A unifying model to explain nirmatrelvir / ritonavir's high efficacy during early treatment and low efficacy as post-exposure prophylaxis, and to predict viral rebound

Shadisadat Esmaeili¹, Katherine Owens¹, Jessica Wagoner², Stephen J. Polyak², and Joshua T. Schiffer¹,²

1. Fred Hutchinson Cancer Center, Vaccine and Infectious Disease Division
2. University of Washington, Department of Medicine

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.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

In a pivotal trial, a 5-day course of oral ritonavir-boosted nirmatrelvir decreased hospitalization and death by 89.1% and reduced nasal viral load by 0.87 log relative to placebo when given early during symptomatic infection to high-risk individuals. Yet, more frequent viral and symptomatic rebound has been observed in community cohorts relative to the clinical trial, and ritonavir-boosted nirmatrelvir failed to achieve efficacy in a post-exposure prophylaxis trial. We developed a mathematical model capturing viral-immune dynamics and nirmatrelvir pharmacokinetics that recapitulated viral loads from the clinical trial. Our results demonstrate that nirmatrelvir IC₅₀ (50% inhibitory concentrations) estimates from in vitro assays are approximately 60-fold lower than the plasma concentration required to reduce viral infection by 50% in humans and that a maximally potent agent would reduce the viral load by approximately 2.5 logs relative to placebo at 5 days. The model produces frequent viral rebound trajectories and identifies that earlier treatment initiation and shorter treatment duration are key predictors of rebound. Extension of early symptomatic treatment duration to 10 days and post-exposure prophylaxis to 15 days, rather than increasing dose or dosing frequency, is predicted to significantly lower the incidence of viral rebound.
Introduction

The SARS-CoV-2 main protease inhibitor nirmatrelvir is a drug plagued by contradictions. In a landmark, randomized, double-blinded placebo-controlled clinical trial with 1364 analyzed individuals, 300 mg of nirmatrelvir boosted with 100 mg ritonavir was given twice daily for five days to high-risk individuals with SARS-CoV-2 infection within 3 days of developing symptoms. Compared to placebo, nirmatrelvir reduced the combined outcome of hospitalization and death by 89%, eliminated death as an outcome, and reduced viral load by 0.87 log after 5 days of treatment\textsuperscript{1}. This critical result prompted the Food and Drug Administration (FDA) to issue an Emergency Use Authorization\textsuperscript{2}. The drug became the most widely prescribed antiviral for SARS-CoV-2 in the United States, likely preventing thousands of hospitalizations and many deaths\textsuperscript{3}. Ritonavir boosted nirmatrelvir was recently licensed by the FDA based on its continued effectiveness and safety\textsuperscript{4}.

However, the use of nirmatrelvir / ritonavir in real-world cohorts has identified viral rebound as a significant issue. Viral rebound occurred in 14.2% of individuals in one large cohort and was usually associated with recrudescence of symptoms, though protection against hospitalization and death appears to be maintained\textsuperscript{5}. Similar rates of viral rebound were observed between molnupiravir and nirmatrelvir suggesting the rebound effect is not drug specific and may pertain to characteristics of SARS-CoV-2 infection and treatment duration\textsuperscript{6}. This high incidence of viral rebound exceeded the 2.3% rate observed in the proof-of-concept trial which did not differ from placebo\textsuperscript{7}.

Despite its high efficacy as an early symptomatic therapy for high-risk individuals, nirmatrelvir / ritonavir was not authorized for use as post-exposure prophylaxis (PEP). In a clinical trial of post-exposure prophylaxis, nirmatrelvir / ritonavir showed only 32% and 37% reductions in symptomatic COVID-19 relative to placebo when given for five or ten days respectively\textsuperscript{8}. However, neither of these results reached statistical significance. Notably, molnupiravir, another drug that reduced hospitalization when given during early symptomatic infection, also failed as post-exposure prophylaxis\textsuperscript{9}. Only long-acting monoclonal antibodies have demonstrated efficacy for post-exposure prophylaxis\textsuperscript{10-12}, but these are no longer active against prevalent circulating strains\textsuperscript{13}.

Early during the COVID-19 pandemic, multiple groups employed mathematical models to predict the outcomes of clinical trials for SARS-CoV-2\textsuperscript{14-20}. These models all accurately predicted that antiviral therapy that was insufficiently potent or given too late during infection might fail to provide clinical benefit\textsuperscript{14-17,19}. Our previous modeling results further suggested that viral rebound may occur and was more likely if a drug was dosed during the pre-symptomatic phase of infection when viral loads are still expanding, as occurs in a post-exposure prophylaxis scenario\textsuperscript{21}. The proposed mechanism of this effect was that reducing viral load may blunt early immune responses and preserve susceptible cells, allowing viral re-expansion upon cessation of treatment that was of insufficient potency to eliminate all infected cells\textsuperscript{22}. The model suggested that this phenomenon could theoretically occur during early symptomatic treatment as well. At the time, we downplayed the significance of model-generated rebound as the phenomenon had yet to be demonstrated clinically. However, models fit to rebound data now suggest a similar mechanism of action to explain viral rebound\textsuperscript{23}.

Here we use an updated model for SARS-CoV-2 viral kinetics that was first validated against a much larger panel of untreated individuals to precisely simulate the virologic
outcomes of the nirmatrelvir / ritonavir trial. We identify that the true \textit{in vivo} potency of nirmatrelvir is approximately 60-fold less than its \textit{in vitro} potency, such that drug levels are sub-therapeutic during a portion of the dosing interval. Viral rebound is observed in our simulations and is more likely when the drug is dosed early during infection and is not reduced with a higher dose or higher dosing frequency. Extended-duration treatment is identified as the best strategy to avoid viral rebound.
Results

Viral Dynamic, Pharmacokinetic, and Pharmacodynamic Mathematical models

To derive parameters for simulating nasal viral loads in the absence of therapy, we used the mechanistic mathematical model that best recapitulated 1510 SARS-CoV-2 infections in a cohort of 2678 SARS-CoV-2 infected individuals from the National Basketball Association cohort (Figure 1a) 24. The model is target-cell limited, with viral load-dependent infectivity. The viral production by infected cells is delayed by an eclipse phase. In keeping with an early interferon-mediated innate immune response, susceptible cells can become refractory to infection based on the total number of productively infected cells but also revert to susceptible at a constant rate. Infected cells are cleared by a density-dependent early immune response in which the lifespan of infected cells decreases as a function of the number of infected cells, and delayed acquired immunity which is activated in a time-dependent fashion. Model parameters were estimated for 589 individuals in the NBA cohort who had confirmed positive symptom status using a mixed-effect population approach implemented in Monolix (Fig S1).

To generate the placebo arm, we simulated the viral load of 500 randomly selected individuals from the symptomatic subgroup of the NBA cohort, using their estimated individual viral load parameters. The mean viral load drop from the baseline recapitulates the mean change from the baseline of the viral load observed in the control arm of nirmatrelvir clinical trial 1 (Error! Reference source not found.a), confirming that our virtual cohort is a good representation of the population studied in the clinical trial.

The confirmed symptomatic population of the NBA cohort consisted of mostly omicron infections (not delta as in the clinical trial). There were only 163 delta cases in the NBA cohort, and they did not have a known time of symptom onset. However, to validate our model, we repeated the simulation with delta cases. For their symptom onset, we randomly assigned all individuals an incubation period selected from a gamma distribution with parameters reported in the literature 25. The delta group represented the control arm of the trial well (Fig S4a).

To reproduce levels of nirmatrelvir, we used a two-compartmental pharmacokinetic (PK) model (Error! Reference source not found.b). Using Monolix and the mixed-effect population approach, we estimated parameters values by fitting the model to the plasma concentration of healthy subjects. The model closely recapitulated observed drug levels following a single dose (Fig S2). The effect of ritonavir as an inhibitor of nirmatrelvir’s metabolism is accounted for in the nirmatrelvir’s clearance rate in the PK model.

For the pharmacodynamic (PD) model, we assumed the efficacy of the drug follows a Hill equation with respect to the drug concentration. We parameterized the Hill equation using in vitro efficacy data collected at different concentrations of nirmatrelvir (Fig S3). The PD model is described in further detail in the Materials and Methods section.

To simulate the treatment arm, we combined VL, pharmacokinetic (PK), and pharmacodynamic (PD) models. Sets of VL parameters for individuals are again drawn from the NBA cohort. The PK and PD parameters for all simulated individuals were randomly drawn from their estimated population distributions. The efficacy of the treatment is calculated from the Hill equation using plasma concentrations of the drug obtained from the PK model. The efficacy of the treatment is used to lower the viral reproduction rate (details in Materials and Methods).

Reduction in in vivo nirmatrelvir potency relative to in vitro
To obtain the PD parameters of nirmatrelvir, we fit the Hill equation to the in vitro efficacy of the drug as a function of its concentration (Fig S3). However, the in vivo potency of a drug is known to be different from values measured in vitro 21,26,27. The potency reduction factor (prf) is defined as the ratio between the in vivo and in vitro IC50. Here the in vivo IC50 is plasma drug concentration required to inhibit viral replication by 50%. To identify the in vivo potency of nirmatrelvir, we estimate the potency reduction factor (prf) that achieves the best fit between our VL+PKPD model and the average drop in viral load of the treatment arm of the clinical trial.

Specifically, to estimate the prf, we simulated the viral load of our virtual cohort of 500 individuals treated with 300 mg of nirmatrelvir twice per day for five days with prf ranging from 1 (no reduction in potency) to 120. The treatment start day was randomly selected from a uniform distribution for each simulated individual to be within 3 days of symptom onset. We fit the average change from baseline in simulated viral load data of the treatment arm to the trial data. We then plotted the coefficient of determination R² of the fit against different prf values (Fig 2c). The best value (prf = 61) was determined by maximizing the R² of the fit and it closely recapitulated the trial data of the treatment arm (Fig 2b). We repeated the simulation 10 times to get the standard error of the prf. In the absence of individual viral load data of the trial’s treatment arm, we could only estimate the average prf and it is expected that the prf value may vary among individuals. The boxplot shown in the lower panel of Fig 2c only represents the standard error of the prf average value and does not reflect individual variability. Using only delta cases with randomly assigned symptom onset, the simulated treatment arm closely fits the trial data with an estimated prf = 69 (Fig S4b,c).

To illustrate the importance of estimating in vivo potency of the drug, we compared the PKPD projection and average change in viral load of treatment arms with prf = 1 (no reduction in potency) and prf = 61. With an approximately 61-fold weaker potency, the drug levels dropped below the therapeutic level shortly after each dose and the antiviral effect subsided in less than a day after the end of treatment leading to an average efficacy of 82% over the first 5 days of treatment (Fig 2d, e). However, the plasma concentration of a perfectly potent drug remained above therapeutic levels for the duration of the treatment with a 5-day average efficacy of 99.99% and the effect persisted for nearly 10 days (Fig 2e). With the assumed in vitro potency level, the same treatment regimen could reduce the viral load by approximately 3.5 logs relative to the placebo compared to the 0.87 log reduction reported in the trial (Fig 2f).

**Frequent viral rebound on nirmatrelvir**

In all cases, simulations were performed from time of infection to 30 days after symptom onset. We monitored viral load continually, and specifically on days 2, 5, and 10 after treatment initiation to match the trial. We defined rebound in the treatment arm as any instance in which a post-treatment viral load exceeded the viral load at the end of the treatment by 1 log. We defined rebound in the control arm as any case with at least two peaks in the viral load trajectory with minimum heights of 3 logs and a second peak higher than its minimum by at least 1 log. By this definition, we observed a rebound in 19% of cases treated with the clinical trial dose and 3% of controls (Fig 3). However, when an equivalent definition of rebound was used as in the trial (1 log increase in viral load 5 days after treatment cessation), the probability of rebound was lower (3.6% if treatment is assumed to be several days after symptoms), equal to that of the controls, and comparable to that observed in the trial (Fig S5).
Limited impact of nirmatrelvir dose or dosing frequency on viral rebound

We next explored different treatment regimens to estimate their impact on lowering viral load and the chance of rebound. We simulated the therapy with 150, 300, 600, and 900 mg doses administered twice per day for 5 days, starting within 3 days post symptom onset. A larger dose decreased viral load more significantly and quickly than 300 mg twice daily. 900 mg of nirmatrelvir reduced the viral load by a mean of 3 logs on day 2 and a mean of 5 logs compared to the control (Fig 3a).

Individual viral loads were highly variable within each treatment group regardless of dose (Fig 3a). This was due to several factors including heterogeneous viral load trajectories (Fig S1) and different timing of treatment. Responses to treatment differed substantially according to viral load trajectory and treatment timing as well (Fig 3b). In nearly every case, the reduction in viral load was greater during the first 5 days of treatment with higher doses though this only impacted viral elimination in certain cases (Fig 3b, i & iv). Sometimes viral load equilibrated to the same level post-treatment regardless of dose (Fig 3b, ii), while in others, highest doses are associated with rebound (Fig 3b, iii). Based on achieving a lower post-treatment viral load nadir, higher doses resulted in a greater likelihood of viral rebound in our simulations (Fig 3c).

Increasing frequency of antiviral dosing had nearly equivalent effects leading to a more rapid reduction in viral load (Fig 4a), heterogeneous effects based on viral load trajectory and timing of treatment (Fig 4b), and increased rate of rebound (Fig 4c).

Early treatment as a predictor of SARS-CoV-2 rebound

We next simulated therapy with four different timings of treatment: post-exposure prophylaxis (PEP): 0-1 day after infection in the pre-symptomatic phase; early treatment: 0-1 day after symptom onset as often occurs in community settings; intermediate treatment: 1-5 days after symptom onset as in the clinical trial; and late treatment: 5-10 days after symptom onset. In all simulations, the administered dosage was 300mg twice per day for 5 days.

Applying treatment as a PEP or shortly after symptoms appeared lowered viral load more substantially relative to control than intermediate or late therapy at days 2 and 5 post treatment, though intermediate and late strategies also significantly lowered viral load relative to control at these timepoints (Fig 5a). However, mean viral load was significantly higher in PEP and early treatment groups versus the control group 10 days after the end of the treatment (Fig 5a), due to high probability of rebound at these timepoints (Fig 5b, c) when the virus is at its initial stages of expanding in the body and before the immune response is established.

Prolongation of treatment to reduce the probability of SARS-CoV-2 rebound

Next, we analyzed the impact of treatment duration on viral rebound. We simulated treatment regimens with 300 mg nirmatrelvir given twice per day for 2, 5, 10, 15, and 20 days. The treatment was again initiated within 3 days after symptoms appear. Fig 6a demonstrates the continuous drop in viral load if treatment is ongoing until the infection is effectively cleared from the body. The viral load distributions of the treatment arms with 15 and 20 days of treatment on days 2, 5, and 10 are the same as the viral load distribution of the treatment arm with 10 days of treatment duration and therefore are not shown. Prolonging treatment duration lowered the chance of viral rebound to the point of almost completely eliminating it if the therapy continued for 15 days (Fig 6b,c).
We next explored the impact of duration on different treatment timing. Prolonging treatment to 15 days for early treatment and 20 days for PEP lowers the viral load close to the limit of detection (1 log) and significantly lowered the probability of rebound (Fig 7).

Differing observed rebound rates resulting from varying timing of sampling and definitions

Differing criteria have been used to define rebound in the previous studies with varying virologic thresholds, as well as timing and frequency of sampling. A rebound has sometimes been flagged when a positive test was observed after a negative test. In the clinical trial, treatment in the treatment arm was started within the first 5 days of symptoms (our intermediate treatment group). Rebound was defined as a 0.5 log increase on day 10 or 14, or both if both data were available, and 2.3% rebound cases were observed. The probability of rebound was higher if both data were available, and 2.3% rebound cases were observed. The probability of rebound in our simulation with a threshold of 0.5 log measured only on day 5 after the end of the treatment was 4.7% but decreased as thresholds for viral rebound increases (Fig S5). This percentage would be lower if treatment started day 2-5 days after symptoms because the probability of rebound is very sensitive to the timing of treatment, and we hypothesize that the participant enrollments were skewed to later during the first 5 days of symptoms in the trial.

In this paper, we recorded the viral load every 0.001 of a day and used 1 log threshold to identify rebound cases. This would be a far more sensitive method to observe rebound and suggests that in the trial and real-world cohorts, rebound is likely more common than observed (Fig S5).

Mechanisms for viral rebound

To understand the mechanisms and factors that might explain the increase in rebound in the PEP and early treatment groups, we simulated four treatment arms with the treatment starting on days 1, 4, 7, and 10 after infection. The start day was fixed for all individuals in each arm to limit the added variability introduced by variable incubation period and timing of treatment relative to symptoms in our previous simulations. The high frequency of rebound in day 1 and day 4 treatment starts are evident from the viral load trajectories (Fig 8 top row), in many individual trajectories (grey lines) as well as the mean viral load (blue line). A second peak after the end of the treatment can also be seen in the dynamics of infected cells (Fig 8 middle row, blue line) as well as the intensity of the early immune response (Fig 8 bottom row).

Applying the treatment earlier during infection (day 1 and day 4 in the case of our simulations) lowered the viral load and subsequently the population of infected and refractory cells, preserving susceptible cells. The ratio of susceptible to refractory cells in the two groups with earlier treatment starting points (day 1 and day 4) was significantly higher than in the control group at equivalent time points (Fig S6). The rate of early immune responses also decayed during the treatment due to a drop in the infected cells. Overall, a weaker immune response and higher availability of susceptible cells lead to a viral rebound after the treatment.

In a parallel manuscript, we subset shedding groups in the NBA cohort according to shedding kinetics using k-means clustering. The groups are ordered based on the area under their viral load curve (AUC) with group 1 having the smallest AUC and group 6 the largest. We simulated treatment with different treatment start days using these 6 groups and identified the highest rebound probability in the earlier treatment groups with the larger AUC (groups 5 and 6) and longer time to peak viral load (groups 3, 5 and 6) prior to antiviral therapy (Fig S7). This
indicates that viral rebound may be more likely in individuals who were destined for more severe infections off therapy.
Discussion

We previously demonstrated for herpes simplex virus-2 30, HIV 31, Ebola virus 26, and for SARS-CoV-221, that it is vital to consider the timing and intensity of the immune response to accurately simulate clinical trials of antiviral agents. If a direct-acting antiviral therapy is given too late during infection, then efficacy is often low because the disease is driven by excess inflammation and cytokine storm. On the other hand, concurrent immune pressure can provide critical assistance for antiviral agents to eliminate viral replication. Accordingly, our previous modeling suggested that extremely early treatment of pre-symptomatic SARS-CoV-2 as occurs with PEP requires higher drug potency than treatment during early symptomatic infection because innate immunity is activated to a greater extent at this slightly later stage of infection and fewer susceptible cells remain 21. It is increasingly clear that the potency and duration of antiviral therapy required to achieve clinical benefit depends strongly on the stage of infection and ongoing intensity of the immune response.

Our prior work also demonstrated that in vitro antiviral drug potency measured in relevant cell culture lines often overestimates in vivo potency in humans 26,27,32. Specifically, the plasma drug level required to achieve 50% inhibition of cellular infections in vivo is higher than the level required to inhibit infection in vitro. The discrepancy between in vitro and in vivo potency can only be assessed by fitting viral dynamic / PK / PD mathematical models to viral load data from clinical trials as we have done here. Traditional PK / PD models which do not account for the dynamics of an immune response on observed viral loads are not sufficient to estimate in vivo potency. Because in vivo potency reduction varies from 2 to 100 depending on the infection and antiviral agents, 26,30,32 in vivo IC50 must be assessed separately in each case.

Here by precisely fitting a combined viral-immune dynamic / PK / PD model to viral load data from placebo and treatment groups in a randomized clinical trial of nirmatrelvir / ritonavir, we merge these two key concepts. We first identify that nirmatrelvir potency is reduced 60-70 fold in vivo relative to in vitro. The mechanistic reasons for this reduction cannot be determined by the model but may include increased in vivo protein binding 33, inhibition of drug delivery from plasma to sites of infection, or differences in cellular uptake and drug metabolism in vivo 34. Nevertheless, our estimated IC50 provides a benchmark plasma level to target in future trials. The PK model also demonstrates that the drug's relatively short half-life allows it to dip to subtherapeutic levels even when dosed twice daily.

Our model also develops a viable hypothesis for why nirmatrelvir is highly effective when given during early symptomatic infection but less so when given as post-exposure prophylaxis. By preventing a high peak viral load approximately 3-5 days after infection, therapy preserves susceptible cells and blunts the immediate, likely innate immune response to SARS-CoV-2. If the virus is not eliminated by an early acquired response along with antiviral pressure, it rebounds to a peak level that is sometimes comparable to the initial peak. We hypothesize that viral rebound occurs more frequently in community settings relative to the clinical trial because infected individuals in the community are often prescribed the drug very early after symptom development whereas in the trial there was a natural 1 to 2-day delay based on the enrollment and consent process. Surprisingly, this short delay may have limited rebound while not affecting the primary endpoints of the trial, a finding supported by recent clinical studies 35. Notably, antiviral therapy is not a risk factor for rebound in our model or in clinical cohorts of individuals.
treated late during infection\textsuperscript{36}. High viral load shedding is also a risk factor for rebound in our model as has been suggested in other studies\textsuperscript{37}.

The model identifies optimal conditions for viral rebound which counterintuitively include early treatment during pre-symptomatic infection which can be exacerbated by higher or more frequent dosing. Both mechanisms occur by suppressing the amount of infection and preserving susceptible cells, limiting development of refractory cells and dampening intensity of the early immune response. The best method to prevent viral rebound is prolonging treatment, with a longer course needed for PEP. This finding is consistent with trials of long-acting monoclonal antibodies which demonstrated efficacy as post-exposure prophylaxis\textsuperscript{10–12}.

Because the model is validated precisely against data, it can be used as a tool to test various treatment strategies for future trials with the ability to vary therapeutic goals, timing of treatment, dose, dosing interval, and duration of therapy. Our prior PD modeling also allows testing of potentially synergistic combination agents and consideration of special hosts such as immunocompromised individuals with persistent infection\textsuperscript{26}. We believe our approach provides a template for optimizing future trial designs with nirmatrelvir and other therapies.

Our model has several limitations. First, nasal viral load may not be a perfect surrogate of disease activity. On the one hand, viral load reduction has been correlated with beneficial clinical outcomes for nirmatrelvir\textsuperscript{1}, molnupiravir\textsuperscript{38}, and monoclonal antibodies\textsuperscript{39}. A recent review shows that viral load reduction is a reasonably good surrogate endpoint\textsuperscript{39}. Moreover, the viral rebound appears to track very closely with the symptomatic rebound in multiple case series\textsuperscript{28}. Yet, early remdesivir treatment provided a profound reduction in hospitalization while not impacting nasal viral load\textsuperscript{40}. Data from non-human primates suggests that the drug has a specific effect on viral loads in the lungs that is not observed in upper airways, a finding that we were also able to capture with models\textsuperscript{21}. Overall, the data suggests that in early treatment trials, a reduction in nasal viral loads beyond that observed in placebo treated individuals is associated with substantial clinical benefit\textsuperscript{1}.

Another limitation is that the model does not account for drug resistance. While there has been limited evidence of \textit{de novo} resistance during nirmatrelvir therapy, serial passage of virus suggests a relatively low barrier, and some viral rebound could, in theory, be with resistant variants. Studies to date suggest very little mutational change between the infecting and rebounding virus\textsuperscript{41–44}.

Our model does not capture the immune response in literal terms. For instance, we do not distinguish innate interferon, antibody and T cell responses as these have not been measured in sufficient longitudinal detail to precisely ascribe viral clearance with different components of the immune response. We structured the model for the early response to roughly map to innate responses as this model term does not allow immune memory. The progression of susceptible cells to a refractory state also diminishes with decreases in viral load. The late immune response in our model has memory, leads to rapid elimination of virus, and is likely to represent acquired immunity. While a more accurate model would discriminate different arms of the immune responses and fit to immune data, ours sufficiently captures the timing and intensity of immune responses for accurate clinical trial simulation.

Finally, it is our opinion that models lacking a spatial component cannot capture the full dynamics of target cell limitation which are influenced by the packing structure of cells, dynamics of viral diffusion, and infection within multiple concurrent micro-environments\textsuperscript{30}. For these reasons, ordinary differential equations may misclassify the relative impact of target cell
limitation and innate immune responses in the period surrounding peak viral load. However, the approach seems sufficient for accurate clinical trial simulation.

In conclusion, our model identifies viable mechanistic underpinning of the high efficacy of nirmatrelvir therapy for early symptomatic SARS-CoV-2 infection, lower efficacy for PEP, and high incidence of viral rebound in a real-world setting. The model also can be used to assess different treatment strategies and suggests prolonging therapy is the optimal method to avoid rebound and maintain potent early antiviral suppression.
Materials and Methods

Study Design
We developed a viral dynamics model recapitulating the viral load data collected from symptomatic individuals in the NBA (National Basketball Association) cohort. We used a two-compartmental model to reproduce the PK data of Nirmatrelvir plus Ritonavir. For the purpose of the simulation, we constructed a virtual cohort by randomly selecting 500 individuals from the NBA cohort and assigning individual PK and PD parameters randomly drawn from their respective inferred distributions. We fit the combined viral dynamics and PK/PD model to the average change in viral load from the baseline of the control and treatment arms of the previously published nirmatrelvir/ritonavir clinical trial. By fitting our model to the control arm, we validated our viral dynamics model and how well our virtual cohort represents the trial control arm. We used the fit to the treatment arm to estimate the potency reduction factor (prf) by maximizing the $R^2$ of the fit. With the estimated prf and in vivo IC$_{50}$ of the drug, we explored different treatment regimens by changing dose, dosing frequency, treatment duration, and treatment timing, to find the most strategy to minimize the probability of rebound.

Viral load data
We used data from the symptomatic subpopulation of the NBA cohort published by Hay et al. The NBA cohort dataset consists of 2875 documented SARS-CoV-2 infections in 2678 people detected through frequent PCR testing regardless of symptoms. 1510 infections in 1440 individuals had at least 4 positive quantitative samples of which 756 infections in 748 individuals had test results through 20 days after detection of infection or two consecutive negative tests prior to day 20 which indicated confirmed elimination of virus. Among these 756 well-documented infections, there were 589 infections with confirmed positive symptom status and recorded onset of symptoms. We used the viral load data from the 589 symptomatic infections to estimate the viral load parameters. In a separate analysis, we included 163 delta virus infections, including those with no documentation of whether there were symptoms.

Clinical trial data
We obtained the average change in viral load data of the control and treatment arms by digitizing Figure 3A of the published nirmatrelvir clinical trial by Hammond et al. The trial included 682 and 697 symptomatic high-risk individuals in the control and treatment arms respectively. The study participants were treated with a placebo or 300mg/100mg nirmatrelvir/ritonavir within three days of symptoms onset as we used in all our simulations unless mentioned otherwise. The treatment was administered twice per day, for five days. Viral load was measured on days 0, 3, 5, 10, and 14 after the treatment start day and adjusted by the baseline viral load.

PK data
PK data of nirmatrelvir (PF-07321332) with ritonavir was obtained by digitizing Figure 4 of the drug’s Emergency Use Authorization document. The data is from a phase I randomized trial by Singh et al. where eight participants (4 fed, 4 not fed) took a single dose of 250 mg nirmatrelvir plus 100 mg ritonavir. The plasma concentrations of the drug in participants were recorded in the next 48 hours after dosing.
PD data

The data on drug efficacy comes from five in vitro dose-response experiments we performed at the University of Washington. The efficacy of Nirmatrelvir in the presence of CP-100356 (an efflux inhibitor) is measured against the delta variant of SARS-CoV2 in Calu-3 cells. The efflux inhibitor is meant to replace the role of ritonavir in vivo. Briefly, Calu 3 cells human lung epithelial were treated with varying concentrations of nirmatrelvir (PF-07321332) in the presence of 2uM CP-100356 prior to infection with SARS-CoV-2 (delta isolate) at a multiplicity of infection of 0.01. Antiviral efficacy and cell viability (of non-infected cells treated with drugs) were assessed as described 47.

Viral dynamics model

We used our model of SARS-CoV-2 dynamics 24 to model the viral load dynamics of symptomatic individuals with SARS-CoV-2 infection. Our model assumes that susceptible cells ($S$) are infected at rate $\beta S V$ by SARS-CoV-2 virions. The infected cells go through a non-productive eclipse phase ($I_E$) before producing viruses and transition to becoming productively infected ($I_P$) at rate $\kappa I_E$. The susceptible cells when encountering the productively infected cells become refractory to infection ($R$) at the rate $\phi I_P S$. Refractory cells revert to a susceptible state at rate $\rho R$. The productively infected cells are cleared at rate $\delta I_P^h + 1$ representing the innate immune response that lacks memory and is proportional to the amount of ongoing infection. If the infection persists longer than $\tau$, the cytotoxic acquired immunity gets involved which in our model is represented by the rate $m(t) I_P$. Finally, free virions are cleared at the rate $\gamma$. Of note, this model was selected against other models in 24 based on superior fit to data and parsimony. The model is written as a set of differential equations has the form,

$$\frac{dS}{dt} = -\beta SV - \phi I_P S + \rho R \quad (1a)$$

$$\frac{dR}{dt} = \phi I_P S - \rho R \quad (1b)$$

$$\frac{dI_E}{dt} = \beta SV - \kappa I_E \quad (1c)$$

$$\frac{dI_P}{dt} = \kappa I_E - \delta I_P^h I_P - m(t) I_P \quad (1d)$$

$$\frac{dV}{dt} = \pi I_P - \gamma V \quad (1e)$$

where $\begin{cases} m(t) = 0 & t < \tau \\ m(t) = m & t \geq \tau \end{cases} \quad (1f)$

To estimate the parameters, we fit the model to the viral load data from the symptomatic individuals in the NBA cohort using a mixed-effect population approach implemented in Monolix.

We start the simulations with $10^7$ susceptible cells. The initial value of the refractory cells is assumed to be zero since the interferon signaling is not active prior to infection. We
further assume there are no infected cells (eclipse or productive) at the beginning of the infection. We estimate the level of inoculum \( (V_0) \) for each individual.

To resolve identifiability issues, we fixed three parameter values, setting the inverse of the eclipse phase duration to \( \kappa = 4 \), the rate of clearance of virions to \( \gamma = 15 \).  

**PK model**

We used a two-compartmental PK model which includes the amount of drug in the GI tract \( (A_{GI}) \), the plasma compartment \( (A_p) \), and the lung \( (A_L) \). The drug is administered orally passes through the GI tract, and gets absorbed into the blood at the rate \( \kappa_a \). The drug then transfers from the blood into the peripheral compartment (or the lung) at the rate \( \kappa_{PL} \). The metabolized drug transfers back into the plasma at the rate \( \kappa_{LP} \) from where it clears from the body at the rate \( \kappa_{CL} \). The model in the form of ordinary differential equations is written as,

\[
\begin{align*}
\frac{dA_{GI}}{dt} &= -\kappa_a A_{GI} \\
\frac{dA_p}{dt} &= \kappa_a A_{GI} + \kappa_{LP} A_L - (\kappa_{CL} + \kappa_{PL}) A_p \\
\frac{dA_L}{dt} &= \kappa_{LP} A_p - \kappa_{LP} A_L
\end{align*}
\]

(2a)  
(2b)  
(2c)

We used Monolix and a mixed-effect population approach to estimate the parameters and their standard deviations. With the initial condition of \( (A_{GI} = Dose, A_p = 0, A_L = 0) \); we fit \( C_p = \frac{A_p}{Vol} \) to the plasma concentration data where \( Vol \) is the estimated plasma volume.

**PD model**

For the pharmacodynamics model we used Hill equation, \( \epsilon(t) = \frac{E_{max} C(t)^n}{C(t)^n + IC_{50}^n} \), where \( C(t) \) is the drug’s concentration in plasma, \( E_{max} \) is the maximum efficacy, \( n \) is the hill coefficient, and \( IC_{50} \) is the drug concentration in plasma required to provide 50% efficacy. We used least-squared fitting to obtain the three parameters and their standard deviations.

**Combined PKPD and VL models**

The plasma concentration of nirmatrelvir obtained from the PK model is used in the PD model to obtain time-dependent efficacy. \( \epsilon(t) \), then, is used to reduce viral production rate, \( \pi \), with the factor of \( (1 - \epsilon(t)) \). Equation 1e is written as,

\[
\frac{dv}{dt} = (1 - \epsilon(t)) \pi I_P - \gamma V
\]

(4)

**Construction of a virtual cohort**

To generate a cohort for our simulated clinical trials, we randomly selected 500 individuals (for each arm of the simulated trial) from the symptomatic subpopulation of the NBA cohort and used their individual viral load parameters estimated by fitting our viral dynamics model to the data. PK parameters of each simulated individual were randomly drawn from the lognormal distributions with their estimated mean and standard deviation inferred from PK data. The PD parameters were also randomly drawn from the normal distribution with the
estimated mean and standard deviation. The standard deviation of the PD parameters represents the accuracy of the assays and not the individual variability.

**Potency reduction factor (prf)**

The potency reduction factor (prf) is defined as,

\[ prf = \frac{IC_{50, \text{in vivo}}}{IC_{50, \text{in vitro}}} \]  \hspace{1cm} (3)

We estimate prf by fitting the change in viral load of the treatment arm of our simulation to the treatment arm of the clinical trial and maximizing the R² of the fit.

**Measuring rebound probability**

A viral load rebound in the treatment arm is defined when the viral load at any time after the treatment exceeds the viral load at the end of the treatment by 1 log. In the control group, viral rebound is defined in patients who have at least two peaks with maximum height of 3logs in their viral load trajectories and the second peak is 1log higher than its local minimum.

**Data availability**

The data analyzed in this work was previously published by Hay et al. and is available on github at [https://github.com/gradlab/SC2-kinetics-immune-history](https://github.com/gradlab/SC2-kinetics-immune-history).

**Acknowledgments**

We thank Judith White for helpful discussions regarding the pharmacodynamics data and modeling. This work was supported by National Institutes of Health (NIH) grants R01AI169427 and R01AI121129.
Figure 1. Schematics of the viral dynamic model and Paxlovid PK-PD two compartmental model. a) The viral dynamic model follows the dynamics of susceptible cells ($S$), refractory cells ($R$), eclipse infected cells ($I_E$), productively infected cells ($I_P$), virus ($V$) and includes the early and late (T-cell mediated) immune responses with rates $\delta^1 h$ and $m(t)$. $\beta$ is the infection rate, $\rho$ is the rate of reversion of refractory cells to susceptible cells. Infected cells produce viruses at the rate $\pi$, and the free viruses are cleared at the rate $\gamma$. b) two-compartmental PK model with oral administration of the drug which models the amounts of the drug in gut tissue ($A_{GI}$), plasma ($A_P$), and the tissue ($A_L$). $K_a$ is the rate of absorption of the drug from gut to plasma. $K_{PL}$ and $K_{LP}$ are the rates of transfer of the drug from plasma to the tissue and back, and $K_{CL}$ is the rate at which the drug clears from the body. $V$ is the estimated plasma volume and $C_P$ is the drug concentration in plasma. $\epsilon(C_P)$ is the drug efficacy that blocks viral production and is calculated using the Hill equation: $\frac{E_{max}C_P^n}{C_P^n + (IC_{50}^{n+1} \times prf)}$ where $E_{max}$ is the maximum efficacy, $n$ is the Hill coefficient, $IC_{50}$ is the concentration of drug in vitro at which viral replication rate is reduced by 50%, prf is the potency reduction factor translating the in vitro potency to in vivo potency.
Figure 2. Lower in vivo potency of nirmatrelvir relative to in vitro potency. (a-b) mean (blue), individual (gray), and ranges (labeled dashed lines) of log10 viral load drop from the baseline of individuals randomly selected from the NBA cohort treated with (a) placebo or (b) five days of nirmatrelvir / ritonavir 300 mg twice daily. The red dots were obtained by digitizing Fig 3a of Hammond et al.1 and model fit was noted by closeness of blue lines to the red dots. (c) R² of the fit of the 10 model simulations per prf to the viral load drop data in light blue and their mean in dark blue. The best model fit was at a potency reduction factor of 61. The boxplot shows the distribution of prf values at which R² is maximum. (d) Drug efficacy when prf=61. Average efficacy was 82% over the 5-day interval with notable drops in antiviral efficacy at drug throughs. (e) Average drug efficacy when prf = 1 vs prf = 61. The drug with no potency reduction has nearly perfect efficacy (average efficacy of 99.99%) over 5 days and has a prolonged post treatment effect. (f) mean log10 viral load drop from baseline of the control arm, treatment arm with prf=61, and treatment arm with prf=1.
Figure 3. Increasing nirmatrelvir dose lowers short term viral load but increases probability of viral rebound. In all scenarios, treatment starts within the first 3 days post-symptoms. (a) log10 viral load at days 2, 5, and 10 after the treatment start day with different doses. p-values are obtained by performing Mann-Whitney U-test between 300 mg group and others and only p-values <0.01 are shown. Viral loads are only reduced by higher dose at days 2 and 5, but not day 10. (b) Examples of viral load trajectories assuming different dose on 4 modeled individuals with equivalent timing of therapy and untreated viral kinetics. (c) The probability of rebound for different doses. The error bars on each column are 95% confidence intervals.
Figure 4. Increasing nirmatrelvir dosing frequency lowers short term viral load but increases probability of viral rebound. In all scenarios, 300 mg treatment starts within the first 3 days post-symptoms. (a) log10 viral load at days 2, 5, and 10 after the treatment start day with different dosing frequency. p-values are obtained by performing Mann-Whitney U-test between 2/day group and others and only values<0.01 are shown. Viral loads are only reduced by higher dosing frequency at days 2 and 5, but not day 10. (b) Samples of viral load trajectories assuming different dosing frequency on 4 modeled individuals with equivalent timing of therapy and untreated viral kinetics. (c) The probability of rebound for different doses. The error bars on each column are 95% confidence interval.
Figure 5. Early timing of therapy initiation is a key risk factor for viral rebound. In all scenarios, the dose is 300 mg twice daily for five days. (a) log10 viral load at days 2, 5, and 10 after the treatment start day with different treatment durations. P-values are obtained by performing Mann-Whitney U-test. At day 10, the treatment group has higher viral loads compared to placebo due to viral rebound in the PEP and early treatment simulations, despite lowering viral loads significantly at days 2 and 5. (b) Samples of viral load trajectories assuming different treatment timing on 4 modeled individuals with equivalent untreated viral kinetics. (c) The probability of rebound for different treatment timing. The error bars on each column are 95% confidence interval.
Figure 6. Prolonging treatment duration limits rebound probability. In all scenarios, treatment starts within the first 3 days post-symptoms and the dose is 300 mg twice daily. (a) log10 viral load at days 2, 5, and 10 after the treatment start day with different treatment durations. p-values are obtained by performing Mann-Whitney U-test and only values <0.01 are shown. At day 10, the control group has equivalent viral loads to 5 days of treatment while 10 days of treatment significantly lowers viral load. (b) Samples of viral load trajectories assuming different treatment durations on 4 modeled individuals with equivalent timing of therapy and untreated viral kinetics. Prolonging therapy often avoids rebound. (c) The probability of rebound for different treatment durations. The error bars on each column are 95% confidence interval.
Figure 7. Post-exposure prophylaxis requires more prolonged therapy than early symptomatic therapy to avoid viral rebound. (a) Probability of rebound and (b) viral load at the end of the treatment as a function of treatment timing and duration.
Figure 8. Early therapy preserves susceptible cells, limits refractory cells, and delays innate immune responses. Simulations are performed using time since infection as a variable rather than based on symptoms as in prior figures to eliminate the slightly confounding impact of variable incubation period. The top row shows the viral load of all individuals (in grey) and the average viral load (in blue). The middle row shows a less substantial depletion of susceptible cells, and lower generation of refractory cells with earlier therapy. The bottom row shows the rate of early, likely innate immune responses $\delta^{in}$ (day$^{-1}$) with biphasic, lower peak responses noted with early therapy and to a lesser extent in day 4 treated individuals.
References


Supplementary Material

Tables:
- Table 1: Population VL parameters
- Table 2: Population PK parameters
- Table 3: PD parameters

Figures
- Fig S1. Mathematical model recapitulation of untreated SARS-CoV-2 kinetics
- Fig S2. PK model recapitulation of plasma concentration of nirmatrelvir
- Fig S3. PD model fitting to mean antiviral efficacy as a function of drug concentration in vitro.
- Fig S4. Clinical trial simulation with delta variant infections
- Fig S5. Sensitivity of viral rebound detection to rebound definition and sampling frequency.
- Fig S6. Preservation of susceptible cells and decreased refractory cell generation with early nirmatrelvir therapy.
- Fig S7. Higher rebound probability in individuals with larger viral area under the curve after early treatment.