

1 Timely vaccine strain selection and 2 genomic surveillance improves 3 evolutionary forecast accuracy of 4 seasonal influenza A/H3N2

5 John Huddleston^{1*} and Trevor Bedford^{1,2}

*For correspondence:
jhuddles@fredhutch.org (JH)

6 ¹Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Center, Seattle, WA,
7 USA; ²Howard Hughes Medical Institute, Seattle, WA, USA

9 **Abstract** For the last decade, evolutionary forecasting models have influenced seasonal
10 influenza vaccine design. These models attempt to predict which genetic variants circulating at
11 the time of vaccine strain selection will be dominant 12 months later in the influenza season
12 targeted by vaccination campaign. Forecasting models depend on hemagglutinin (HA) sequences
13 from the WHO's Global Influenza Surveillance and Response System to identify currently
14 circulating groups of related strains (clades) and estimate clade fitness for forecasts. However,
15 the average lag between collection of a clinical sample and the submission of its sequence to the
16 Global Initiative on Sharing All Influenza Data (GISAID) EpiFlu database is ~3 months. Submission
17 lags complicate the already difficult 12-month forecasting problem by reducing understanding of
18 current clade frequencies at the time of forecasting. These constraints of a 12-month forecast
19 horizon and 3-month average submission lags create an upper bound on the accuracy of any
20 long-term forecasting model. The global response to the SARS-CoV-2 pandemic revealed that
21 modern vaccine technology like mRNA vaccines can reduce how far we need to forecast into the
22 future to 6 months or less and that expanded support for sequencing can reduce submission lags
23 to GISAID to 1 month on average. To determine whether these recent advances could also
24 improve long-term forecasts for seasonal influenza, we quantified the effects of reducing forecast
25 horizons and submission lags on the accuracy of forecasts for A/H3N2 populations. We found
26 that reducing forecast horizons from 12 months to 6 or 3 months reduced average absolute
27 forecasting errors to 25% and 50% of the 12-month average, respectively. Reducing submission
28 lags provided little improvement to forecasting accuracy but decreased the uncertainty in current
29 clade frequencies by 50%. These results show the potential to substantially improve the accuracy
30 of existing influenza forecasting models by modernizing influenza vaccine development and
31 increasing global sequencing capacity.

33 Introduction

34 Seasonal influenza virus infections cause approximately half a million deaths per year (*World Health
35 Organization, 2014*). Vaccination provides the best protection against hospitalization and death,
36 but the rapid evolution of the influenza surface protein hemagglutinin (HA) allows viruses to es-
37 cape existing immunity and requires regular updates to influenza vaccines (*Petrova and Russell,
38 2018*). The World Health Organization (WHO) meets twice a year to decide on vaccine updates for
39 the Northern and Southern Hemispheres (*Morris et al., 2018*). The dominant influenza vaccine plat-

It is made available under a [CC-BY 4.0 International license](https://creativecommons.org/licenses/by/4.0/) .

40 form is an inactivated whole virus vaccine grown in chicken eggs (*Wong and Webby, 2013*) which
41 takes 6 to 8 months to develop and contains a single representative vaccine virus per seasonal
42 influenza subtype including A/H1N1pdm, A/H3N2, and B/Victoria (*Morris et al., 2018*). As a result,
43 the WHO must select a single immunologically-representative virus per subtype approximately
44 12 months before the peak of the next influenza season. These selections depend on the diver-
45 sity of currently circulating phylogenetic clades, groups of influenza viruses that all share a recent
46 common ancestor. The WHO's understanding of that genetic diversity comes from HA sequences
47 collected by the WHO's Global Influenza and Surveillance and Response System (GISRS) (*Hay and*
48 *McCauley, 2018*) and submitted to the Global Initiative on Sharing All Influenza Data (GISAID) EpiFlu
49 database (*Shu and McCauley, 2017*). The fastest evolving influenza subtype A/H3N2 accumulates
50 3–4 HA amino acid substitutions per year (*Smith et al., 2004; Kistler and Bedford, 2023*) such that
51 the clades circulating 12 months after the vaccine decision can be antigenically distinct from clades
52 that were circulating at the time of the decision.

53 Given the 12-month lag between the decision to update an influenza vaccine and the peak of the
54 following influenza season, the vaccine composition decision is commonly framed as a long-term
55 forecasting problem (*Lässig et al., 2017*). For this reason, the decision process is partially informed
56 by computational models that attempt to predict the genetic composition of seasonal influenza
57 populations 12 months in the future based on current genetic and phenotypic data (*Morris et al.,*
58 *2018*). The earliest of these models predicted future influenza populations from HA sequences
59 alone (*Łuksza and Lässig, 2014; Neher et al., 2014; Steinbrück et al., 2014*). Recent models include
60 phenotypic data from serological experiments (*Morris et al., 2018; Huddleston et al., 2020; Meijers*
61 *et al., 2023, 2024*), but these models still heavily rely on HA sequences to determine the viruses
62 circulating at the time of a forecast. Unfortunately, the average lag between collection of a seasonal
63 influenza A/H3N2 HA sample and submission of its sequence had been ~3 months in the era prior
64 to the SARS-CoV-2 pandemic (*Figure 1A*). While long-term forecasting models continue to improve
65 technically, the constraints of a 12-month forecast horizon and the availability of enough recent,
66 representative HA sequences impose an upper bound on the accuracy of long-term forecasts.

67 The global response to the SARS-CoV-2 pandemic in 2020 showed the speed with which we can
68 develop new vaccines and capture real-time viral genetic diversity. Decades of research on mRNA
69 vaccines enabled the development of multiple effective vaccines a year after the emergence of
70 SARS-CoV-2 (*Mulligan et al., 2020; Baden et al., 2021*). This mRNA-based vaccine platform also
71 enabled the approval of booster vaccines targeting Omicron only 3 months after the recommen-
72 dation of an Omicron-based vaccine candidate (*Grant et al., 2023*). In parallel to vaccine devel-
73 opment, expanded funding and capacity building for viral genome sequencing enabled unprece-
74 dented dense sampling of a pathogen's genetic diversity over a short period of time (*Chen et al.,*
75 *2022*). By 2021, the average time between collection of a SARS-CoV-2 sample and submission of the
76 sample's genome sequence to GISAID EpiCoV database had decreased to approximately 1 month
77 (*Brito et al., 2022*). This reduction in submission lags reflects both increased emergency funding
78 and the sustained efforts by more public health organizations to adopt best practices for genomic
79 epidemiology (*Kalia et al., 2021; Black et al., 2020*). Assessments of SARS-CoV-2 short-term fore-
80 casts have shown how such reductions in forecast horizon and submission lags can improve the
81 accuracy of short-term forecasts and real-time estimates of clade frequencies (*Abousamra et al.,*
82 *2024*).

83 These technological and societal changes in response to SARS-CoV-2 suggest that we could re-
84 alistically expect the same outcomes for seasonal influenza. Work on mRNA vaccines for influenza
85 viruses dates back over a decade (*Petsch et al., 2012; Brazzoli et al., 2016; Pardi et al., 2018; Feld-*
86 *man et al., 2019*). A switch from the current egg-based inactivated virus vaccines to mRNA vaccines
87 could reduce the time between vaccine design decisions and the peak influenza season from 12
88 months to 6 months. Similarly, the expanded global capacity for sequencing SARS-CoV-2 genomes
89 could reasonably extend to broader and more rapid genomic surveillance for seasonal influenza,
90 reducing submission lags from 3 months to 1 month on average. Even in the years immediately

It is made available under a [CC-BY 4.0 International license](https://creativecommons.org/licenses/by/4.0/) .

91 after the onset of the SARS-CoV-2 pandemic, we have observed a trend toward a reduced average
92 submission lag of 2.5 months that we would expect from increased global capacity for genome
93 sequencing (*Figure 1—figure Supplement 1*).

94 In this work, we tested the effects of similar reductions in forecast horizons and submission
95 lags on the accuracy of long-term forecasts for seasonal influenza. Building on our previously pub-
96 lished forecasting framework (*Huddleston et al., 2020*), we performed a retrospective analysis of
97 HA sequences from simulated and natural A/H3N2 populations. For each population type, we pro-
98 duced forecasts from 12, 9, 6, and 3 months prior to a given influenza season (*Figure 1A*). We made
99 each forecast under three different submission lag scenarios including a realistic lag (3 months on
100 average), an ideal lag (1 month on average), and no lag (*Figure 1B*). First, we measured the accu-
101 racy and precision of forecasts under these different scenarios by calculating the genetic distance
102 between predicted and observed future populations using the same earth mover’s distance metric
103 that we originally used to train our forecasting models (*Rubner et al., 1998*). Next, we calculated
104 the effect of forecast horizon and submission lags on clade frequencies which are the values we
105 use to communicate predictions to WHO decision-makers (*Huddleston et al., 2024*). We quanti-
106 fied the effect of reduced submission lags on initial clade frequencies, and we calculated forecast
107 accuracy as the difference between predicted and observed clade frequencies of future popula-
108 tions. Finally, we calculated the relative improvement in forecast accuracy produced by different
109 realistic interventions including reduced vaccine development time, reduced submission lags, and
110 the combination of both. In this way, we show the potential to improve the accuracy of existing
111 long-term forecasting models and, thereby, the quality of vaccine design decisions by simplifying
112 the forecasting problem through realistic societal changes.

113 Results

114 Reducing forecast horizons and submission lags decreases distances between pre- 115 dicted and observed future populations

116 Previously, we trained long-term forecasting models that minimized the genetic distance between
117 predicted and observed future populations of HA sequences (*Huddleston et al., 2020*). We pre-
118 dicted each population 12 months in the future based on the frequencies and fitness estimates
119 of HA sequences in the current population. We calculated the distance between predicted and
120 observed future populations with the earth mover’s distance metric (*Rubner et al., 1998*). This
121 metric provided an average genetic distance between amino acid sequences of the two popula-
122 tions weighted by the frequencies of sequences in each population. This approach allowed us to
123 measure forecasting accuracy without first defining phylogenetic clades, a process that can borrow
124 information from the future or change clade definitions between initial and future timepoints. We
125 identified the best forecasting models as those that minimized this distance between populations.
126 The most accurate sequence-only model for the 12-month forecast horizon estimated fitness with
127 local branching index (LBI) (*Neher et al., 2014*) and mutational load (*Łuksza and Lässig, 2014*). As a
128 positive control, we calculated the posthoc empirical fitness of each initial population based on the
129 composition of the corresponding future population. These empirical fitnesses provided the lower
130 bound on the earth mover’s distance which represented the number of amino acid substitutions
131 accumulated between populations.

132 To understand the effects of reducing forecast horizons and submission lags on long-term fore-
133 cast accuracy, we produced forecasts 3, 6, 9, and 12 months into the future using HA sequences
134 available at each initial timepoint under each submission lag scenario including no lag, ideal lag
135 (~1-month average), and realistic lag (~3-month average) (*Figure 1, Figure 1—figure Supplement 2,*
136 *Figure 1—figure Supplement 3*). For both natural and simulated populations, we assigned ideal and
137 realistic lags to each sequence from the modeled distributions in *Figure 1B*. This approach allowed
138 us to assign uncorrelated lag values to both population types while avoiding the biases associated
139 with historical submission patterns for natural A/H3N2 HA sequences. For natural A/H3N2 pop-

It is made available under a [CC-BY 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

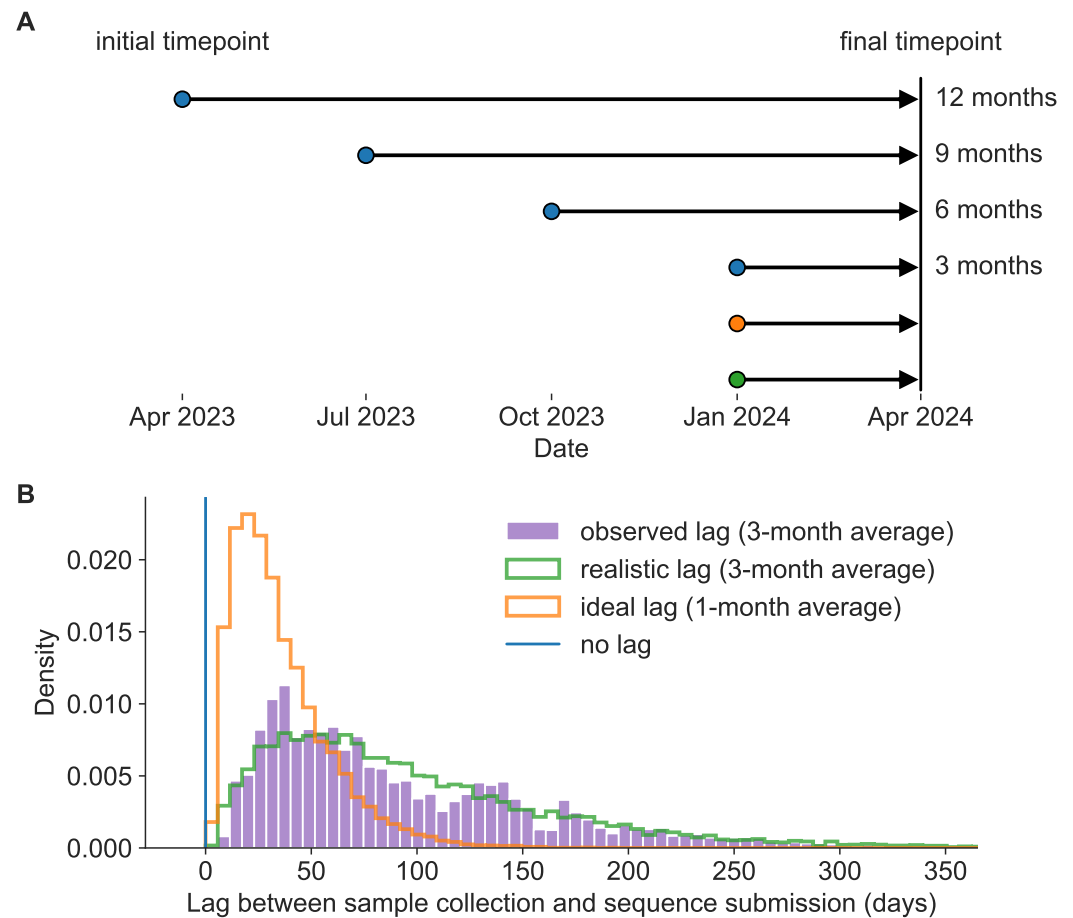


Figure 1. Model of forecast horizons and submission lags. A) Long-term forecasting models historically predicted 12 months into the future from April and October because of the time required to develop and distribute a new vaccine (Łuksza and Lässig, 2014). We tested three additional shorter forecast horizons in three-month intervals of 9, 6, and 3 months prior to the same time in the future season. For each forecast horizon, we calculated the accuracy of forecasts under each of the three submission lags reflected above including no lag (blue), realistic lag (green), and ideal lag (orange). B) Observed lags in days between collection of viral samples and submission of corresponding HA sequences to GISAID (purple) for samples collected in 2019 have a mean of 98 days (approximately 3 months). A gamma distribution fit to the observed lag distribution with a similar mean and shape (green) represents a realistic submission lag that we sampled from to assign “submission dates” to simulated and natural A/H3N2 populations. A gamma distribution with a mean that is one third of the realistic distribution (orange) represents an ideal submission lag analogous to the 1-month average observed lags for SARS-CoV-2 genomes. Retrospective analyses including fitting of forecasting models typically filter HA sequences by collection date instead of submission dates in which case there is no lag (blue).

Figure 1—figure supplement 1. Distribution of submission lags in days for the pre-pandemic era (2019-2020) and pandemic era (2022-2023)

Figure 1—figure supplement 2. Number and proportion of A/H3N2 sequences available per timepoint and lag type

Figure 1—figure supplement 3. Number and proportion of simulated A/H3N2-like sequences available per timepoint and lag type

Figure 1—source data 1. Distribution of lags between sample collection and sequence submission in pre-pandemic and pandemic eras; see <https://zenodo.org/records/13742375>

140 ulations, we used the best sequence-only forecasting model, LBI and mutational load, which we
141 previously trained on 12-month forecasts without any submission lag. For simulated A/H3N2-like

It is made available under a [CC-BY 4.0 International license](https://creativecommons.org/licenses/by/4.0/) .

142 populations, we used the observed fitness per sample provided by the simulator. For each forecast
143 horizon and submission lag type, we calculated the earth mover's distance between the predicted
144 future populations under the given lag scenario and the observed future populations without any
145 lag in sequence availability. As a control, we also calculated the optimal distance between initial
146 and future populations based on posthoc empirical fitness of the initial population. We anticipated
147 that reducing either the forecast horizon or the submission lag would reduce the distance to the
148 future in amino acids (AAs), representing increased accuracy of the forecasting models.

149 We found that reducing the forecast horizon from the current standard of 12 months linearly
150 reduced the distance to the future population predicted by the LBI and mutational load model
151 (**Figure 2**). Under the all three submission lag scenarios, the distance to the future reduced by ap-
152 proximately 1 AA on average for each 3-month reduction in forecast horizon (**Table 1**). We observed
153 the greatest average reduction in distance to the future (~1.4 AAs) between the 6- and 3-month
154 forecast horizons. Reducing the forecast horizon also noticeably reduced the variance per time-
155 point in predicted future populations across all lag scenarios (**Figure 2**). For example, the standard
156 deviation of distances to the future reduced from ~2.6 AAs at the 12-month horizon to ~1 AA at the
157 3-month horizon (**Table 1**). We observed the same patterns for forecasts of simulated A/H3N2-like
158 populations (**Figure 2—figure Supplement 1**) and optimal distances to the future for natural and
159 simulated populations (**Figure 2—figure Supplement 2** and **Figure 2—figure Supplement 3**). Thus,
160 reducing how far we have to predict into the future increased both forecast accuracy and precision.

Horizon	Distance to future (mean +/- std dev AAs)		
	No lag	Ideal lag	Realistic lag
3	2.91 +/- 0.86	3.32 +/- 0.96	3.85 +/- 1.05
6	4.44 +/- 1.39	4.74 +/- 1.54	5.03 +/- 1.66
9	5.48 +/- 2.05	5.84 +/- 2.14	6.04 +/- 2.15
12	6.45 +/- 2.72	6.77 +/- 2.80	6.78 +/- 2.61

Table 1. Distance to the future in amino acids (mean +/- standard deviation AAs) by forecast horizon (in months) and submission lag for A/H3N2 populations.

161 In contrast, we found that reducing submission lags from a ~3-month average lag in the re-
162 alistic scenario to a ~1-month average lag in the ideal scenario had a weaker effect on distance
163 to the future. At the 12-month forecast horizon, the ideal and realistic lag scenarios produced
164 similar predictions, with the only noticeable improvement observed under the scenario without
165 any submission lags (**Figure 2**). As the forecast horizon decreased, the effect of submission lags
166 appeared more prominent, with the greatest effect of reduced lags observed at the 3-month fore-
167 cast horizon. However, the average improvement from the realistic to the ideal submission lag
168 scenario at the 3-month horizon was still only ~0.3 AAs (**Table 1**). Reducing submission lags also
169 had little effect on the variance per timepoint in predicted future populations. Interestingly, we
170 observed a stronger effect of reducing submission lags in simulated A/H3N2-like populations, with
171 the best average improvement between realistic and ideal lags of ~0.7 AAs at the 3-month horizon
172 (**Figure 2—figure Supplement 1**). As with natural A/H3N2 populations, the effect of reducing sub-
173 mission lags appeared to increase as the forecast horizon decreased. These results indicate that
174 reducing submission lags may have little effect under the current 12-month forecast approach
175 used for influenza vaccine composition, but reducing submission lags should become increasingly
176 important as we forecast from closer to future influenza populations.

177 **Reducing submission lags improves estimates of current clade frequencies**

178 Although the distance between predicted and observed future populations in amino acids provides
179 an unbiased metric to optimize forecasting models, in practice, we use these models to forecast

It is made available under a [CC-BY 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

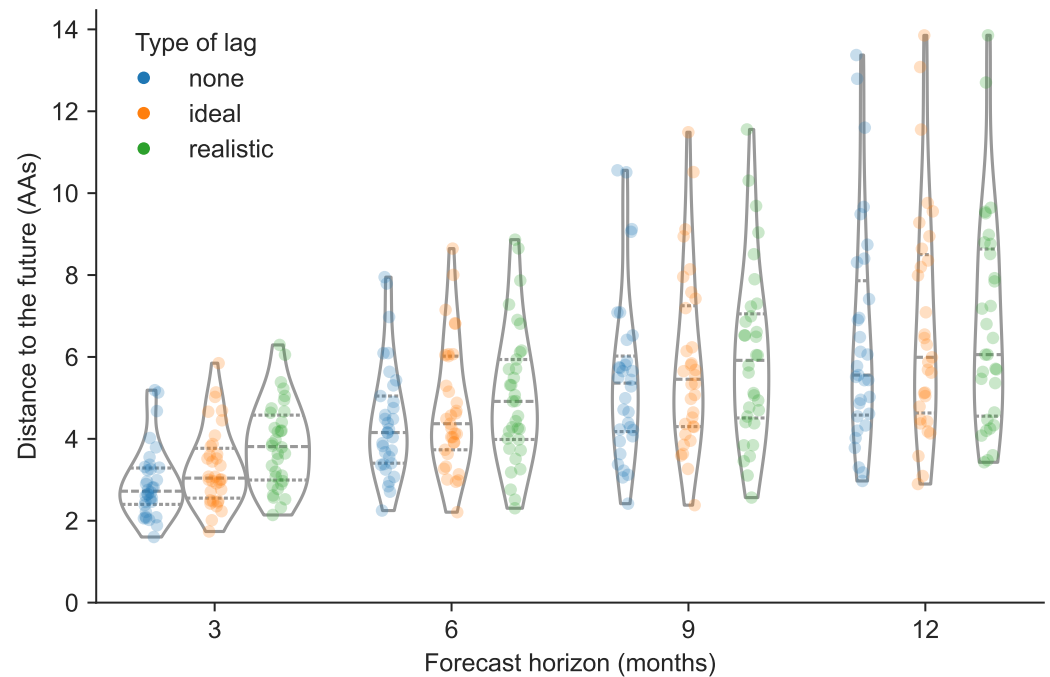


Figure 2. Distance to the future per timepoint (AAs) for natural A/H3N2 populations by forecast horizon and submission lag type based on forecasts from the local branching index (LBI) and mutational load model. Each point represents a future timepoint whose population was predicted from the number of months earlier corresponding to the forecast horizon. Points are colored by submission lag type including forecasts made with no lag (blue), an ideal lag (orange), and a realistic lag (green).

Figure 2—figure supplement 1. Distance to the future for simulated A/H3N2-like populations

Figure 2—figure supplement 2. Optimal distance to the future for natural A/H3N2 populations

Figure 2—figure supplement 3. Optimal distance to the future for simulated A/H3N2-like populations

Figure 2—source data 1. Distances to the future for natural A/H3N2 populations.

Figure 2—source code 1. Jupyter notebook used to produce this figure and the supplemental figure lives in `workflow/notebooks/plot-distances-to-the-future-by-delay-type-and-horizon-for-population.py.ipynb`.

180 clade frequencies. We predict each clade's future frequency as the sum of predicted future frequen-
181 cies for each HA sequence in the clade. We calculate these sequence-specific future frequencies
182 as the initial sequence frequency times the estimated sequence fitness (Łuksza and Lässig, 2014;
183 Huddleston et al., 2020). Given the importance of initial clade frequencies in these forecasts, we
184 tested the effect of submission lags on current clade frequency estimates. For each timepoint and
185 clade with a frequency greater than zero under the scenario without lags, we calculated the clade
186 frequency error as the difference between clade frequency without submission lags and the fre-
187 quency with either an ideal or realistic lag. Positive error values represented underestimation of
188 current clades, while negative values represented overestimation.

189 Across all clade frequencies, we found that errors in current clade frequencies for A/H3N2 ap-
190 peared normally distributed with lower variance in the ideal lag scenario than under realistic lags
191 (Figure 3A and B). Of the 822 clades under the scenario without lags, 613 (75%) had a frequency less
192 than 10%, representing small, emerging clades. The remaining 209 (25%) had a frequency of 10%
193 or greater, representing larger clades that could be more likely to succeed. To understand whether
194 lags had different effects on these small and large clades, respectively, we inspected clades from
195 these latter two groups separately. For small clades, errors under ideal lags ranged from -4% to
196 4% with a standard deviation of 1%, while realistic lags produced errors ranging from -8% to 7%
197 with a standard deviation of 2% (Figure 3C). We did not observe a bias toward underestimation or

It is made available under a [CC-BY 4.0 International license](https://creativecommons.org/licenses/by/4.0/) .

198 overestimation of initial small clade frequencies under either lag scenario. For large clades, errors
199 under ideal lag ranged from -9% to 14% with a standard deviation of 3% (**Figure 3D**). Errors un-
200 der realistic lags ranged from -16% to 29% with a standard deviation of 6%. We observed a slight
201 bias toward underestimation of large clades under the realistic lag scenario, with a median error
202 of 1%. These results show that reducing submission lags for natural A/H3N2 populations from a
203 3-month average to a 1-month average could reduce the bias toward underestimated large clade
204 frequencies and reduce the standard deviation of all current clade frequency errors by 50%.

205 Lagged submissions similarly affected clade frequencies for simulated A/H3N2-like populations
206 (**Figure 3—figure Supplement 1**). Small clade errors under ideal lags ranged from -4% to 6% (stan-
207 dard deviation of 1%) and under realistic lags ranged from -9% to 8% (standard deviation of 2%)
208 (**Figure 3—figure Supplement 1C**). For large clades, errors under ideal lags ranged from -8% to 18%
209 (standard deviation of 3%) and under realistic lags from -14% to 40% (standard deviation of 7%)
210 (**Figure 3—figure Supplement 1D**). As with natural A/H3N2 populations, we observed a slight bias
211 in simulated populations under realistic lags toward underestimation of large clade frequencies
212 with a median error of 2%. We also observed a similar reduction in standard deviation of cur-
213 rent frequency errors for these simulated A/H3N2-like populations when switching from realistic
214 to ideal submission lags.

215 **Reducing forecast horizons increases the accuracy and precision of clade frequency** 216 **forecasts**

217 Next, we estimated the effects of different forecast horizons and submission lags on the accu-
218 racy of clade frequency forecasts. As with the current clade frequency analysis, we analyzed small
219 clades (<10% initial frequency) and large clades ($\geq 10\%$ initial frequency) separately. For each com-
220 bination of initial timepoint, future timepoint, and lag scenario (**Figure 1A**), we calculated initial
221 and predicted future frequencies for all clades present under the given lag and then calculated the
222 corresponding observed future frequencies without lag for clades that descended from the clades
223 present at the initial timepoint. We calculated the error in forecast frequencies as the difference
224 between predicted future frequencies under the given lag scenario and observed future frequen-
225 cies without any lag. We used absolute forecast errors to evaluate forecast accuracy and overall
226 forecast errors to evaluate forecast bias.

227 Absolute forecast errors trended strongly toward values less than 30% with long tails reaching
228 80% for both small and large clades (**Figure 4**). Each 3-month reduction of the forecast horizon lin-
229 early reduced the variance in forecast errors, but mean and median absolute errors only improved
230 after reducing the forecast horizon below 9 months (**Figure 4** and **Table 2**). For small clades, reduc-
231 ing the forecast horizon most noticeably reduced the range of errors, while reducing submission
232 lags had little effect (**Figure 4A**). For large clades, almost all decreases in forecast horizon and sub-
233 mission lag (except lags at the 12-month horizon) reduced the standard deviation of absolute fore-
234 cast errors (**Figure 4B**). Overall, reducing the forecast horizon had a greater effect on the mean,
235 median, and standard deviation of absolute forecast errors than reducing submission lags. For
236 example, the standard deviation of absolute errors at the 12-month horizon under realistic sub-
237 mission lags was 23%, while the standard deviation for the 6-month horizon under realistic lags
238 was 14% (**Table 2**). In contrast, the standard deviation at the 12-month horizon under ideal submis-
239 sion lags did not change from the realistic lags at 23%, and the average absolute error increased by
240 1% from 20%. For all other forecast horizons, reducing the submission lags from realistic to ideal
241 only reduced the mean and standard deviation of absolute errors by 1–2%. We observed the same
242 general patterns in simulated populations (**Figure 4—figure Supplement 1**).

243 The majority of forecast frequency errors appeared to be normally distributed, indicating lit-
244 tle bias toward over- or underestimating future clade frequencies (**Figure 4—figure Supplement 2**
245 and **Figure 4—figure Supplement 3**). This pattern matched our expectation that at any given initial
246 timepoint the overestimation of one clade's future frequency must cause an underestimation of
247 another current clade's future frequency. However, we observed a long tail of small clades with

It is made available under a [CC-BY 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

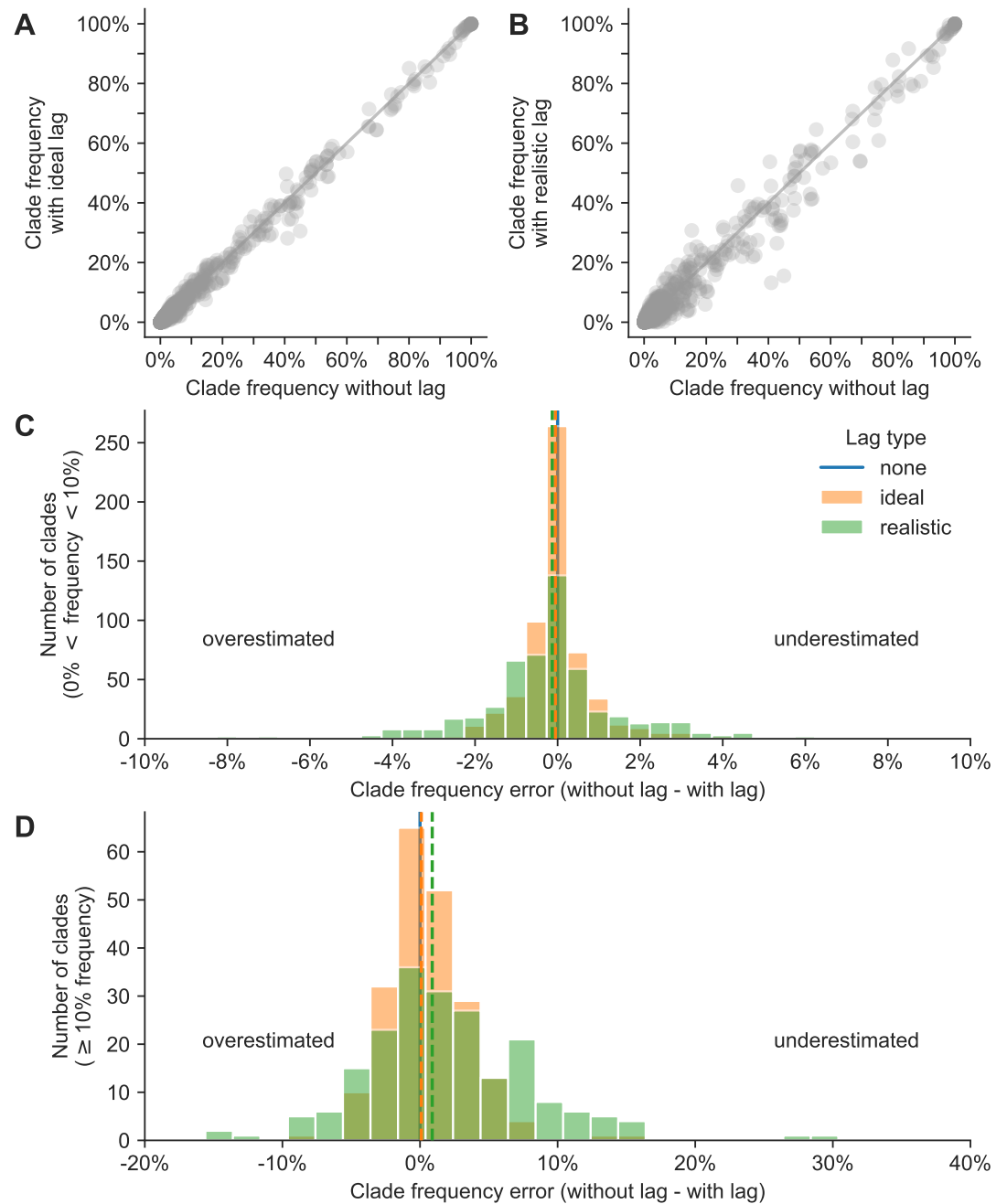


Figure 3. Clade frequency errors for natural A/H3N2 clades at the same timepoint calculated as the difference between clade frequencies without submission lag and corresponding frequencies with either A) ideal or B) realistic submission lags. Distributions of frequency errors appear normally distributed in both lag scenarios for both C) small clades (>0% and <10% frequency) and D) large clades ($\geq 10\%$). Dashed lines indicate the median error from the distribution of the lag type with the same color.

Figure 3—figure supplement 1. Current clade frequency errors for simulated A/H3N2-like populations

Figure 3—source data 1. Current and future clade frequencies for natural A/H3N2 populations by forecast horizon and submission lag type.

Figure 3—source code 1. Jupyter notebook used to produce this figure and the supplemental figure lives in `workflow/notebooks/plot-current-clade-frequency-errors-by-delay-type-for-populations.py.ipynb`.

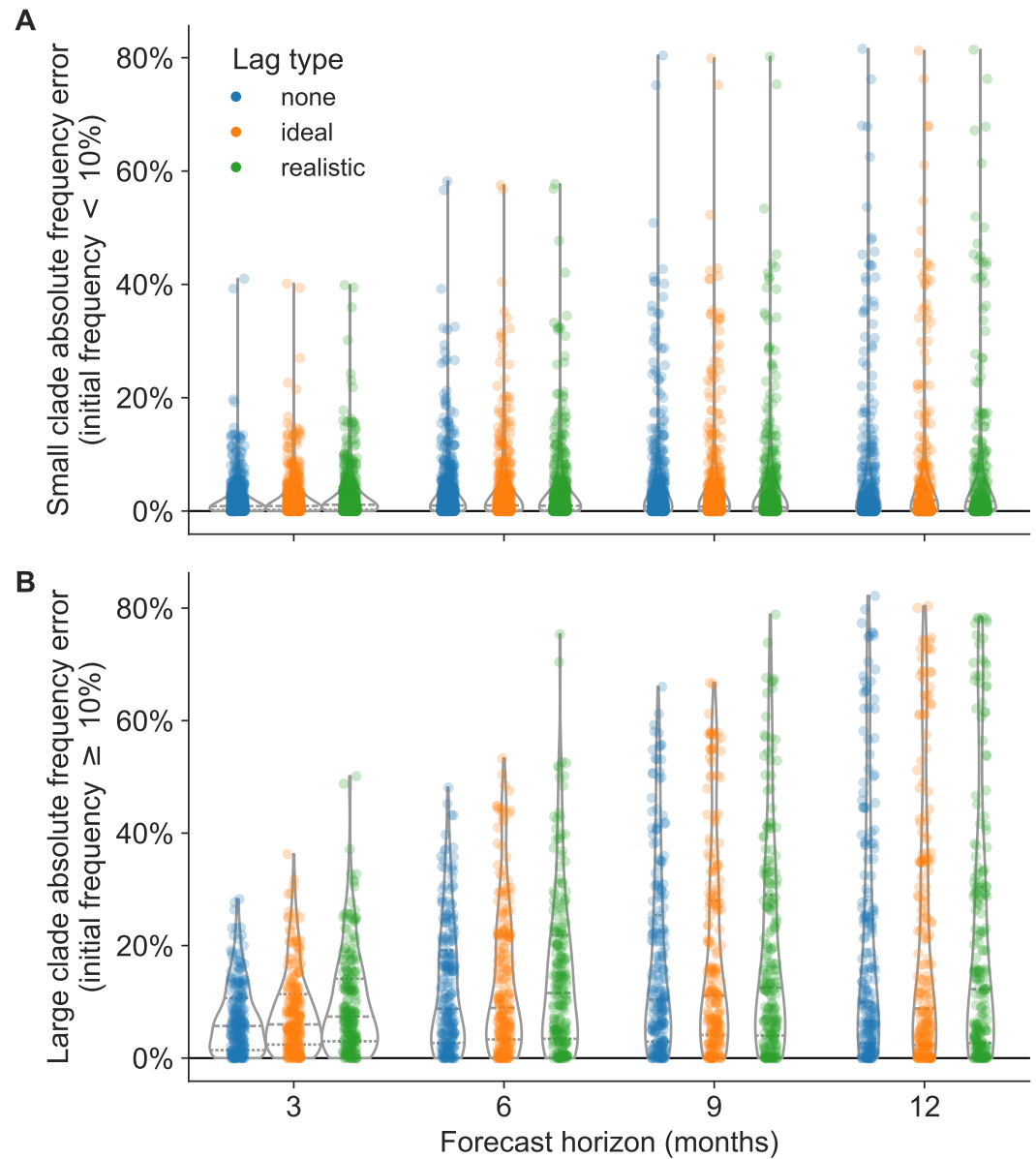


Figure 4. Absolute forecast clade frequency errors for natural A/H3N2 populations by forecast horizon in months and submission lag type (none, ideal, or observed) for A) small clades (<10% initial frequency) and B) large clades (≥10% initial frequency).

Figure 4—figure supplement 1. Absolute forecast clade frequency errors for simulated A/H3N2-like populations.

Figure 4—figure supplement 2. Forecast clade frequency errors for natural A/H3N2 populations.

Figure 4—figure supplement 3. Forecast clade frequency errors for simulated A/H3N2-like populations.

Figure 4—source code 1. Jupyter notebook used to produce this figure and the supplemental figures lives in [workflow/notebooks/plot-forecast-clade-frequency-errors-by-delay-type-and-horizon-for-population.py](#). [ipython](#)

249 growth of small clades remains more difficult than predicting their decline (**Figure 4—figure Sup-**
250 **plement 2A**). The strongest effect of reducing submission lags was the reduction in maximum error,
251 corresponding to reduction in underestimation of large clades. The switch from realistic to ideal
252 lags at 12-, 9-, 6-, and 3-month horizons reduced the maximum forecast error by 4%, 21%, 22%,
253 and 14%, respectively (**Table 2**). These results show that reducing submission lags can substantially

Horizon	Lag type	Clade frequency error (%)					Absolute frequency error (%)		
		Mean	Median	Std Dev	Min	Max	Mean	Median	Std Dev
3	none	1	0	9	-28	28	7	6	6
3	ideal	1	0	11	-32	36	8	6	7
3	realistic	1	0	13	-31	50	10	7	9
6	none	1	0	17	-48	45	12	9	11
6	ideal	1	0	19	-50	53	13	9	13
6	realistic	1	0	20	-52	75	15	12	14
9	none	0	-1	23	-66	59	16	10	17
9	ideal	1	-1	25	-67	58	18	11	18
9	realistic	1	-1	26	-67	79	19	12	19
12	none	0	0	30	-82	76	20	10	22
12	ideal	1	0	31	-80	74	21	9	23
12	realistic	0	0	31	-78	78	20	12	23

Table 2. Errors in clade frequencies between observed and predicted values by forecast horizon (in months) and submission lag for A/H3N2 clades with an initial frequency $\geq 10\%$ under the given lag scenario.

254 lower the upper bound for forecasting errors.

255 **Reduced vaccine development time provides the best improvement in forecast ac-**
 256 **curacy of available realistic interventions**

257 Although we have investigated the effects of a range of forecast horizons and submission lags, not
 258 all of these scenarios are currently realistic. The most we can hope to reduce the forecast horizon
 259 with current mRNA vaccine technology is from 12 months to 6 months and the most we could re-
 260 duce submission lags would be from an average of 3 months to 1 month (*Grant et al., 2023*). In
 261 practice, we wanted to know how much a reduction in forecast horizon or submission lag could
 262 improve the accuracy of forecasts to each future timepoint. To determine the effects of realistic
 263 interventions on forecast accuracy, we inspected the reduction in total absolute forecast error per
 264 future timepoint associated with improved vaccine development (reducing forecast horizon from
 265 12 months to 6 months), improved genomic surveillance (reducing lags from a 3-month average to
 266 1 month), and the combination of both improvements. We selected all forecasts with a 12-month
 267 horizon and a realistic lag, to represent current forecast conditions or “the status quo”. For the
 268 same future timepoints present in the status quo conditions, we selected the corresponding fore-
 269 casts for a 6-month horizon and a realistic lag, a 12-month horizon and an ideal lag, and 6-month
 270 horizon and an ideal lag. Since forecasts between different initial and future timepoints could be
 271 represented by different clades, we could not compare forecasts for specific clades between in-
 272 terventions. Instead, we calculated the total absolute clade frequency error per future timepoint
 273 under each intervention and calculated the improvement in forecast accuracy as the difference in
 274 total error between the status quo and each intervention. In addition to this clade-based analy-
 275 sis, we also estimated effects of interventions on the difference in distance to the future between
 276 different scenarios for both estimated and empirical fitnesses. For all analyses, positive values
 277 represented improved forecast accuracy under a given intervention scenario and negative values
 278 represented a reduction in accuracy.

279 Both interventions with improved vaccine development increased forecast accuracy for the ma-
 280 jority of future timepoints (*Figure 5, Table 3, and Figure 5—figure Supplement 1*). Improving vac-
 281 cine development alone increased total forecast accuracy by 53% on average, while the addition
 282 of improved genomic surveillance under that 6-month forecast horizon increased total forecast ac-
 283 curacy by 54% on average. In contrast, the intervention that only improved genomic surveillance

It is made available under a [CC-BY 4.0 International license](https://creativecommons.org/licenses/by/4.0/) .

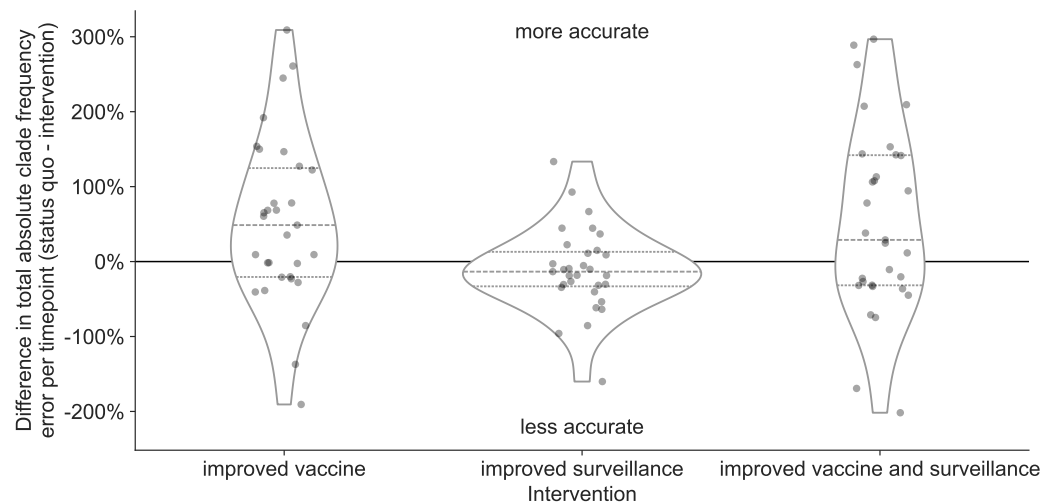


Figure 5. Improvement of clade frequency errors for A/H3N2 populations between the status quo (12-month forecast horizon and realistic submission lags) and realistic interventions of improved vaccine development (reducing 12-month to 6-month forecast horizon), improved surveillance (reducing submission lags from 3 months on average to 1 month), or a combination of both interventions. We measured improvements from the status quo as the difference in total absolute clade frequency error per future timepoint. Positive values indicate increased forecast accuracy, while negative values indicate decreased accuracy. Each point represents the improvement of forecasts for a specific future timepoint under the given intervention. Horizontal dashed lines indicate median improvements. Horizontal dotted lines indicate upper and lower quartiles of improvements.

Figure 5—figure supplement 1. Distribution of total absolute clade frequency errors summed across clades per future timepoint for A/H3N2 populations.

Figure 5—figure supplement 2. Improvement of clade frequency errors for simulated A/H3N2-like populations.

Figure 5—figure supplement 3. Distribution of total absolute clade frequency errors summed across clades per future timepoint for simulated A/H3N2-like populations.

Figure 5—figure supplement 4. Improvement of distances to the future (AAs) for A/H3N2 populations between the status quo (12-month forecast horizon and realistic submission lags) and realistic interventions.

Figure 5—figure supplement 5. Improvement of distances to the future (AAs) for simulated A/H3N2-like populations between the status quo (12-month forecast horizon and realistic submission lags) and realistic interventions.

Figure 5—source data 1. Differences in total absolute clade frequency error per future timepoint and clade between the status quo and realistic interventions for A/H3N2 populations.

Figure 5—source code 1. Jupyter notebook used to produce effects of interventions on total absolute clade frequency errors

`workflow/notebooks/plot-forecast-clade-frequency-errors-by-delay-type-and-horizon-for-population.py.ipynb`

Figure 5—source code 2. Jupyter notebook used to produce effects of interventions on distances to the future lives in

`workflow/notebooks/plot-distances-to-the-future-by-delay-type-and-horizon-for-population.py.ipynb`

284 decreased forecast accuracy by an average of 11%. Based on the distributions of total absolute
285 forecast error per future timepoint, we would expect improved genomic surveillance to improve
286 forecast accuracy at a forecast horizon of 3 months (**Figure 5—figure Supplement 1**). We observed
287 similar effects of interventions in simulated A/H3N2-like populations except that the average ef-
288 fect of reducing submission lags alone was positive for these populations (**Figure 5—figure Sup-
289 plement 2** and **Figure 5—figure Supplement 3**). When we calculated the effects of interventions on
290 distances to the future instead of total absolute clade frequency errors, we observed the same pat-

Intervention	Forecast accuracy improvement (%)			Timepoints improved	
	Mean	Median	Std Dev	Total	Proportion
improved vaccine	53	49	112	19	0.61
improved surveillance	-11	-13	56	10	0.32
improved vaccine and surveillance	54	29	124	18	0.58

Table 3. Improvement in A/H3N2 clade frequency forecast accuracy under realistic interventions of improved vaccine development (reducing 12-month to 6-month forecast horizon), improved surveillance (reducing submission lags from 3 months on average to 1 month), or a combination of both interventions. We measured improvements from the status quo (12-month forecast horizon and 3-month average submission lag) as the difference in total absolute clade frequency error per future timepoint and the number and proportion of future timepoints for which forecasts improved under the intervention.

291 terns for natural and simulated populations (*Figure 5—figure Supplement 4* and *Figure 5—figure*
292 *Supplement 5*). Based on these results, the single most valuable intervention we could make to im-
293 prove forecast accuracy would be to reduce the forecast horizon to 6 months or less through more
294 rapid vaccine development. However, as we reduce the forecast horizon, reducing submission lags
295 should have a greater effect on improving forecast accuracy.

296 We hypothesized that the decrease in average accuracy of natural A/H3N2 forecasts under the
297 improved genomic surveillance intervention could reflect the bias of the LBI and mutational load
298 fitness metrics. For example, we previously showed how LBI fitness estimates can overestimate
299 the future growth of large clades (*Huddleston et al., 2020*). Adding more sequences at initial time-
300 points where LBI already overestimates clade success could increase the LBI of those clades and
301 exacerbate the overestimation. To test this hypothesis, we calculated the effects of the same inter-
302 ventions on the optimal distances to the future for both natural and simulated populations. Since
303 optimal distances reflected the empirical fitnesses of the initial populations, the effects of inter-
304 ventions should be independent of biases from fitness metrics. We expected all interventions to
305 maintain or improve the optimal distance to the future without any cases where an intervention
306 decreased accuracy. As expected, all interventions improved on the optimal distance to the future
307 for both populations (*Figure 6* and *Figure 6—figure Supplement 1*). For natural A/H3N2 popula-
308 tions, the average improvement of the vaccine intervention was 1.1 AAs and the improvement of
309 the surveillance intervention was 0.27 AAs or approximately 25% of the vaccine intervention. The
310 average improvement of both interventions was only slightly less than additive at 1.28 AAs. These
311 results confirmed the relatively stronger effect of reducing forecast horizons compared to submis-
312 sion lags. They also confirmed that reducing submission lags can improve forecasts under optimal
313 forecasting conditions. For this reason, we expect that simultaneous improvements to forecasting
314 models and genomic surveillance will have a mutually beneficial effect on forecast accuracy.

315 Discussion

316 In this work, we showed that decreasing the time to develop new vaccines for seasonal influenza
317 A/H3N2 and decreasing submission lags of HA sequences to public databases improves our esti-
318 mates of future and current populations, respectively. We confirmed that forecasts became more
319 accurate and more precise with each 3-month reduction in forecast horizon from the status quo
320 of 12 months. Although decreasing submission lags only marginally improved long-term forecast
321 accuracy, shorter lags increased the accuracy of current clade frequency estimates, reduced the
322 bias toward underestimating current and future frequencies of larger clades, and improved fore-
323 casts 3 months into the future. Under a realistic scenario where a faster vaccine development
324 timeline allowed us to forecast from 6 months before the next season, we found a 53% average
325 improvement in forecasts of total absolute clade frequency and a 25% reduction in average ab-
326 solute forecast frequency errors for large clades from 20% to 15%. We confirmed these effects

It is made available under a [CC-BY 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

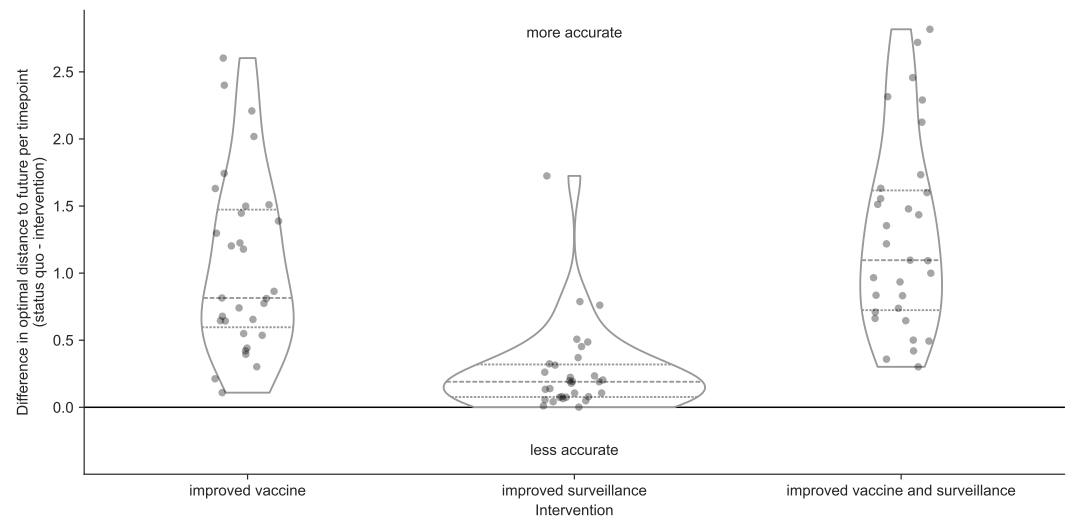


Figure 6. Improvement of optimal distances to the future (AAs) for A/H3N2 populations between the status quo (12-month forecast horizon and realistic submission lags) and realistic interventions of improved vaccine development (reducing 12-month to 6-month forecast horizon), improved surveillance (reducing submission lags from 3 months on average to 1 month), or a combination of both interventions. We measured improvements from the status quo as the difference in optimal distances to the future per future timepoint. Positive values indicate increased forecast accuracy, while negative values indicate decreased accuracy. Each point represents the improvement of forecasts for a specific future timepoint under the given intervention. Horizontal dashed lines indicate median improvements. Horizontal dotted lines indicate upper and lower quartiles of improvements.

Figure 6—figure supplement 1. Improvement of optimal distances to the future (AAs) for simulated A/H3N2-like populations between the status quo (12-month forecast horizon and realistic submission lags) and realistic interventions.

Figure 6—source data 1. Differences in optimal distances to the future per future timepoint between the status quo and realistic interventions for A/H3N2 populations.

Figure 6—source code 1. Jupyter notebook used to produce optimal effects of interventions on distances to the future lives in [workflow/notebooks/plot-distances-to-the-future-by-delay-type-and-horizon-for-population.py.ipynb](#).

327 with a previously validated forecasting model using both simulated and natural populations and
328 two different metrics of forecast accuracy including earth mover's distances between populations
329 and clade frequencies. We expect that decreasing forecast horizons and submission lags will have
330 similar relative effect sizes in other forecasting models, too.

331 Even without these recommended improvements to vaccine development and sequence sub-
332 missions, these results inform important next steps to improve forecasting models. Current and
333 future frequency estimates should be presented with corresponding uncertainty intervals. From
334 this work, we know that our current frequency estimates for large clades ($\geq 10\%$ frequency) un-
335 der realistic submission lags have a wide range of errors (-16% to 29%). Similarly, the range of
336 12-month forecast frequency errors under realistic lags include overestimates by up to 78% and
337 underestimates up to 78%. Long-term forecasts with incomplete current data are highly uncertain
338 by their nature. To support informed decisions about vaccine updates, we must communicate that
339 uncertainty of the present and future to decision-makers. One simple immediate strategy to pro-
340 vide these uncertainty estimates is to estimate current and future clade frequencies from count
341 data with multinomial probability distributions. Another immediate improvement would be to de-
342 velop models that can use all available data in a way that properly accounts for geographic and
343 temporal biases. Current models based on phylogenetic trees need to evenly sample the diversity
344 of currently circulating viruses to produce unbiased trees in a reasonable amount of time. Models

It is made available under a [CC-BY 4.0 International license](https://creativecommons.org/licenses/by/4.0/) .

345 that could estimate sample fitness and compare predicted and future populations without trees
346 could use more available sequence data and reduce the uncertainty in current and future clade
347 frequencies. Finally, we could improve existing models by changing the start and end times of our
348 long-term forecasts. We could change our forecasting target from the middle of the next season
349 to the beginning of the season, reducing the forecast horizon from 12 to 9 months. We could also
350 start forecasting from one month prior to the current date to minimize the effect of submission
351 lags on our estimates of the current global influenza population.

352 Despite the small effect that reducing sequence submission lags had on long-term forecasting
353 accuracy, we still see a need to continue funding global genomic surveillance at higher levels than
354 the pre-pandemic period. Compared to estimates of current viral diversity, forecasts of future
355 influenza populations only represent one component of the overall decision-making process for
356 vaccine development. For example, virologists must choose potential vaccine candidates from the
357 diversity of circulating clades well in advance of vaccine composition meetings to have time to
358 grow virus in cells and eggs and measure antigenic drift with serological assays (*Morris et al., 2018*;
359 *Loes et al., 2024*). Similarly, prospective measurements of antigenic escape from human sera allow
360 researchers to predict substitutions that could escape global immunity (*Lee et al., 2019*; *Greaney*
361 *et al., 2022*; *Welsh et al., 2023*). The finding of even a few sequences with a potentially important
362 antigenic substitution could be enough to inform choices of vaccine candidate viruses. Finally, our
363 results here reflect uncorrelated submission lags for each sequence, but actual lags can strongly
364 correlate between sequences from the same originating and submitting labs. These correlated
365 lags could further decrease the accuracy of frequency estimates beyond our more conservative
366 estimates. More rapid sequence submission will improve our understanding of the present and
367 give decision-makers more choices for new vaccines. Such reductions in submission lags depend
368 on substantial, sustained funding and capacity building globally.

369 **Methods and Materials**

370 **Selection of natural influenza A/H3N2 HA sequences**

371 We downloaded all A/H3N2 HA sequences and metadata from GISAID's EpiFlu database (*Shu and*
372 *McCauley, 2017*) as of November 2023. We evenly sampled sequences geographically and tem-
373 porally as previously described (*Huddleston et al., 2020*). Briefly, we selected 90 sequences per
374 month, evenly sampling from major continental regions (Africa, Europe, North America, China,
375 South Asia, Japan and Korea, Oceania, South America, Southeast Asia, and West Asia) and exclud-
376 ing sequences labeled as egg-passaged or missing complete date annotations. For our forecasting
377 analyses, we selected sequences collected between April 1, 2005 and October 1, 2019.

378 **Simulation of influenza A/H3N2-like HA sequences**

379 We simulated A/H3N2-like populations as previously described (*Huddleston et al., 2020*). Briefly,
380 we simulated A/H3N2 HA sequences with SANTA-SIM (*Jariani et al., 2019*) for 10,000 generations
381 or 50 years at 200 generations per year. We discarded the first 10 years of simulated data as a
382 burn-in period and used the next 30 years of the remaining data for our analyses. We sampled 90
383 viruses per month to match the sampling density of natural populations.

384 **Estimating and assigning submission lags**

385 We estimated the lag between sample collection and submission of A/H3N2 hemagglutinin (HA)
386 sequences to the GISAID EpiFlu database (*Shu and McCauley, 2017*) by calculating the difference in
387 GISAID-annotated submission date and collection date in days for samples collected between Jan-
388 uary 1, 2019 and January 1, 2020 and with a submission date prior to October 1, 2020. We selected
389 this period of time as representative of modern genomic surveillance efforts prior to changes in cir-
390 culation patterns of influenza caused by the SARS-CoV-2 pandemic. Of the 104,392 HA sequences
391 in GISAID EpiFlu, 11,222 (11%) were collected during this period with a mean submission lag of 98

It is made available under a [CC-BY 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

392 days (~3 months) and a median lag of 74 days. Only 11% of sequences (N=1,210) were submitted
393 within 4 weeks of collection, and only 36% (N=4,057) were submitted within 8 weeks (*Figure 1A*,
394 purple).

395 We modeled the shape of the observed lag distribution as a gamma distribution using a max-
396 imum likelihood fit from SciPy 1.10.1 (*Virtanen et al., 2020*). With this approach, we estimated a
397 shape parameter of 1.76, a scale parameter of 53.18, and location parameter of 3.98. The product
398 of these shape and scale values corresponded to a mean lag of 93.76 days (*Figure 1A*, green). To
399 assign realistic submission lags to each sample in our analysis, we randomly sampled from this
400 gamma distribution and calculated a “realistic submission date” by adding the sampled lag in days
401 to the observed collection date. This approach allowed us to assign realistic lags to natural and sim-
402 ulated populations without the biases and autocorrelations associated with historical submission
403 patterns across different submitting labs.

404 Based on the observed rapid submission of SARS-CoV-2 genomes during the first years of the
405 pandemic, we expected that an achievable “ideal” submission lag for seasonal influenza sequences
406 would have a 1-month average lag instead of the observed ~3-month lag from the pre-pandemic
407 period. We modeled this ideal submission lag distribution by dividing the gamma shape parameter
408 by 3 to get a value of 0.59 and a corresponding mean lag of 31.25 days (*Figure 1A*, orange). This ap-
409 proach effectively shifted the realistic gamma toward zero, while maintaining the relatively longer
410 upper tail of the distribution. To assign ideal submission lags to each sample in our analysis, we
411 randomly sampled from this modified gamma distribution and added the sampled lag in days to
412 the observed collection date. Additionally, we required that each sample’s “ideal” lag be less than
413 or equal to its “realistic” lag.

414 To estimate the effect of increased global sequencing capacity associated with the response to
415 the SARS-CoV-2 pandemic, we summarized the lag distribution for sequences submitted to GISAID
416 EpiFlu between January 1, 2022 and January 1, 2023. During this period, global influenza circulation
417 had rebounded to its prepandemic level and 26,394 HA sequences were collected. The mean and
418 median submission lags during this period were 76 and 62 days, respectively, representing a trend
419 toward reduced lags compared to the prepandemic era (*Figure 1—figure Supplement 1*).

420 **Phylogenetic inference**

421 We inferred time-scaled phylogenetic trees for HA sequences as previously described (*Huddleston*
422 *et al., 2020*). Briefly, we aligned sequences with MAFFT v7.520 (*Katoh et al., 2002; Katoh and Stand-*
423 *ley, 2013*) using the `augur align` command in Augur v22.3.0 (*Huddleston et al., 2021*). We inferred
424 phylogenies with IQ-TREE v2.2.3 (*Nguyen et al., 2014*) using the `augur tree` command with IQ-TREE
425 parameters of `-ninit 2 -n 2 -me 0.05` and a general time reversible (GTR) model. We inferred
426 time-resolved phylogenies with TreeTime v0.10.1 (*Sagulenko et al., 2018*) with the `augur refine`
427 command.

428 **Forecasting with different forecast horizons**

429 We tested the effect of forecasting future influenza populations at forecast horizons of 3, 6, 9, and
430 12 months (*Figure 1B*). Previously, we produced forecasts every 6 months starting from October
431 1 and April 1 and predicting 12 months into the future (*Huddleston et al., 2020*). To support fore-
432 casts in 3-month intervals, we produced annotated time trees for 6 years of HA sequences every
433 3 months with data available up to the first day of January, April, July, and October. We produced
434 these trees for each timepoint with three different lag scenarios: no lag, ideal lag, and realistic lag.
435 For each scenario, we selected sequences for analysis at a given timepoint based on their collection
436 date, ideal submission date, or realistic submission date, respectively. This experimental design
437 produced forecasts for three lag types at each of the four forecast horizons (e.g., *Figure 1B*, blue,
438 green, and orange initial timepoints for the 3-month forecast horizon).

439 Since reliable submission dates were not available prior to April 2005, our analysis of natural
440 A/H3N2 sequences spanned from April 1, 2005 to October 1, 2019. To simplify the data required for

It is made available under a [CC-BY 4.0 International license](https://creativecommons.org/licenses/by/4.0/) .

441 these analyses, we produced forecasts of natural A/H3N2 populations with our best sequence-only
442 model from our prior work (*Huddleston et al., 2020*), a composite model based on local branch-
443 ing index (LBI) (*Neher et al., 2014*) and mutational load (*Łuksza and Lässig, 2014*). For simulated
444 A/H3N2-like populations, we produced forecasts with the “true fitness” model that relies on the
445 normalized fitness value of each simulated sample.

446 Each forecast generated a predicted future frequency per sequence in the initial timepoint's
447 tree. As in our prior work, we calculated the earth mover's distance (*Rubner et al., 1998*) between
448 the predicted and observed future populations using HA amino acid sequences from initial and
449 future timepoints, predicted future frequencies from the initial timepoint, and observed future fre-
450 quencies from future timepoint. For the future timepoint, we used data from the “no lag” scenario
451 as our truth set, regardless of the lag scenario for the initial timepoint. This design allowed us to
452 measure the effect of ideal and realistic submission lags on forecast accuracy relative to a scenario
453 with no lags.

454 **Defining clades**

455 Official clade definitions do not exist for all time periods of our analysis of A/H3N2 populations and
456 do not exist at all for simulated A/H3N2-like populations. Therefore, we defined clades *de novo* for
457 both population types with the same clade assignment algorithm used to produce “subclades”
458 for recent seasonal influenza vaccine composition meeting reports (*Huddleston et al., 2024*). The
459 complete algorithm description and implementation is available at [https://github.com/neherlab/](https://github.com/neherlab/flu_clades)
460 [flu_clades](https://github.com/neherlab/flu_clades). Briefly, the algorithm scores each node in a phylogenetic tree based on three criteria
461 including the number of child nodes descending from the current node, the number of epitope
462 substitutions on the branch leading to the current node, and the number of amino acid mutations
463 since the last clade assigned to an ancestor in the tree. After assigning and normalizing scores, the
464 algorithm traverses the tree in preorder, assigning clade labels to each internal node whose score
465 exceeds a predefined threshold of 1.0. Clade labels follow a hierarchical nomenclature inspired
466 by Pangolin (*O'Toole et al., 2021*) such that the first clade in the tree is named “A” and its first
467 immediate descendant is named “A.1”. For each population type, we applied this algorithm to a
468 single phylogeny representing all HA sequences present in our analysis. This approach allowed us
469 to produce a single clade assignment per sequence and easily identify related sequences between
470 initial and future timepoints using the hierarchical clade nomenclature.

471 **Estimating current and future clade frequencies**

472 We estimated clade frequencies with a kernel density estimation (KDE) approach as previously de-
473 scribed (*Huddleston et al., 2020*) with the `augur frequencies` command (*Huddleston et al., 2021*).
474 Briefly, we represented each sequence in a given phylogeny by a Gaussian kernel with a mean
475 at the sequence's collection date and a variance of two months. We estimated the frequency of
476 each sequence at each timepoint by calculating the probability density function of each KDE at that
477 timepoint and normalizing the resulting values to sum to one.

478 We calculated clade frequencies for each initial timepoint in our analysis by first summing the
479 frequencies of individual sequences in a given timepoint's tree by the clade assigned to each se-
480 quence and then summing the frequencies for each clade and its descendants to obtain nested
481 clade frequencies. To inspect the effects of submission lags on clade frequency estimates, we
482 calculated the clade frequency error per timepoint and clade by subtracting the clade frequency
483 estimated with ideal or realistic lagged sequence submission from the corresponding clade fre-
484 quency without lags. We compared the effects of submission lags for clades of different sizes by
485 filtering clades by their frequency estimated without lags to small clades (>0% and <10%) and large
486 clades ($\geq 10\%$).

487 To estimate the accuracy of clade frequency forecasts, we needed to calculate the predicted
488 and observed future clade frequencies for each combination of lag type, initial timepoint, and
489 future timepoint in the analysis. We calculated predicted future frequencies for all clades that

It is made available under a [CC-BY 4.0 International license](https://creativecommons.org/licenses/by/4.0/) .

490 existed at given initial timepoint and lag type by first summing the predicted future frequency per
491 sequence by the clade assigned to each sequence and then summing the predicted frequencies for
492 each clade and its descendants. Clades that existed at any given future timepoint were not always
493 represented at a corresponding initial timepoint either because the clades had not emerged yet
494 or sequences for those clades had a lagged submission. For this reason, we calculated observed
495 future clade frequencies in a multi-step process. First, we calculated the frequencies of clades
496 observed at the future timepoint without submission lag by summing the individual frequencies of
497 all sequences in each clade. Then, we mapped each future clade to its most derived ancestral clade
498 that circulated at the initial timepoint by progressively removing suffixes from the future clade's
499 label until we found a match in the initial timepoint. For example, if the future timepoint had a
500 clade named A.1.1.3 and the initial timepoint had the ancestral clade A.1, we would test for the
501 presence of A.1.1.3, A.1.1, and A.1 at the initial timepoint until we found a match. The hierarchical
502 nature of the clade assignment algorithm guaranteed that each future clade mapped directly to a
503 clade at each initial timepoint and lag type. Finally, we summed the frequencies of future clades
504 by their corresponding initial clades to get the observed future frequencies of clades circulating
505 at the initial timepoint. We calculated the accuracy of clade frequency forecasts as the difference
506 between the predicted and observed future clade frequencies.

507 **Data and software availability**

508 Sequence data are available from the GISAID EpiFlu Database using accessions provided in Supple-
509 mental File S1. Source code for the analysis workflow and manuscript are available in the project's
510 GitHub repository (<https://github.com/blab/flu-forecasting-delays>). Supplemental data are available
511 on Zenodo at DOI [10.5281/zenodo.13742375](https://doi.org/10.5281/zenodo.13742375).

512 **Acknowledgments**

513 We gratefully acknowledge the authors, originating and submitting laboratories of the sequences
514 from the GISAID EpiFlu Database (*Shu and McCauley, 2017*) on which this research is based. A list
515 of sequence accessions, authors, and labs appear in the Supplemental Material. We thank Katie
516 Kistler and Marlin Figgins for their comments on early versions of this manuscript and Richard A.
517 Neher for the development of tools for hierarchical clade nomenclature. This work was funded by
518 NIAID R01 AI165821-01. TB is a Howard Hughes Medical Institute Investigator.

519 **Author contributions**

520 JH designed and implemented experiments, analyzed results, and wrote the manuscript. TB edited
521 the manuscript.

522 **Competing interests**

523 The authors declare that no competing interests exist.

524 **Supplemental Files**

525 **Supplemental File S1.** GISAID accessions and metadata including originating and submitting labs
526 for natural strains used across all timepoints.

527 **References**

- 528 **Abousamra E**, Figgins M, Bedford T. Fitness models provide accurate short-term forecasts of SARS-CoV-2
529 variant frequency. *PLOS Computational Biology*. 2024 09; 20(9):1–20. <https://doi.org/10.1371/journal.pcbi.1012443>, doi: [10.1371/journal.pcbi.1012443](https://doi.org/10.1371/journal.pcbi.1012443).
- 531 **Baden LR**, El Sahly HM, Essink B, Kotloff K, Frey S, Novak R, Diemert D, Spector SA, Roupheal N, Creech CB,
532 McGettigan J, Khetan S, Segall N, Solis J, Brosz A, Fierro C, Schwartz H, Neuzil K, Corey L, Gilbert P, et al.
533 Efficacy and Safety of the mRNA-1273 SARS-CoV-2 Vaccine. *N Engl J Med*. 2021 Feb; 384(5):403–416.

It is made available under a [CC-BY 4.0 International license](https://creativecommons.org/licenses/by/4.0/) .

- 534 **Black A**, MacCannell DR, Sibley TR, Bedford T. Ten recommendations for supporting open pathogen genomic
535 analysis in public health. *Nat Med*. 2020 Jun; 26(6):832–841.
- 536 **Brazzoli M**, Magini D, Bonci A, Buccato S, Giovani C, Kratzer R, Zurli V, Mangiavacchi S, Casini D, Brito LM, De Gre-
537 gorio E, Mason PW, Ulmer JB, Geall AJ, Bertholet S. Induction of Broad-Based Immunity and Protective Efficacy
538 by Self-amplifying mRNA Vaccines Encoding Influenza Virus Hemagglutinin. *J Virol*. 2016 Jan; 90(1):332–344.
- 539 **Brito AF**, Semenova E, Dudas G, Hassler GW, Kalinich CC, Kraemer MUG, Ho J, Tegally H, Githinji G, Agoti CN,
540 Matkin LE, Whittaker C, Howden BP, Sintchenko V, Zuckerman NS, Mor O, Blankenship HM, de Oliveira T, Lin
541 RTP, Siqueira MM, et al. Global disparities in SARS-CoV-2 genomic surveillance. *Nat Commun*. 2022 Nov;
542 13(1):7003.
- 543 **Chen Z**, Azman AS, Chen X, Zou J, Tian Y, Sun R, Xu X, Wu Y, Lu W, Ge S, Zhao Z, Yang J, Leung DT, Domman
544 DB, Yu H. Global landscape of SARS-CoV-2 genomic surveillance and data sharing. *Nat Genet*. 2022 Apr;
545 54(4):499–507.
- 546 **Feldman RA**, Fuhr R, Smolenov I, Mick Ribeiro A, Panther L, Watson M, Senn JJ, Smith M, Almarsson Ö, Pujar
547 HS, Laska ME, Thompson J, Zaks T, Ciaramella G. mRNA vaccines against H10N8 and H7N9 influenza viruses
548 of pandemic potential are immunogenic and well tolerated in healthy adults in phase 1 randomized clinical
549 trials. *Vaccine*. 2019 May; 37(25):3326–3334.
- 550 **Grant R**, Sacks JA, Abraham P, Chunsuttiwat S, Cohen C, Figueroa JP, Fleming T, Fine P, Goldblatt D, Hasegawa H,
551 MacIntyre CR, Memish ZA, Miller E, Nishioka S, Sall AA, Sow S, Tomori O, Wang Y, Van Kerkhove MD, Wambo
552 MA, et al. When to update COVID-19 vaccine composition. *Nat Med*. 2023 Apr; 29(4):776–780.
- 553 **Greaney AJ**, Starr TN, Bloom JD. An antibody-escape estimator for mutations to the SARS-CoV-2 receptor-
554 binding domain. *Virus Evol*. 2022; 8(1):veac021.
- 555 **Hay AJ**, McCauley JW. The WHO global influenza surveillance and response system (GISRS)-A future perspective.
556 *Influenza Other Respir Viruses*. 2018 Sep; 12(5):551–557.
- 557 **Huddleston J**, Barnes JR, Rowe T, Xu X, Kondor R, Wentworth DE, Whittaker L, Ermetal B, Daniels RS, McCauley
558 JW, Fujisaki S, Nakamura K, Kishida N, Watanabe S, Hasegawa H, Barr I, Subbarao K, Barrat-Charlaix P, Neher
559 RA, Bedford T. Integrating genotypes and phenotypes improves long-term forecasts of seasonal influenza
560 A/H3N2 evolution. *eLife*. 2020 Sep; 9:e60067. <https://doi.org/10.7554/eLife.60067>, doi: 10.7554/eLife.60067.
- 561 **Huddleston J**, Bedford T, Chang J, Lee J, Neher RA. Seasonal influenza circulation patterns and projections
562 for February 2024 to February 2025. *Zenodo*. 2024 Mar; <https://doi.org/10.5281/zenodo.10846007>, doi:
563 [10.5281/zenodo.10846007](https://doi.org/10.5281/zenodo.10846007).
- 564 **Huddleston J**, Hadfield J, Sibley TR, Lee J, Fay K, Ilcisin M, Harkins E, Bedford T, Neher RA, Hodcroft EB. Augur:
565 a bioinformatics toolkit for phylogenetic analyses of human pathogens. *J Open Source Softw*. 2021; 6(57).
- 566 **Jariani A**, Warth C, Deforche K, Libin P, Drummond AJ, Rambaut A, Matsen IV FA, Theys K. SANTA-SIM: simulating
567 viral sequence evolution dynamics under selection and recombination. *Virus Evolution*. 2019 Mar; 5(1).
- 568 **Kalia K**, Saberwal G, Sharma G. The lag in SARS-CoV-2 genome submissions to GISAID. *Nat Biotechnol*. 2021
569 Sep; 39(9):1058–1060.
- 570 **Katoh K**, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance
571 and usability. *Mol Biol Evol*. 2013 Apr; 30(4):772–780.
- 572 **Katoh K**, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based
573 on fast Fourier transform. *Nucleic Acids Research*. 2002 07; 30(14):3059–3066. [https://doi.org/10.1093/nar/
574 gkf436](https://doi.org/10.1093/nar/gkf436), doi: 10.1093/nar/gkf436.
- 575 **Kistler KE**, Bedford T. An atlas of continuous adaptive evolution in endemic human viruses. *Cell Host Microbe*.
576 2023; 31(11):1898–1909.
- 577 **Lässig M**, Mustonen V, Walczak AM. Predicting evolution. *Nat Ecol Evol*. 2017 Feb; 1(3):77.
- 578 **Lee JM**, Eguia R, Zost SJ, Choudhary S, Wilson PC, Bedford T, Stevens-Ayers T, Boeckh M, Hurt AC, Lakdawala SS,
579 Hensley SE, Bloom JD. Mapping person-to-person variation in viral mutations that escape polyclonal serum
580 targeting influenza hemagglutinin. *Elife*. 2019 Aug; 8.

It is made available under a [CC-BY 4.0 International license](https://creativecommons.org/licenses/by/4.0/) .

- 581 **Loes AN**, Tarabi RAL, Huddleston J, Touyon L, Wong SS, Cheng SMS, Leung NHL, Hannon WW, Bedford T, Cobey
582 S, Cowling BJ, Bloom JD. High-throughput sequencing-based neutralization assay reveals how repeated vac-
583 cinations impact titers to recent human H1N1 influenza strains. *bioRxiv*. 2024 Mar; .
- 584 **Łuksza M**, Lässig M. A predictive fitness model for influenza. *Nature*. 2014 Mar; 507(7490):57–61.
- 585 **Meijers M**, Ruchnewitz D, Eberhardt J, Łuksza M, Lässig M. Population immunity predicts evolutionary trajec-
586 tories of SARS-CoV-2. *Cell*. 2023 Nov; 186(23):5151–5164.
- 587 **Meijers M**, Ruchnewitz D, Eberhardt J, Karmakar M, Łuksza M, Lässig M. Concepts and methods for predict-
588 ing viral evolution. *bioRxiv*. 2024; <https://www.biorxiv.org/content/early/2024/05/02/2024.03.19.585703>, doi:
589 [10.1101/2024.03.19.585703](https://doi.org/10.1101/2024.03.19.585703).
- 590 **Morris DH**, Gostic KM, Pompei S, Bedford T, Łuksza M, Neher RA, Grenfell BT, Lässig M, McCauley JW. Predictive
591 modeling of influenza shows the promise of applied evolutionary biology. *Trends Microbiol*. 2018 Oct; .
- 592 **Mulligan MJ**, Lyke KE, Kitchin N, Absalon J, Gurtman A, Lockhart S, Neuzil K, Raabe V, Bailey R, Swanson KA, Li
593 P, Koury K, Kalina W, Cooper D, Fontes-Garfias C, Shi PY, reci Ö, Tompkins KR, Walsh EE, Frenck R, et al. I/II
594 study of COVID-19 RNA vaccine BNT162b1 in adults. *Nature*. 2020 Oct; 586(7830):589–593.
- 595 **Neher RA**, Russell CA, Shraiman BI. Predicting evolution from the shape of genealogical trees. *Elife*. 2014 Nov;
596 3:e03568.
- 597 **Nguyen LT**, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: A fast and effective stochastic algorithm for
598 estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*. 2014 11; 32(1):268–274.
599 <https://doi.org/10.1093/molbev/msu300>, doi: 10.1093/molbev/msu300.
- 600 **O'Toole A**, Scher E, Underwood A, Jackson B, Hill V, McCrone JT, Colquhoun R, Ruis C, Abu-Dahab K, Taylor B,
601 Yeats C, du Plessis L, Maloney D, Medd N, Attwood SW, Aanensen DM, Holmes EC, Pybus OG, Rambaut A.
602 Assignment of epidemiological lineages in an emerging pandemic using the pangolin tool. *Virus Evol*. 2021;
603 7(2):veab064.
- 604 **Pardi N**, Parkhouse K, Kirkpatrick E, McMahon M, Zost SJ, Mui BL, Tam YK, ó K, Barbosa CJ, Madden TD, Hope
605 MJ, Krammer F, Hensley SE, Weissman D. Nucleoside-modified mRNA immunization elicits influenza virus
606 hemagglutinin stalk-specific antibodies. *Nat Commun*. 2018 Aug; 9(1):3361.
- 607 **Petrova VN**, Russell CA. The evolution of seasonal influenza viruses. *Nature Reviews Microbiology*. 2018 Jan;
608 16(1):47–60. <https://doi.org/10.1038/nrmicro.2017.118>, doi: [10.1038/nrmicro.2017.118](https://doi.org/10.1038/nrmicro.2017.118).
- 609 **Petsch B**, Schnee M, Vogel AB, Lange E, Hoffmann B, Voss D, Schlake T, Thess A, Kallen KJ, Stitz L, Kramps T.
610 Protective efficacy of in vitro synthesized, specific mRNA vaccines against influenza A virus infection. *Nat*
611 *Biotechnol*. 2012 Dec; 30(12):1210–1216.
- 612 **Rubner Y**, Tomasi C, Guibas LJ. A metric for distributions with applications to image databases.
613 In: *Sixth International Conference on Computer Vision (IEEE Cat. No.98CH36271)*; 1998. p. 59–66. doi:
614 [10.1109/ICCV.1998.710701](https://doi.org/10.1109/ICCV.1998.710701).
- 615 **Sagulenko P**, Puller V, Neher RA. TreeTime: Maximum-likelihood phylodynamic analysis. *Virus Evolution*. 2018
616 01; 4(1). <https://doi.org/10.1093/ve/vex042>, doi: 10.1093/ve/vex042.
- 617 **Shu Y**, McCauley J. GISAID: Global initiative on sharing all influenza data – from vision to reality. *Eurosurveillance*.
618 2017; 22(13).
- 619 **Smith DJ**, Lapedes AS, de Jong JC, Bestebroer TM, Rimmelzwaan GF, Osterhaus ADME, Fouchier RAM. Mapping
620 the antigenic and genetic evolution of influenza virus. *Science*. 2004 Jul; 305(5682):371–376.
- 621 **Steinbrück L**, Kligen TR, McHardy AC. Computational prediction of vaccine strains for human influenza A
622 (H3N2) viruses. *J Virol*. 2014 Oct; 88(20):12123–12132.
- 623 **Virtanen P**, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, Burovski E, Peterson P, Weckesser
624 W, Bright J, van der Walt SJ, Brett M, Wilson J, Millman KJ, Mayorov N, Nelson ARJ, Jones E, Kern R, Larson E,
625 Carey CJ, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*. 2020
626 Mar; 17(3):261–272. <https://www.nature.com/articles/s41592-019-0686-2>, doi: 10.1038/s41592-019-0686-2,
627 number: 3 Publisher: Nature Publishing Group.

It is made available under a [CC-BY 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

628 **Welsh FC**, Eguia RT, Lee JM, Haddock HK, Galloway J, Chau NVV, Loes AN, Huddleston J, Yu TC, Le MQ, Nhat NTD,
629 Thanh NTL, Greninger AL, Chu HY, Englund JA, Bedford T, Matsen FA, Boni MF, Bloom JD. Age-dependent
630 heterogeneity in the antigenic effects of mutations to influenza hemagglutinin. *bioRxiv*. 2023 Dec; .

631 **Wong SS**, Webby RJ. Traditional and new influenza vaccines. *Clin Microbiol Rev*. 2013 Jul; 26(3):476–492.

632 **World Health Organization**. Seasonal influenza fact sheet. Available at <http://www.who.int/mediacentre/factsheets/fs211/en/>; 2014.

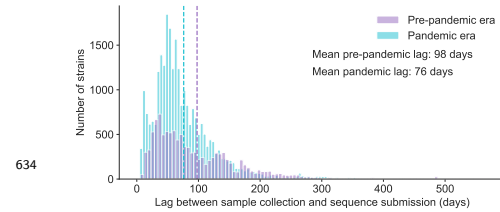


Figure 1—figure supplement 1. Distribution of submission lags in days for the pre-pandemic era (2019-2020 in blue) and pandemic era (2022-2023 in orange). Vertical dashed lines represent mean lags for each distribution.

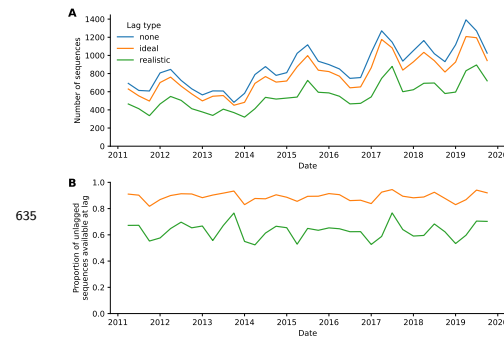


Figure 1—figure supplement 2. A) Number of A/H3N2 sequences available per timepoint and lag type. B) Proportion of all A/H3N2 sequences without lag per timepoint and lag type.

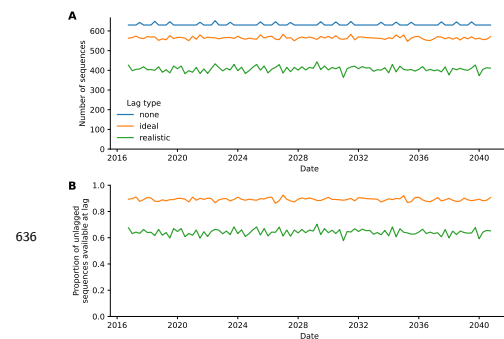


Figure 1—figure supplement 3. A) Number of simulated A/H3N2-like sequences available per timepoint and lag type. B) Proportion of all simulated A/H3N2-like sequences without lag per timepoint and lag type.

It is made available under a [CC-BY 4.0 International license](https://creativecommons.org/licenses/by/4.0/) .

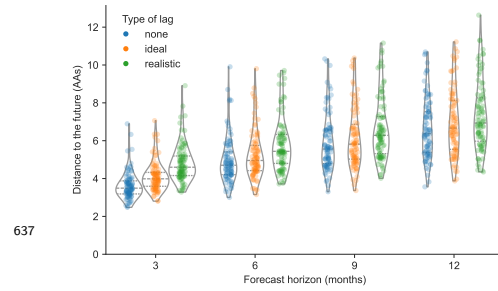


Figure 2—figure supplement 1. Distance to the future per timepoint (AAs) for simulated A/H3N2-like populations by forecast horizon and submission lag type based on forecasts from the “true fitness” model.

Figure 2—figure supplement 1—source data 1. Distances to the future for simulated A/H3N2-like populations; see <https://zenodo.org/records/13742375>

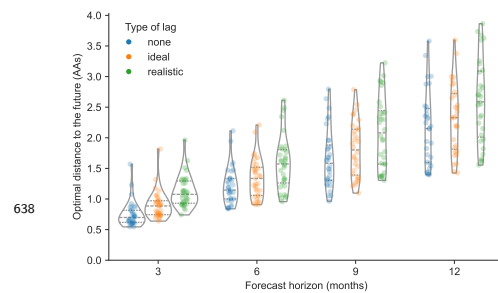


Figure 2—figure supplement 2. Optimal distance to the future per timepoint (AAs) for natural A/H3N2 populations by forecast horizon and submission lag type based on posthoc empirical fitness of the initial population.

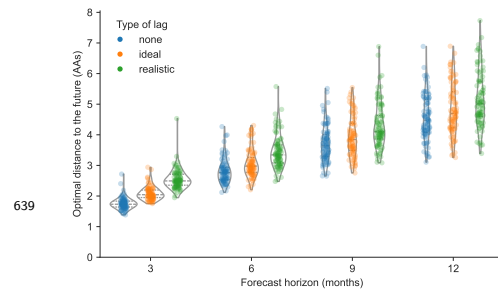


Figure 2—figure supplement 3. Optimal distance to the future per timepoint (AAs) for simulated A/H3N2-like populations by forecast horizon and submission lag type based on posthoc empirical fitness of the initial population.

It is made available under a [CC-BY 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

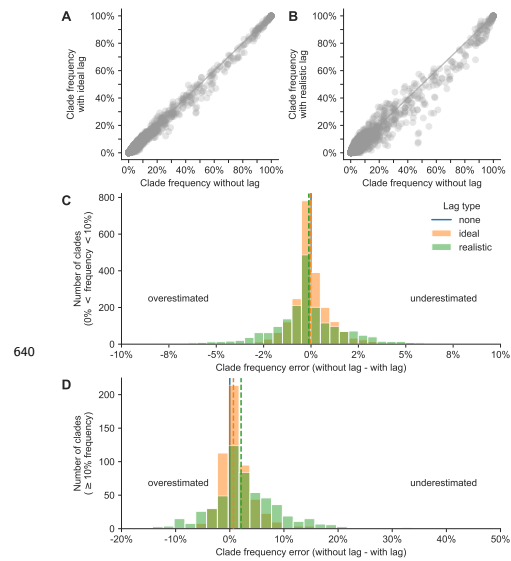


Figure 3—figure supplement 1. Clade frequency errors between simulated A/H3N2-like HA populations with ideal or realistic submission lags and populations without any submission lag.

Figure 3—figure supplement 1—source data 1. Current and future clade frequencies for simulated A/H3N2-like populations by forecast horizon and submission lag type; see <https://zenodo.org/records/13742375>

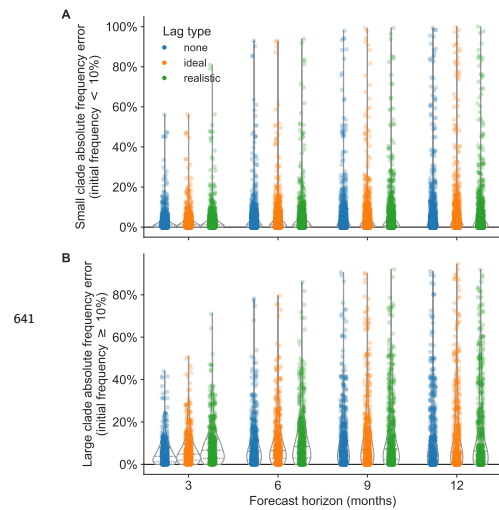


Figure 4—figure supplement 1. Absolute forecast clade frequency errors for simulated A/H3N2-like HA populations by forecast horizon in months and submission lag type (none, ideal, or realistic) for A) small clades (<10% initial frequency) and B) large clades (≥10% initial frequency).

It is made available under a [CC-BY 4.0 International license](https://creativecommons.org/licenses/by/4.0/) .

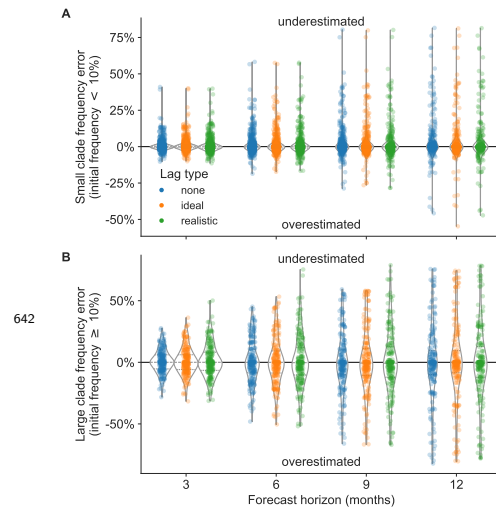


Figure 4—figure supplement 2. Forecast clade frequency errors for natural A/H3N2 HA populations by forecast horizon in months and submission lag type (none, ideal, or realistic) for A) small clades (<10% initial frequency) and B) large clades (≥10% initial frequency).

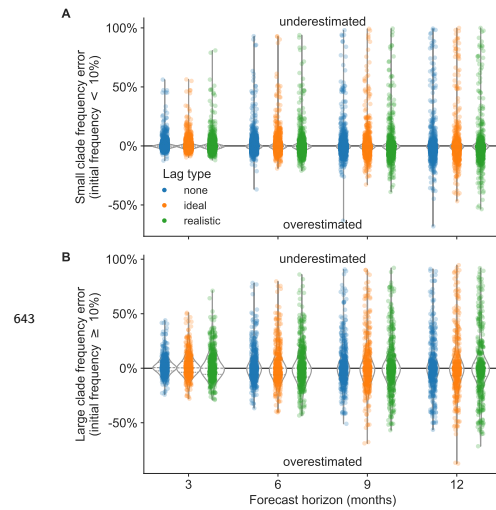


Figure 4—figure supplement 3. Forecast clade frequency errors for simulated A/H3N2-like HA populations by forecast horizon in months and submission lag type (none, ideal, or realistic) for A) small clades (<10% initial frequency) and B) large clades (≥10% initial frequency).

It is made available under a [CC-BY 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

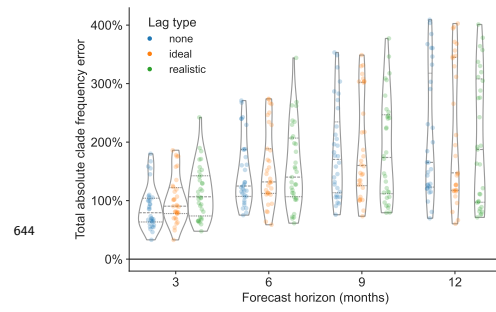


Figure 5—figure supplement 1. Distribution of total absolute clade frequency errors summed across clades per future timepoint for A/H3N2 populations. We calculated the effects of interventions as the difference between these values per future timepoint under the status quo (12-month forecast horizon and realistic submission lag) and specific interventions.

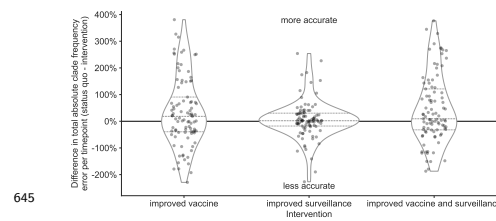


Figure 5—figure supplement 2. Improvement of clade frequency errors for simulated A/H3N2-like populations between the status quo and realistic interventions.

Figure 5—figure supplement 2—source data 1. Differences in total absolute clade frequency error per future timepoint and clade between the status quo and realistic interventions for simulated A/H3N2-like populations; see <https://zenodo.org/records/13742375>

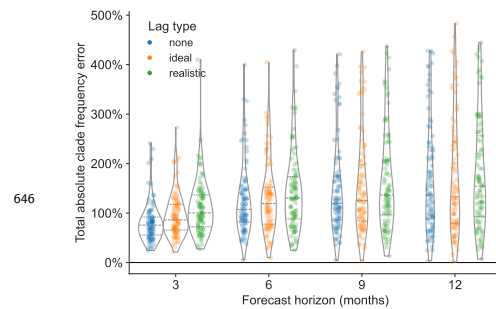
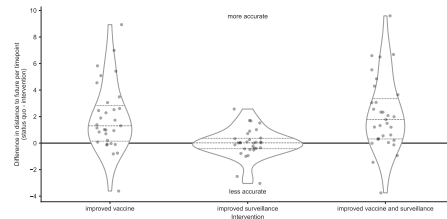


Figure 5—figure supplement 3. Distribution of total absolute clade frequency errors summed across clades per future timepoint for simulated A/H3N2-like populations.

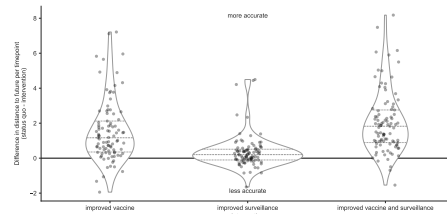
It is made available under a [CC-BY 4.0 International license](https://creativecommons.org/licenses/by/4.0/).



647

Figure 5—figure supplement 4. Improvement of distances to the future (AAs) for A/H3N2 populations between the status quo (12-month forecast horizon and realistic submission lags) and realistic interventions. The effects of interventions are the differences between distances to the future per future timepoint under the status quo and specific interventions.

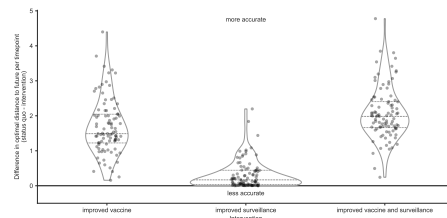
Figure 5—figure supplement 4—source data 1. Improvement of distances to the future per future timepoint for A/H3N2 populations; see <https://zenodo.org/records/13742375>



648

Figure 5—figure supplement 5. Improvement of distances to the future (AAs) for simulated A/H3N2-like populations between the status quo (12-month forecast horizon and realistic submission lags) and realistic interventions. The effects of interventions are the differences between distances to the future per future timepoint under the status quo and specific interventions.

Figure 5—figure supplement 5—source data 1. Improvement of distances to the future per future timepoint for simulated A/H3N2-like populations; see <https://zenodo.org/records/13742375>



649

Figure 6—figure supplement 1. Improvement of optimal distances to the future (AAs) for simulated A/H3N2-like populations between the status quo (12-month forecast horizon and realistic submission lags) and realistic interventions.

Figure 6—figure supplement 1—source data 1. Improvement of optimal distances to the future per future timepoint for simulated A/H3N2-like populations; see <https://zenodo.org/records/13742375>