.102	0.873	0.938	0.975	0.997	1.060	0.828	0.892	0.915	0.952	0.998
	νf=0.25	1.020	1.083	1.103	1.125	1.167	0.998	1.060	1.082	1.103	1.147	0.957	1.020	1.042	1.063	1.125	0.913	0.977	0.998	1.022	1.082
	νf=0.30	1.080	1.123	1.145	1.147	1.167	1.058	1.103	1.123	1.145	1.167	1.020	1.082	1.102	1.123	1.167	0.980	1.040	1.062	1.083	1.145
	νf=0.01	0.133	0.153	0.173	0.173	0.213	0.132	0.153	0.173	0.173	0.213	0.132	0.153	0.173	0.173	0.193	0.132	0.153	0.172	0.173	0.193
α=0.8	νf=0.05	0.380	0.440	0.462	0.483	0.543	0.377	0.437	0.458	0.478	0.525	0.372	0.417	0.437	0.458	0.518	0.350	0.393	0.413	0.435	0.477
	νf=0.10	0.633	0.698	0.723	0.760	0.823	0.612	0.677	0.698	0.722	0.785	0.572	0.633	0.657	0.678	0.738	0.527	0.588	0.608	0.630	0.690
	νf=0.15	0.805	0.872	0.895	0.932	0.997	0.763	0.830	0.867	0.890	0.953	0.720	0.783	0.805	0.827	0.890	0.658	0.718	0.740	0.762	0.822
	νf=0.20	0.915	0.980	1.017	1.038	1.085	0.887	0.938	0.973	0.995	1.043	0.827	0.890	0.912	0.933	0.997	0.763	0.825	0.848	0.870	0.933
	νf=0.25	0.998	1.060	1.082	1.103	1.147	0.957	1.018	1.040	1.062	1.107	0.912	0.975	0.997	1.018	1.080	0.865	0.913	0.935	0.973	1.037
α=0.8	νf=0.30	1.058	1.103	1.123	1.145	1.167	1.018	1.063	1.083	1.105	1.147	0.975	1.037	1.058	1.080	1.125	0.932	0.995	1.017	1.038	1.083
	νf=0.01	0.132	0.153	0.173	0.173	0.213	0.132	0.153	0.173	0.173	0.213	0.132	0.153	0.172	0.173	0.193	0.130	0.152	0.172	0.173	0.193
	νf=0.05	0.377	0.435	0.457	0.478	0.525	0.373	0.417	0.438	0.460	0.520	0.352	0.397	0.417	0.438	0.497	0.328	0.370	0.392	0.412	0.453
	νf=0.10	0.612	0.677	0.698	0.722	0.783	0.575	0.635	0.668	0.693	0.738	0.545	0.605	0.627	0.647	0.692	0.483	0.523	0.543	0.563	0.605
	νf=0.15	0.763	0.828	0.863	0.888	0.952	0.722	0.782	0.803	0.825	0.887	0.673	0.718	0.740	0.760	0.820	0.608	0.652	0.672	0.693	0.737
α=0.8	νf=0.20	0.872	0.933	0.955	0.978	1.040	0.823	0.887	0.908	0.930	0.975	0.762	0.810	0.842	0.865	0.910	0.695	0.740	0.762	0.783	0.842
	νf=0.25	0.953	1.015	1.037	1.058	1.103	0.907	0.953	0.975	0.997	1.058	0.845	0.892	0.913	0.937	0.997	0.780	0.827	0.850	0.885	0.932
	νf=0.30	1.015	1.060	1.082	1.102	1.145	0.953	1.017	1.038	1.058	1.103	0.910	0.957	0.978	1.002	1.060	0.847	0.908	0.930	0.952	1.015

		$\beta=0.5$					$\beta=0.6$					$\beta=0.7$					$\beta=0.8$				
		1st	25th	50th	75th	99th	1st	25th	50th	75th	99th	1st	25th	50th	75th	99th	1st	25th	50th	75th	99th
$\alpha=0.5$	$vf=0.01$	0,130	0,172	0,210	0,212	0,252	0,130	0,172	0,210	0,212	0,292	0,130	0,172	0,210	0,212	0,252	0,130	0,172	0,210	0,212	0,252
	$vf=0.05$	0,545	0,630	0,672	0,712	0,797	0,508	0,630	0,670	0,712	0,797	0,508	0,625	0,633	0,673	0,793	0,503	0,587	0,628	0,668	0,752
	$vf=0.10$	0,835	0,918	0,960	1,000	1,083	0,798	0,918	0,958	1,000	1,043	0,795	0,878	0,918	0,960	1,043	0,753	0,838	0,878	0,918	1,003
	$vf=0.15$	0,960	1,042	1,043	1,083	1,083	0,958	1,042	1,043	1,083	1,083	0,920	1,002	1,042	1,043	1,083	0,917	1,000	1,002	1,042	1,083
	$vf=0.20$	1,003	1,083	1,083	1,083	1,083	1,002	1,083	1,083	1,083	1,083	1,000	1,042	1,083	1,083	1,083	0,960	1,042	1,083	1,083	1,083
	$vf=0.25$	1,042	1,083	1,083	1,083	1,083	1,042	1,083	1,083	1,083	1,083	1,042	1,083	1,083	1,083	1,083	1,042	1,083	1,083	1,083	1,083
$\alpha=0.6$	$vf=0.30$	1,083	1,083	1,083	1,083	1,083	1,043	1,083	1,083	1,083	1,083	1,042	1,083	1,083	1,083	1,083	1,042	1,083	1,083	1,083	1,083
	$vf=0.01$	0,130	0,172	0,210	0,212	0,292	0,130	0,172	0,210	0,212	0,252	0,128	0,170	0,210	0,212	0,252	0,128	0,170	0,208	0,210	0,252
	$vf=0.05$	0,543	0,630	0,670	0,712	0,797	0,507	0,625	0,635	0,673	0,793	0,503	0,588	0,628	0,668	0,752	0,463	0,545	0,587	0,627	0,710
	$vf=0.10$	0,798	0,918	0,958	1,000	1,043	0,795	0,878	0,920	0,960	1,043	0,753	0,838	0,878	0,920	1,002	0,710	0,795	0,835	0,877	0,960
	$vf=0.15$	0,958	1,003	1,043	1,083	1,083	0,918	1,002	1,042	1,043	1,083	0,917	1,000	1,002	1,042	1,083	0,837	0,958	1,000	1,002	1,083
	$vf=0.20$	1,002	1,083	1,083	1,083	1,083	1,000	1,042	1,083	1,083	1,083	0,958	1,042	1,043	1,083	1,083	0,958	1,000	1,042	1,083	1,083
$\alpha=0.7$	$vf=0.25$	1,042	1,083	1,083	1,083	1,083	1,042	1,083	1,083	1,083	1,083	1,000	1,083	1,083	1,083	1,083	1,000	1,042	1,083	1,083	1,083
	$vf=0.30$	1,043	1,083	1,083	1,083	1,083	1,042	1,083	1,083	1,083	1,083	1,042	1,083	1,083	1,083	1,083	1,042	1,083	1,083	1,083	1,083
	$vf=0.01$	0,130	0,172	0,210	0,212	0,292	0,130	0,172	0,210	0,212	0,252	0,130	0,170	0,210	0,212	0,252	0,128	0,170	0,210	0,210	0,252
	$vf=0.05$	0,507	0,593	0,635	0,673	0,793	0,503	0,588	0,628	0,670	0,752	0,465	0,547	0,587	0,627	0,708	0,422	0,503	0,543	0,583	0,627
	$vf=0.10$	0,793	0,878	0,918	0,960	1,042	0,752	0,837	0,877	0,918	1,002	0,710	0,793	0,835	0,877	0,960	0,667	0,750	0,792	0,833	0,917
	$vf=0.15$	0,918	1,000	1,042	1,043	1,083	0,877	0,960	1,000	1,042	1,083	0,835	0,918	0,958	1,000	1,083	0,792	0,875	0,917	0,958	1,042
$\alpha=0.8$	$vf=0.20$	1,000	1,042	1,083	1,083	1,083	0,958	1,042	1,042	1,083	1,083	0,917	1,000	1,042	1,042	1,083	0,875	0,958	1,000	1,042	1,083
	$vf=0.25$	1,042	1,083	1,083	1,083	1,083	1,000	1,042	1,083	1,083	1,083	1,000	1,042	1,083	1,083	1,083	0,958	1,042	1,042	1,083	1,083
	$vf=0.30$	1,042	1,083	1,083	1,083	1,083	1,042	1,083	1,083	1,083	1,083	1,000	1,083	1,083	1,083	1,083	1,000	1,042	1,083	1,083	1,083
	$vf=0.01$	0,128	0,170	0,210	0,212	0,252	0,128	0,170	0,210	0,212	0,252	0,128	0,170	0,208	0,210	0,252	0,127	0,168	0,170	0,210	0,252
	$vf=0.05$	0,503	0,587	0,628	0,668	0,752	0,463	0,547	0,587	0,628	0,710	0,425	0,503	0,543	0,583	0,627	0,380	0,460	0,500	0,502	0,542
	$vf=0.10$	0,753	0,835	0,877	0,917	1,000	0,708	0,792	0,833	0,837	0,918	0,627	0,710	0,752	0,792	0,875	0,583	0,667	0,708	0,710	0,792
	$vf=0.15$	0,877	0,958	1,000	1,002	1,083	0,833	0,917	0,958	0,960	1,042	0,753	0,875	0,877	0,917	1,000	0,708	0,792	0,833	0,875	0,958
	$vf=0.20$	0,958	1,002	1,042	1,083	1,083	0,917	1,000	1,000	1,042	1,083	0,875	0,918	0,958	1,000	1,083	0,792	0,875	0,917	0,958	1,042
	$vf=0.25$	1,000	1,042	1,083	1,083	1,083	0,958	1,042	1,042	1,083	1,083	0,917	1,000	1,042	1,042	1,083	0,875	0,958	1,000	1,042	1,083
	$vf=0.30$	1,042	1,083	1,083	1,083	1,083	1,000	1,042	1,083	1,083	1,083	0,958	1,042	1,042	1,083	1,083	0,917	1,000	1,042	1,042	1,083

Table 2c. Informed sequential pooling, pool size $s=24$. Values of the 1st, the 25th, the 50th (median), the 75th, and the 99th percentiles of T/N obtained in the set of 5,000 simulations, for all combinations of α and β in $\{0.5, 0.6, 0.7, 0.8\}$, for some values of the virus frequency ($vf=0.01, 0.10, 0.15, 0.20, 0.25, 0.30$).

Supplementary Material

In this Section, we list some Figures pertaining to the results of the simulations. In the main paper, only a selection of the plots could be included.

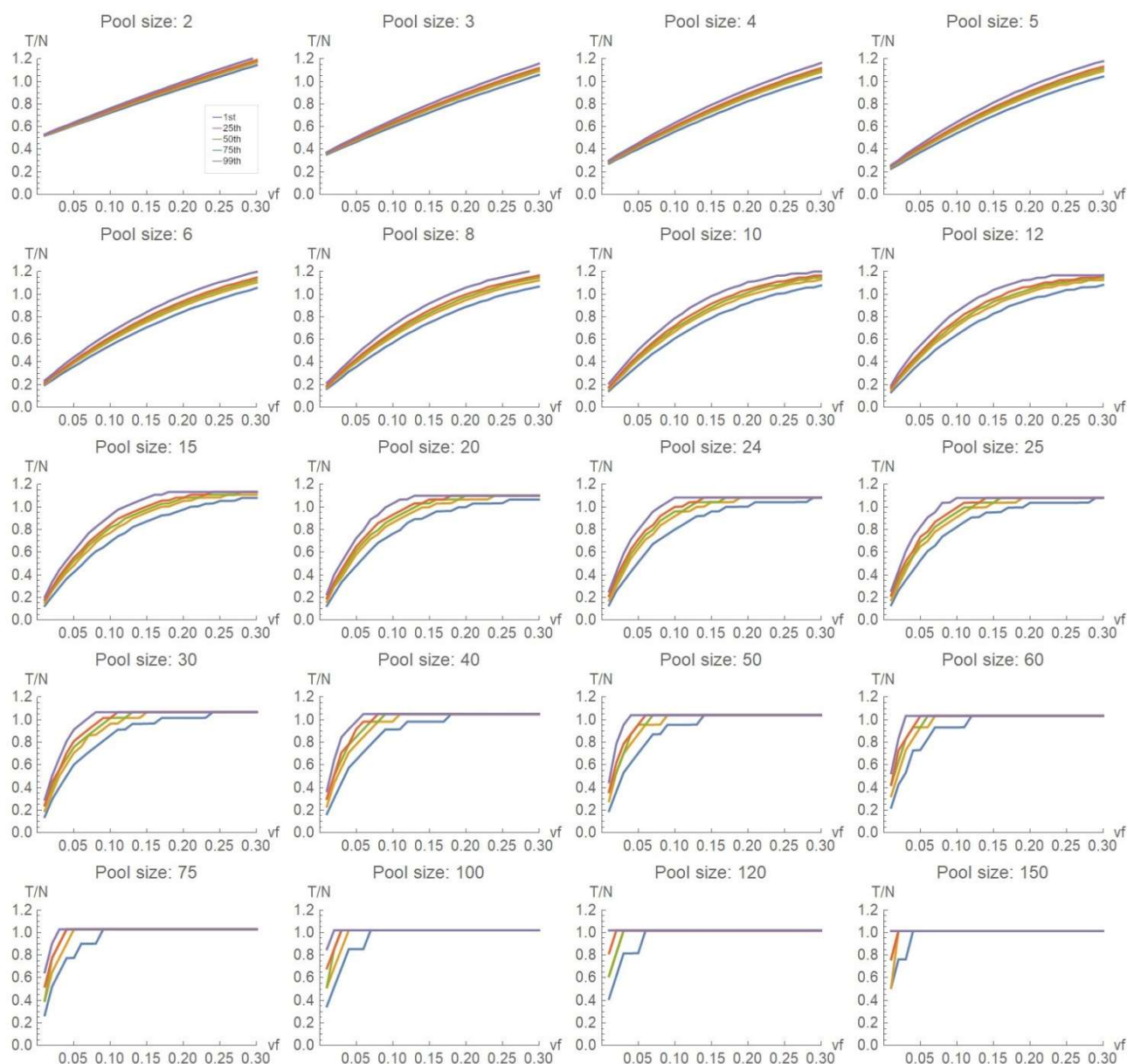


Figure S.1. Random sequential pooling. The curves plotted represent the 1st, the 25th, the 50th (median), the 75th, and the 99th percentiles of T/N obtained in the set of 5,000 simulations, for a three values of the pool size s from 2 to 150.

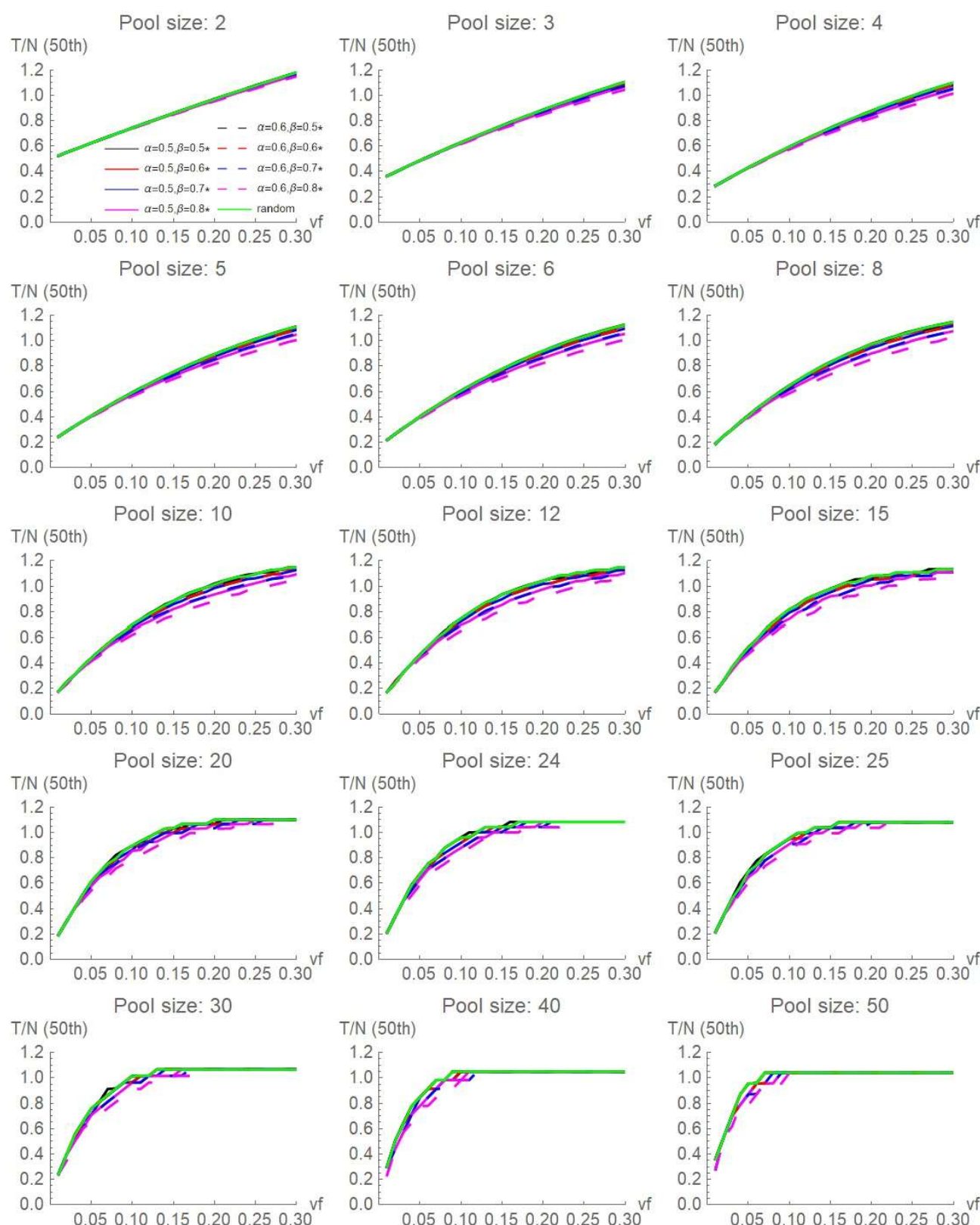


Figure S2. Informed sequential pooling. The curves plotted represent the 50th percentile (median) of T/N obtained in the set of 5,000 simulations, for values of the pool size s ranging from 2 to 50. The plots were obtained with $\alpha=0.5$ and $\alpha=0.6$, combined with $\beta=0.5, 0.6, 0.7, 0.8$.

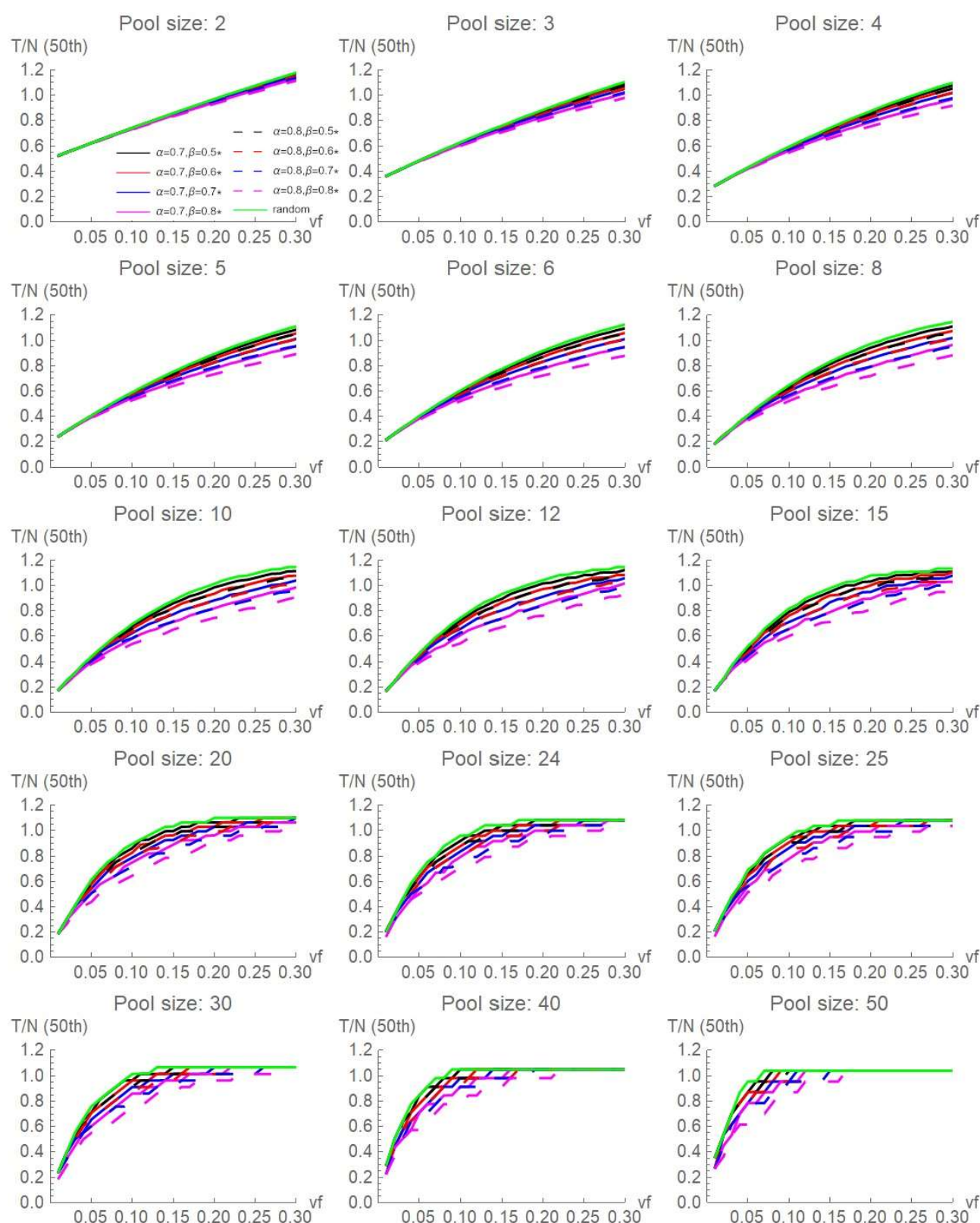


Figure S3. Informed sequential pooling. The curves plotted represent the 50th percentile (median) of T/N obtained in the set of 5,000 simulations, for values of the pool size s ranging from 2 to 50. The plots were obtained with $\alpha=0.7$ and $\alpha=0.8$, combined with $\beta=0.5, 0.6, 0.7, 0.8$.