

## An Application of Nowcasting Methods: Cases of Norovirus during the Winter 2023/2024 in England

**Authors:** Jonathon Mellor<sup>1\*</sup>, Maria L Tang<sup>1</sup>, Emilie Finch<sup>1,2</sup>, Rachel Christie<sup>1</sup>, Oliver Polhill<sup>1</sup>, Christopher E Overton<sup>1,3</sup>, Ann Hoban<sup>4</sup>, Amy Douglas<sup>4</sup>, Sarah R Deeny<sup>1</sup>, Thomas Ward<sup>1</sup>

1. Data Analytics and Surveillance Group, UK Health Security Agency, London, UK
2. Centre for Mathematical Modelling of Infectious Diseases, London School of Hygiene and Tropical Medicine, London, UK
3. Department of Mathematical Sciences, University of Liverpool, Liverpool, UK
4. Gastrointestinal Infections, Food Safety and One Health Division, UK Health Security Agency, London, UK

\*Corresponding author: [Jonathon.mellor@ukhsa.gov.uk](mailto:Jonathon.mellor@ukhsa.gov.uk)

### Abstract

#### *Background*

Norovirus is a leading cause of acute gastroenteritis, adding to strain on healthcare systems. Diagnostic test reporting of norovirus is often delayed, resulting in incomplete data for real-time surveillance.

#### *Methods*

To nowcast the real-time case burden of norovirus a generalised additive model, semi-mechanistic Bayesian joint process and delay model, and Bayesian structural time series model including syndromic surveillance data were developed. These models were evaluated over weekly nowcasts using a probabilistic scoring framework.

#### *Results*

Modelling current cases clearly outperforms a simple heuristic approach. Models that harnessed a time delay correction had higher skill, overall, relative to forecasting techniques. However, forecasting approaches were found to be more reliable in the event of temporally changeable reporting patterns. The incorporation of norovirus syndromic surveillance data was not shown to improve model skill in this nowcasting task, which may be indicative poor correlation between the indicator and norovirus incidence.

#### *Interpretation*

Analysis of surveillance data enhanced by nowcasting reporting delays improves understanding over simple model assumptions, which is important for real-time decision making. The structure of the modelling approach needs to be informed by the patterns of the reporting delay and can have large impacts on operational performance and insights produced.

## 1 **Introduction**

2 Norovirus is a gastrointestinal RNA virus causing symptoms of nausea, vomiting and  
3 diarrhoea. Norovirus often causes outbreaks in enclosed settings, such as hospitals, care  
4 homes and nurseries [1], causing a substantial burden on health systems, particularly over  
5 winter [2] [3]. Norovirus transmission was limited during lockdown periods of the SARS-CoV-  
6 2 pandemic response, followed by resurgent spreading when normative population mixing  
7 patterns resumed to pre-pandemic levels [4]. Norovirus is a constantly evolving pathogen  
8 with antigenic drift and shift [5] leading to strain replacement events periodically [6] [7] and  
9 resulting in short-lived immunity. These events cause large outbreaks and elevated  
10 transmission, highlighting the importance of monitoring and improving the timeliness of  
11 insights for taking public health action.

12 Norovirus surveillance in England uses data from multiple national surveillance systems.  
13 These include norovirus positive laboratory reports from confirmed cases, of which a subset  
14 undergo molecular typing, as well as notifications of outbreaks [8]. There is a reporting  
15 delay between diagnostic test administration and reporting to the national surveillance  
16 data, partially attributable to norovirus not being a Schedule 2 notifiable causative agent in  
17 legislation [9]. Due to this lag, the national official statistics surveillance reports truncate the  
18 time period shown by one week to remove partially complete data [8].

19 Norovirus is an excellent candidate for the application of nowcasting methods due to the  
20 inherent delay in case reporting as a non-priority pathogen. Research has been conducted  
21 on short term projections using statistical methods [10] [11], though there is limited  
22 exploration of correcting for time delays in norovirus cases. Norovirus incidence is highly  
23 stochastic, with a partially seasonal pattern and high heterogeneity between localised  
24 outbreaks and national trends, making it challenging to predict. Building on nowcasting  
25 research applied during the SARS-CoV-2 pandemic [12] [13] modelling can be explored to  
26 improve understanding of the real-time norovirus dynamics.

27 In this paper, we explored the reporting delay for norovirus cases in England over the  
28 2023/2024 winter. We evaluated a range of methods for nowcasting using different model  
29 structures, guide signals, and assumptions about data completeness to consider the trade-  
30 offs between different approaches applicable to norovirus and beyond.

## 31 **Methods**

### 32 **Data**

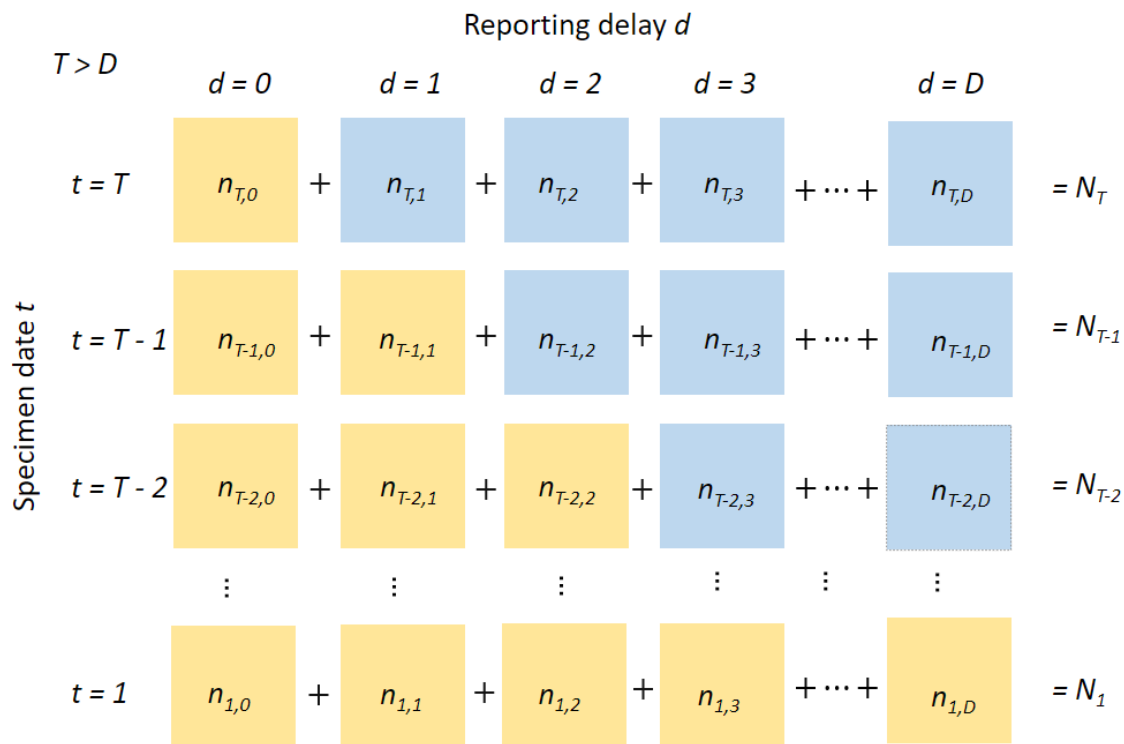
#### 33 **Norovirus Cases**

34 Individual test results were extracted from the Second Generation Surveillance Service  
35 (SGSS) database in UKHSA (UK Health Security Agency) [14] for positive norovirus tests  
36 conducted in England. The database only stores information on positive laboratory test  
37 results uploaded by frontline diagnostic laboratories, with a sampling bias towards health  
38 and social care settings. We deduplicated tests to keep the first test per patient infection  
39 episode. Under the legislation positive norovirus diagnostic tests are required to be notified

40 to the UKHSA, but not required within 7 days of testing [15]. Cases followed a day-of-week  
41 periodicity shown in Supplementary Figure 1.

42 In this analysis we focused on two main time events for each test. Firstly, the specimen date  
43  $t$  defines when the specimen was collected from the infected individual for testing.  
44 Secondly, the report date  $t_r$  defines the date the record is ingested into SGSS, thereby  
45 notifying UKHSA of the norovirus case through national surveillance. As symptom onset  
46 dates are not reported, the specimen date is the most epidemiologically relevant event.  
47 Despite being impacted by time to treatment, the specimen date gave the least delayed  
48 representation of the epidemic's progression compared with other available time events for  
49 each test. The difference between report date and specimen date  $d = t_r - t$  is the  
50 reporting delay.

51 To model the epidemic and corresponding delay distributions, we aggregated the data by  $t$   
52 and  $d$  to construct a so-called data "reporting triangle" [16], illustrated in Figure 1. The  
53 reporting triangle is an array with elements  $n_{t,d}$ , for  $t \in [1, T]$  and  $d \in [0, D]$ , where  $T$  is the  
54 maximum length of the specimen date time series  $t$  and  $D$  is the maximum reporting delay.  
55 The element  $n_{t,d}$  represents the number of tests collected on the  $t^{\text{th}}$  day of the specimen  
56 date time series that were reported after  $d$  days. In theory,  $D$  could be very large. However,  
57 in practice most reporting delays are under 10 days. Therefore, for this analysis, we assume  
58 a maximum possible reported delay of 50, though each model may assume a shorter value.  
59 In real-time, cases  $n_{t,d}$  cannot exist when  $d > T - t$  which introduces a right truncation.  
60 Therefore, for cases at  $t = T$  only cases with  $d = 0$  can be known, with other values  
61  $1 \leq d \leq D$  unknown. The quantity of most interest used to inform decision making and  
62 proactive communications was the total cases by specimen date  $N_t$ . The reporting triangle is  
63 therefore collapsed into  $N_t = \sum_{d=0}^D n_{t,d}$ . To support operational needs, these daily counts  
64 are also aggregated to weekly levels for ease of interpretation.



65

66 *Figure 1. Illustration of the reporting delay data triangle structure, with elements of the 2-dimensional array. Horizontal*  
 67 *axis represents the report delay and vertical axis the specimen date. Complete data per specimen date correspond to the*  
 68 *sum of each row across the reporting delays. Each cell represents the case count for a given specimen date and reporting*  
 69 *delay. Case counts are unknown in real-time when  $d > T - t$ , represented here by blue cells.*

## 70 **NHS 111 Online Pathways**

71 While there is a delay in case reporting, other data sources are complete in real-time and  
 72 rapidly available. These data were be leveraged to inform case prediction. NHS 111 Online  
 73 Pathways is an online triage service in the UK used to give non-emergency healthcare  
 74 guidance to individuals [17]. Users are routed to appropriate guidance given input  
 75 information about their symptoms. We transformed these inputs into symptom categories,  
 76 , and calculate counts of symptom triages, , by time, and symptom category.  
 77 Symptom categories and groupings are given in Supplementary Table 1, with visualisations  
 78 of the trends in Supplementary Figure 2 & 3.

## 79 **Models**

80 The aim of our nowcasting models was to estimate the expected complete number of cases  
 81 that have been collected during the most recent 7 days, . Some models harness the  
 82 partial reporting of recent cases correcting for the delay distribution, others ignore this  
 83 partial reporting. We aimed to select methods that perform well against the norovirus  
 84 dynamics observed. Models were tuned for appropriate parameter selection over the 4-  
 85 week period using weeks ending 8 October 2023 to 29 October 2023, then applied to the  
 86 remainder of the weeks to 10 March 2024, to avoid parameter selection using evaluation  
 87 data. Models are tuned based on the average daily scores for the most recent 7 days, as  
 88 outlined in the evaluation section. Model structures and assumptions are given in Table 1.

## 89 Baseline

90 To contextualise the performance of the models, we implemented a simple baseline  
91 approach to compare against. We assumed each predicted day will be equal to the observed  
92 count the previous week giving an autocorrelated prediction with day-of-week effects.

93 The central estimate is set as  $\bar{N}_t = \sum_{d=0}^{T-(t-7)} n_{t-7,d}$ , which corresponds to the reported data  
94 from the seven days prior – matching the weekly reporting cycle in surveillance. Most  
95 norovirus cases were reported with  $d \leq 7$  and as such this method gives predictions of near  
96 complete case numbers. We did not consider uncertainty within the baseline method. For  
97 application of the scoring methodology, prediction intervals are required. Therefore, for the  
98 baseline model the prediction intervals were assumed equal to the central estimate.

## 99 Generalised Additive Model

100 We used a generalised additive model (GAM) utilising partially reported data, based on a  
101 nowcasting model for mpox [18] [19]. This estimated the total number of cases with  
102 specimen date  $t$ ,  $\bar{N}_t$ , as the sum of known data that has already been reported,  $n_{t,d}$ , for  
103 reporting delays  $d \in [0, T - t]$ , and estimates for the unknown data yet to be reported,  
104  $\bar{n}_{t,d}$ , for reporting delays  $d \in [T - t + 1, D]$ , i.e.

$$105 \quad \bar{N}_t = \sum_{d=0}^{T-t} n_{t,d} + \sum_{d=T-t+1}^D \bar{n}_{t,d} \quad (1)$$

106 As  $n_{t,d}$  is known,  $\bar{N}_t$  has a natural lower bound of  $\sum_{d=0}^{T-t} n_{t,d}$ . The unknown data was modelled  
107 with a negative binomial distribution accounting for the non-negative integer values and  
108 overdispersion. Using the mean and variance parameterisation,

$$n_{t,d} \sim \text{NegBin}(\mu_{t,d}, \mu_{t,d} + \mu_{t,d}^2/k)$$

109 with dispersion parameter  $k$ . We use a log link function to model the exponential epidemic  
110 process, where  $\mu_{t,d}$  depends on both  $t$  and  $d$  according to

$$111 \quad \log(\mu_{t,d}) = \beta_0 + s_1(t) + s_2(d) + \omega_1(\text{wday}(t)) + \omega_2(\text{wday}(t + d)).$$

112 where  $\beta_0$  is a constant. We assumed that the number of cases vary smoothly over specimen  
113 date  $t$  and number of days delay  $d$  as  $s_1(t)$  and  $s_2(d)$ , with random day-of-week effects  
114  $\omega_1(\text{wday}(t))$  and  $\omega_2(\text{wday}(t + d))$  respectively. The model was fitted in  $R$  using the *gam*  
115 function from the *mgcv* package [20]. 1000 burn-in and posterior samples were drawn from  
116 the model using the *gratia* package [21] with a Metropolis-Hastings sampler. Samples were  
117 aggregated to  $\bar{N}_t$  (eqn. 1), with prediction intervals taken using quantiles of these samples.  
118 Models were fit to the past 56 days, with cubic regression basis functions every  $l = 7$  days  
119 for  $s_1(t)$  and  $s_2(d)$ , and a maximum reporting delay  $D = 14$ . Model tuning is outlined in  
120 Supplementary Section 2.

## 121 Epinowcast

122 We also used a Bayesian hierarchical nowcasting framework implemented in the *epinowcast*  
123 package [22], with the implementation described below. This approach builds on earlier  
124 nowcasting approaches [23] [24]. As with the “GAM” model, the estimate for the total

125 number of tests with specimen date  $t$ ,  $\bar{N}_t$ , is the sum of known data,  $n_{t,d}$ , and estimates for  
 126 the unknown data,  $\bar{n}_{t,d}$  (eqn. 1).

127 Here  $n_{t,d} | N_t$  follows a multinomial distribution with a probability vector  $(p_{t,d})$  that is  
 128 estimated jointly with the expected number of final reported cases. This differs from the  
 129 “GAM” model approach, where each  $n_{t,d}$  is independent. We used the default  
 130 implementation of modelling expected final reported cases as a first order random walk,

$$E[N_t] = \lambda_t$$

$$\log(\lambda_t) \sim \text{Normal}(\log(\lambda_{t-1}), \sigma^\lambda)$$

131  $\log(\lambda_0) \sim \text{Normal}(\log(N_0 + 1), 1)$

132  $\sigma^\lambda \sim \text{HalfNormal}(0, 1)$ .

133 The instantaneous growth rate  $r_t$  is defined as the log of the expected number of final  
 134 reported tests between time  $t$  and  $t - 1$ .  $r_t$  is then modelled on the log scale by a daily  
 135 random effect  $\omega_1(t)$  and a random effect for the day of the week  $\omega_2(\text{wday}(t))$ , to account  
 136 for weekly periodicity in the underlying data.

$$\log(r_t) = \omega_1(t) + \omega_2(\text{wday}(t))$$

137 Within *epinowcast* the delay distribution  $(p_{t,d})$  is then defined as a discrete time hazard  
 138 model where:

139 
$$h_{t,d} = P(\text{delay} = d | \text{day} \geq d, W_{t,d}).$$

140 Here, the hazard is determined by a design matrix  $W_{t,d}$  including a baseline delay  
 141 distribution and time- and delay- specific covariates which affect the reporting delay. We  
 142 assume the probability of reporting  $p_{t,d}^i$  follows a discretised log-normal distribution where  
 143 the log mean and log standard deviation are modelled with a daily random effect (the model  
 144 default).

$$p_{t,d}^i \sim \text{LogNormal}(\mu_t, \nu_t)$$

145 where the parametric logit hazard  $\gamma_{td}$  is given by

$$\gamma_{td} = \text{logit} \left( \frac{p'_{t,d}}{\left(1 - \sum_{d'=0}^{d-1} p'_{t,d'}\right)} \right)$$

146 We also use a constant non-parametric logit hazard such that:

$$\epsilon_{t,d} = \beta_0$$

147 The overall hazard is then modelled as  $\text{logit}(h_{t,d}) = \gamma_{td} + \epsilon_{t,d}$

148 To estimate final observed reported cases a negative binomial observation model is used  
 149 where:

$$\bar{n}_{t,d} | \lambda_t, p_{t,d} \sim \text{NegBin}(\lambda_t \times p_{t,d}, \phi), t = 1, \dots, T$$

150 and the overdispersion parameter  $\phi$  is estimated with a prior of

$$\frac{1}{\sqrt{\phi}} \sim \text{HalfNormal}(0,1)$$

151 and  $\bar{N}_t$  is given by (eqn. 1).

152 Unlike the “GAM” model, this approach introduces parametric, discrete, and truncated  
153 distributions for the reporting delay, better reflecting the reporting measurements. Models  
154 are fit in *stan* with *cmdstan* [25] using the Hamiltonian Monte Carlo (HMC) with NUTS (No-  
155 U-Turn Sampler). We ran 1000 iterations for warm-up and 1000 post-warmup iterations. A  
156 maximum reporting delay of 7 days, with a training length of 35 was selected. Model tuning  
157 and prior specification are outlined in Supplementary Section 3.

### 158 **Bayesian Structural Time Series**

159 We employed a flexible Bayesian structural time series (BSTS) modelling approach to  
160 produce a nowcast without harnessing partial reported case counts. The time series  $N_t$  is  
161 truncated by 7 days, with the unknown daily counts estimated in a traditional forecasting  
162 approach. The BSTS allows for a state space specification with decomposition of time  
163 varying dynamics including trend, seasonality and regression effects [26]. We create two  
164 models using the *bsts* R package [27], one without regressors, the second using 111 online  
165 indicators.

166 The first model “BSTS” is defined by the following state space equations, where at time  $t$ ,  
167 we have mean  $\mu_t$ , slope  $\delta_t$  and seasonal component  $\tau_t$ , with a season as  $S = 7$  days to  
168 capture the day-of-week effects.

$$169 \quad \log(\lambda_t) = \mu_t + \tau_t \quad \text{where } N_t \sim \text{Poisson}(\lambda_t) \quad (2)$$

170 The equation for the mean  $\mu_t$  is given by

$$171 \quad \mu_{t+1} = \mu_t + \delta_t + \eta_{0,t} \quad \text{with } \eta_{0,t} \sim \mathcal{N}(0, \sigma_\mu) \quad (3)$$

172 and the slope,

$$173 \quad \delta_{t+1} = \delta_t + \eta_{1,t} \quad \text{and } \eta_{1,t} \sim \mathcal{N}(0, \sigma_\delta). \quad (4)$$

174 Lastly the seasonality component is determined via dummy regression variables,

$$\tau_{t+1} = - \sum_{s=1}^{S-1} \tau_{t-s+1} + \eta_{2,t}$$

$$175 \quad \text{with } \eta_{2,t} \sim \mathcal{N}(0, \sigma_\tau). \quad (5)$$

176 This ensures that the seasonal component  $\tau_t$  accounts for the cumulative seasonal effects  
177 over the specified period  $S$ , in our case one week. Therefore,  $\log(\lambda_t)$  follows a local linear  
178 trend with seasonality, where the mean and slope of the trend are assumed to follow  
179 random walks. For the “BSTS” model, a training length of 60 days was chosen, with upper  
180 limits of  $\exp(\sigma_\mu)$  and  $\exp(\sigma_\delta)$  equal to 1.1. Model tuning is outlined in Supplementary

181 Section 4. The models were fit via Gibbs sampling MCMC, run for 50,000 iterations with  
 182 2,000 burn in.

183 To produce the second model “BSTS + NHS 111 online” we update the observational  
 184 equation (1) to include the  $i$  regressor symptom category scaled counts  $x_{i,t}$  in  $\mathbf{x}_t$

185 
$$\log(\lambda_t) = \mu_t + \tau_t + \beta^T \mathbf{x}_t, \quad \text{where } N_t \sim \text{Poisson}(\lambda_t).$$

186 The  $\beta_i$  values are estimated using spike and slab priors [28] centred on zero to allow for  
 187 sensible variable selection. For the “BSTS + NHS 111 online” model we choose a training  
 188 length of 150 days, 5 expected regression coefficients (through the spike and slab prior), and  
 189 an upper limit for  $\exp(\sigma_\mu)$  of 1.01 and  $\exp(\sigma_\delta)$  of 1.1. Model tuning analysis is given in  
 190 Supplementary Section 4.

191 **Models Overview**

Property	Baseline	BSTS	BSTS + NHS 111 online	GAM	epinowcast
Uses partial reported data	No	No	No	Yes	Yes
Parametric reporting delay distribution	-	-	-	No	Yes
Supplementary indicator signal	No	No	Yes	No	No
Bayesian	No	Yes	Yes	No	Yes
Parameter estimation method	-	Gibbs	Gibbs	REML, sampling via Metropolis-Hastings	HMC with NUTS
Approximate runtime per week (fitting and post-processing)	0.1s	50s	1 min 45s	10s	10 min
Posterior samples (burn-in)	-	50,000 (2,000)	50,000 (2,000)	1,000 (1,000)	1,000 (1,000)

192 *Table 1. Summary of key model structures, assumptions, and characteristics to compare for each model.*

193 **Evaluation**

194 To compare the different nowcasting approaches we employ multiple scoring methods in a  
 195 probabilistic framework. The interval coverage is a measure of probabilistic calibration,  
 196 telling us the proportion of observations that are within given prediction interval ranges – in  
 197 our case 50% and 90%. From the interval coverage we calculate the coverage deviation, the  
 198 average difference between the measured interval coverage and the specified interval  
 199 value, with a coverage deviation nearer zero being preferred. The (weighted) interval score  
 200 (WIS) is a proper scoring rule composed of sharpness and under/overprediction, giving an  
 201 overall measure of performance where low values are better. The weighted interval skill

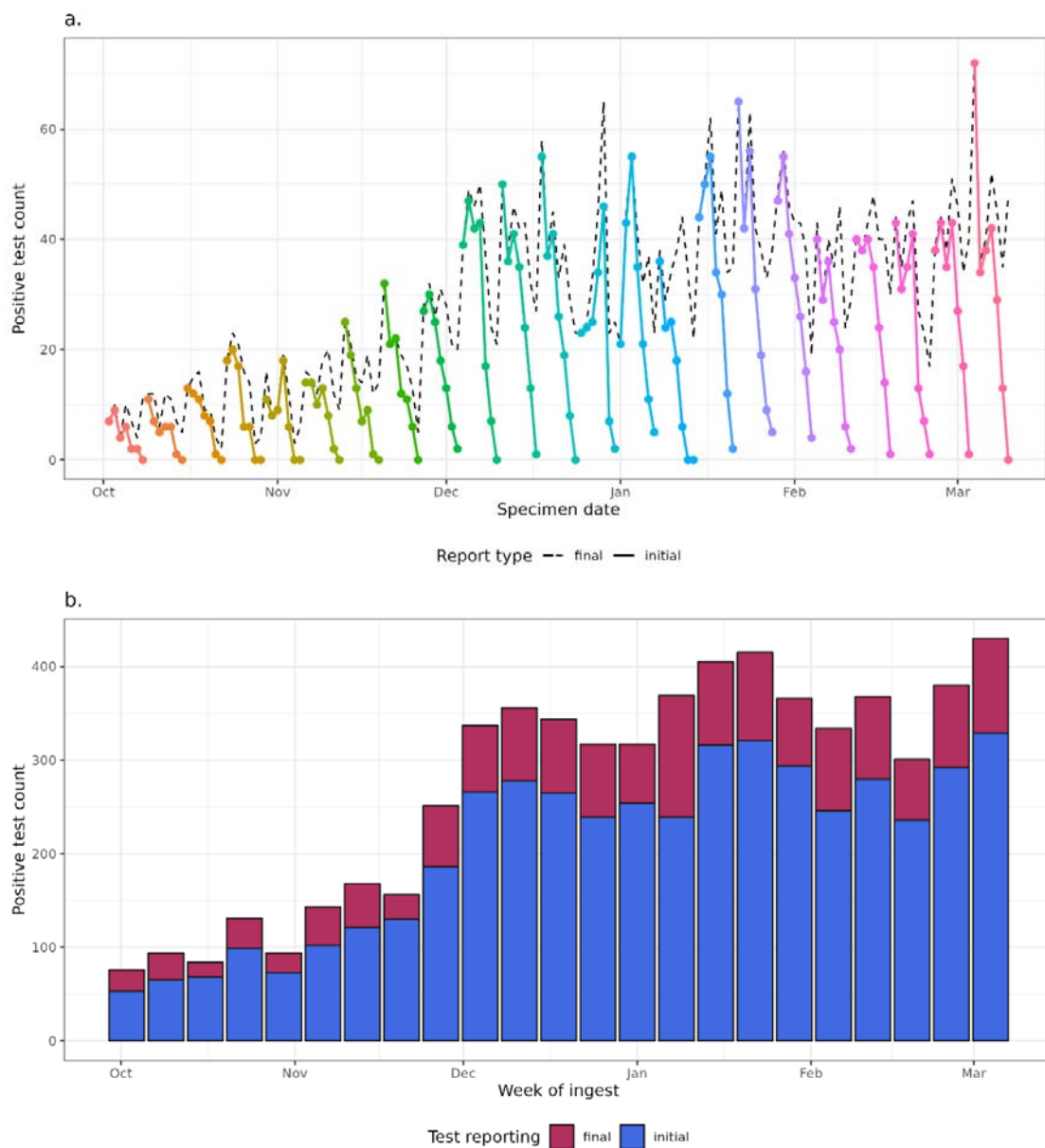


202 score is calculated as  $WISS_{\text{model}} = 1 - \frac{WIS_{\text{model}}}{WIS_{\text{baseline}}}$  where  $WISS_{\text{model}} > 0$  corresponds to a  
203 model better than the “baseline” model. The bias is a relative measure of  
204 under/overprediction telling us if the models systematically estimate high or low, with lower  
205 bias models having a score nearer zero. The median absolute error gives an average of the  
206 absolute difference between central prediction and true data. The scoring is conducted  
207 using the *scoringutils* package [29]. The estimates are scored at daily and weekly  
208 aggregations, as well as explored by nowcast horizon  $h$ , where  $h = T - t$  in our case is the  
209 day-of-week predicted. Since the data is uploaded weekly, the nowcast horizon  $h$   
210 corresponds to a unique day-of-week where *Sunday* will be a nowcast horizon of 0 days, and  
211 *Monday* will have a nowcast horizon of 6 days.

## 212 **Results**

213 Winter 2023/2024 followed the seasonal trend of increasing cases from September  
214 onwards, reaching a stable trend from December 2023 onwards. The difference between  
215 final and initial cases is largest in the most recent days each week, as expected, with  $n_{t,0}$   
216 near zero (Figure 2a). Across each week approximately 20% of the data are revisions (cases  
217 added the following week). These revisions can change the narrative of the real-time trend  
218 without correction (Figure 2b). The distribution of  $d$  shows few reports on  $d = 0$ , a peak at  
219 1-2 days and most reports within 7 days (Figure 3). The time varying reporting delay is given

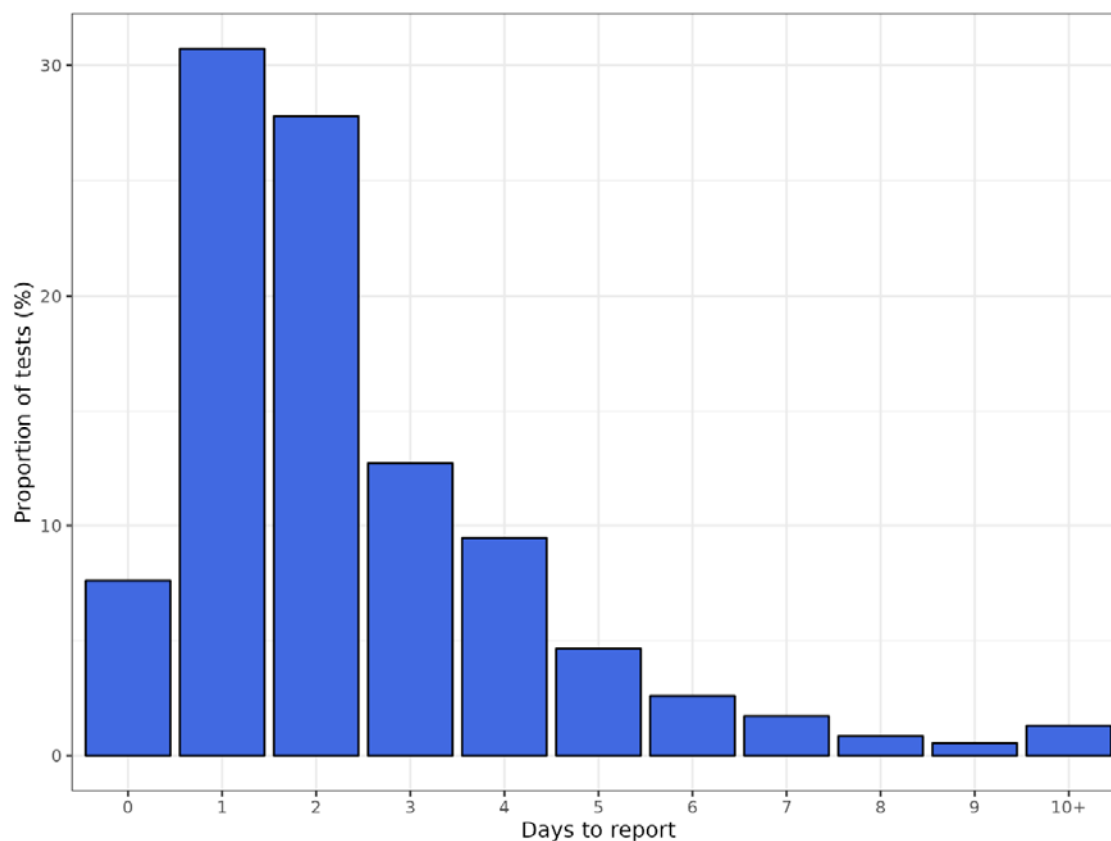
220 in Supplementary Figure 4, showing limited variation.



221

222 *Figure 2. The backfilling of norovirus tests over the Winter 2023/2024 season. (a.) daily counts of tests at different*  
223 *snapshots of reporting, showing the most recent observed counts are substantially lower than the final revised data. (b.)*  
224 *weekly counts of tests at each ingest and final revisions. The end date for each week was taken as a Sunday, to produce a*  
225 *nowcast of data from the previous week.*

226



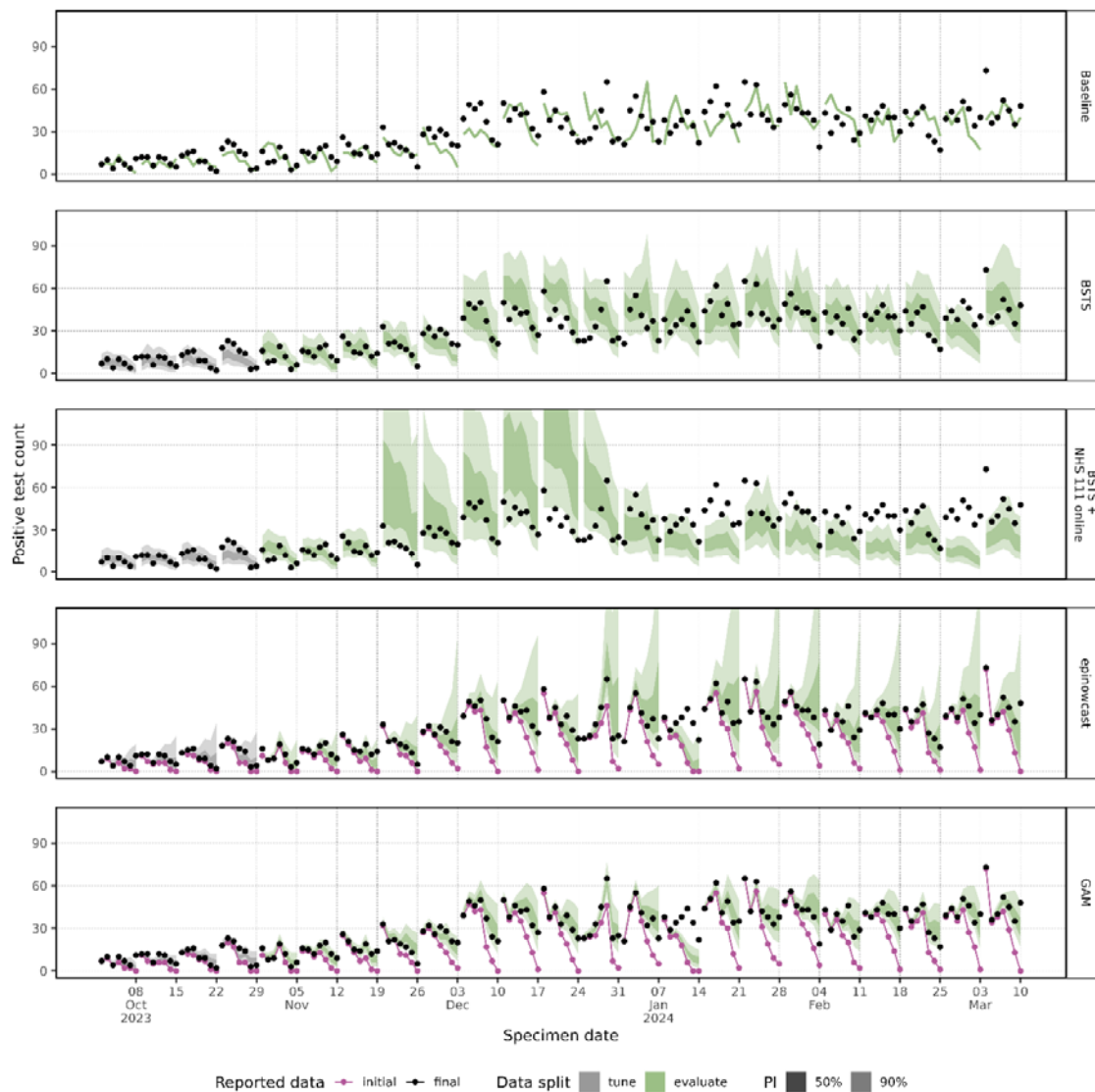
227

228 *Figure 3. Time delay distribution of days between specimen date and report date. Includes complete data from 02-10-2023*  
229 *to 10-03-2024.*

230 The daily and weekly nowcasts are shown over the tuning and evaluation time periods  
231 (Figure 4 & 5). Both the “GAM” and “epinowcast” models show increasing uncertainty  
232 towards the most recent date where data is more incomplete. The models using the  
233 partially complete data underpredict the complete cases in the week ending 14 January  
234 2024, which we also see in the weekly estimates (Figure 5), though the “BSTS” is not  
235 impacted in this way. The uncertainty in the weekly estimate varies substantially by model,  
236 though the “baseline” model has no associated uncertainty. The BSTS models have wide  
237 prediction intervals compared to the “GAM”, with the “epinowcast” model prediction  
238 intervals being skewed towards higher values.

239

240



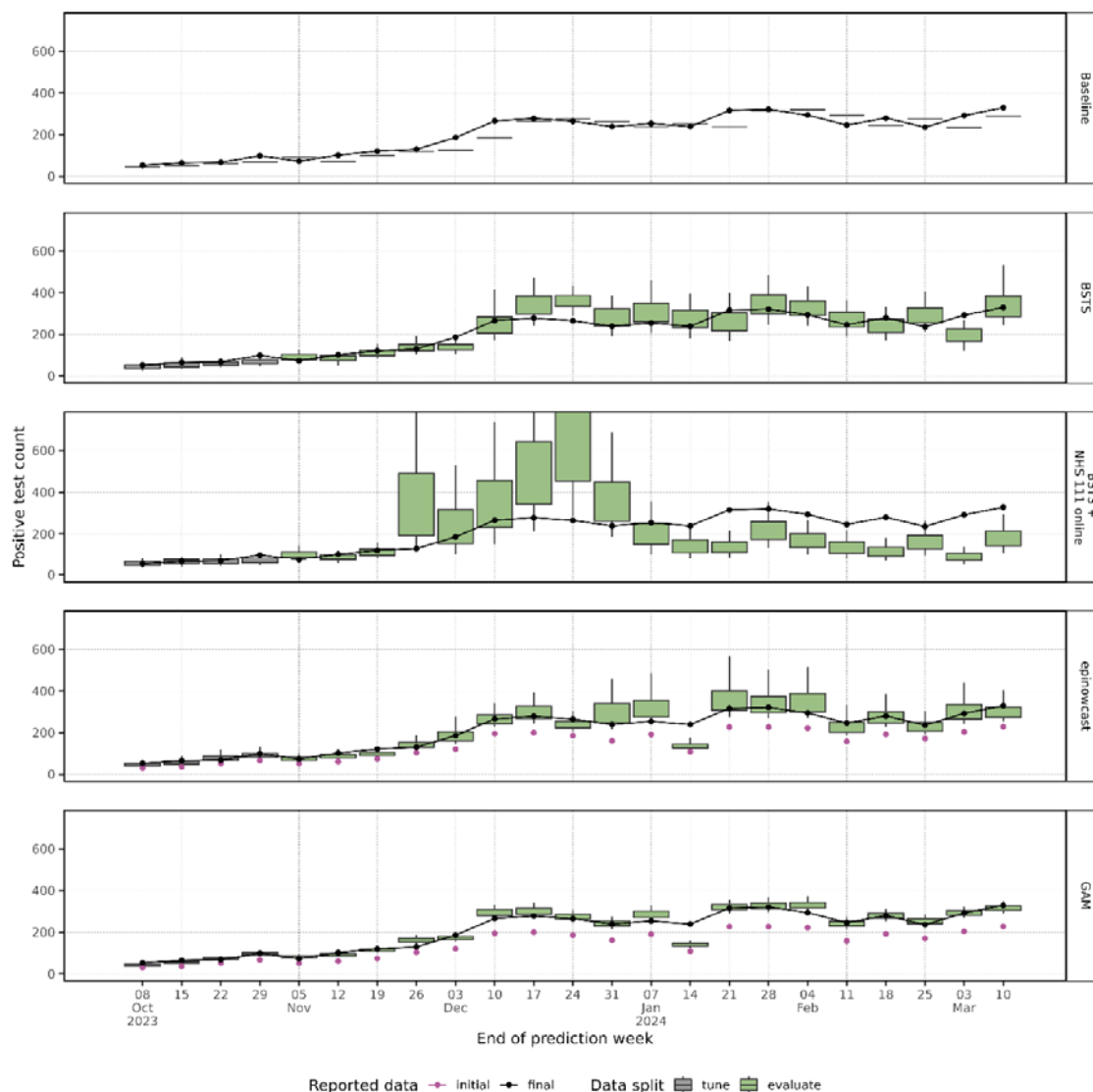
241

242

243

Figure 4. Daily predictions from all models with 50% and 90% prediction intervals against initial and final reported count of tests.

244



245

246 *Figure 5. Weekly predictions from all models with 50% (box) and 90% (whiskers) prediction intervals against initial and final*  
 247 *reported count of tests. The weekly predictions are created as the sum of sample predictions per week.*

248 The overall daily and weekly evaluation scores are shown in Table 2. The “baseline” model  
 249 has high WIS, expected given its small interval width. The partial reporting delay models  
 250 “epinowcast” and “GAM” outperform other models across WIS and MAE, generally  
 251 overpredicting, when other models are underpredicting. The “BSTS” model performs better  
 252 than the baseline across all daily metrics, whereas the “BSTS + NHS 111 online” performs  
 253 broadly worst. Across daily and weekly scoring the “BSTS” model has the best calibration  
 254 with lowest coverage deviation, though other models have similar values. Notably, the  
 255 “GAM” and “epinowcast” models over and underpredict respectively.

256

257

258

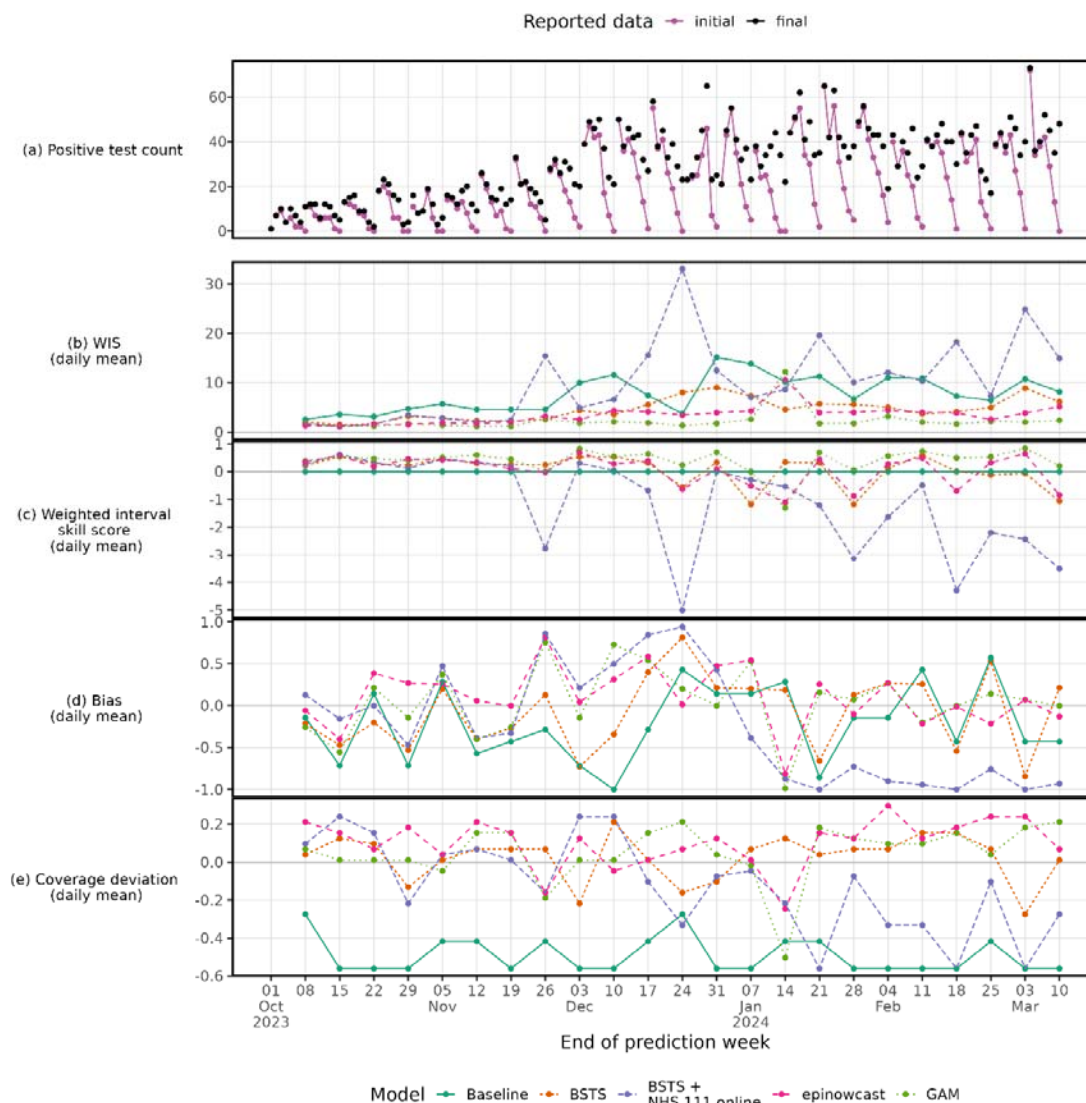
259

Temporal granularity	Model	WIS	Median absolute error	Bias	Coverage deviation
daily	Baseline	7.73	7.73	-0.21	-0.49
daily	BSTS	4.57	7.18	-0.07	<b>0.03</b>
daily	BSTS + NHS 111 online	10.28	15.36	-0.24	-0.12
daily	epinowcast	3.03	4.07	-0.31	0.08
daily	GAM	<b>2.29</b>	<b>3.39</b>	<b>0.05</b>	0.05
weekly	Baseline	29.74	29.74	-0.39	-0.56
weekly	BSTS	21.19	34.04	-0.05	<b>-0.06</b>
weekly	BSTS + NHS 111 online	67.44	100.35	-0.30	-0.20
weekly	epinowcast	15.61	22.35	-0.19	0.08
weekly	GAM	<b>11.56</b>	<b>16.00</b>	<b>0.04</b>	-0.11

260 *Table 2. Breakdown of overall model scores by temporal granularity. The daily granularity shows the average daily score*  
 261 *over the time series. The weekly granularity shows the average weekly score over the time series. The most optimal score by*  
 262 *temporal granularity and scoring metric is in bold.*

263 Over the evaluation period the “GAM”, “BSTS” and “epinowcast” models have improved  
 264 skill over the baseline model in most but not all weeks (Figure 6c). For much of the time  
 265 series, the “BSTS+NHS 111 online” model has higher WIS than the baseline model (Figure  
 266 6b). The “GAM” and “epinowcast” models have bias > 0 during the epidemic growth phase,  
 267 indicating overprediction (Figure 6c). The week of 14 January 2024 the “epinowcast” and  
 268 “GAM” perform markedly worse than other weeks, where initial reported data is  
 269 particularly low. Further scoring at daily and weekly levels are given in Supplementary  
 270 Figures 5 & 6.

271

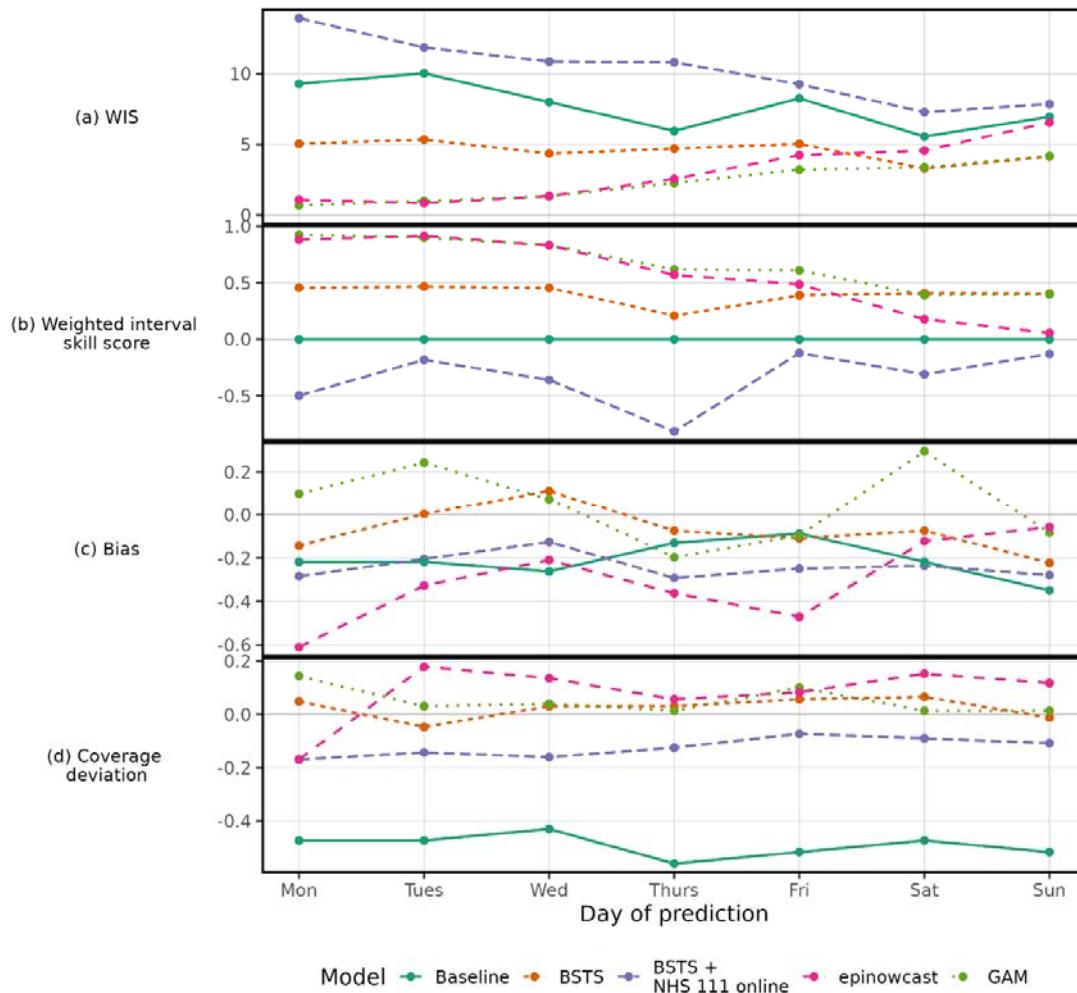


272

273 *Figure 6. Daily count of final and initial reported tests (a) with daily mean model scores for each prediction week. The*  
 274 *Weighted Interval Score (b), Weighted Interval Skill Score (c), Bias (d) and Coverage deviation (e) are given across models*  
 275 *and time.*

276 By breaking down by the day-of-week (and therefore nowcast horizon, in our case) we can  
 277 explore how varying data completeness affects model performance. Relative to “baseline”  
 278 the “BSTS” model exhibits a flat skill across days (Figure 7a), whereas the relative skill of the  
 279 “GAM” and “epinowcast” gets deteriorates towards the end of the week (Figure 7b). The  
 280 “baseline” consistently underpredicts, while “epinowcast” underpredicts at the start of the  
 281 week but becomes less biased toward Sunday (Figure 7c). Compared to the “BSTS” model,  
 282 the improved performance of the “GAM” model is primarily due to lower WIS early in the  
 283 prediction week when data is more complete.

284



285

286 *Figure 7. Model scores averaged over each day of prediction. A Monday has near complete data, whereas a Sunday has*  
 287 *many cases not yet reported. The scores are the average over the evaluation period.*

288

289

290 **Discussion**

291 Norovirus contributes substantially to health service winter pressures through hospital  
 292 outbreaks, reduced bed availability and staff absences. As such, timely surveillance is crucial  
 293 for situational awareness, particularly to understand changes in the epidemic curve in the  
 294 context of delayed reporting. In this work we applied a range of nowcasting approaches to  
 295 norovirus cases, with the aim of understanding the current epidemic state.. We have shown  
 296 that harnessing partially complete data outperforms a truncate-and-forecast approach, but  
 297 the performance can be sensitive to the consistency of case reporting, which is challenging  
 298 in frontline health protection. The delay in reporting impacts the analysis of trends in  
 299 national surveillance, so it is important official reporting exclude these partially reported  
 300 days, though nowcasting can support decision making in real-time. The nowcasting problem



301 presented is a straightforward application of time delay correction, with a small average  
302 delay, a single test type, and without considering regional or age-related variation. This may  
303 partially explain the strong performance of approximate methods in the scoring.

304 Nowcasting approaches are increasingly used to predict case counts by accounting for  
305 delays in reporting, and have been crucial in the recent COVID-19 pandemic and mpox  
306 outbreak [12] [24] [18] [30]. In this analysis, we apply several modelling approaches from  
307 the epidemic literature to this problem. We compare a well-principled Bayesian  
308 implementation, *epinowcast*, which jointly models a reporting delay distribution with an  
309 underlying process model, and a more approximate but highly flexible and computationally  
310 efficient GAM-based model. We also consider a Bayesian structural time series approach,  
311 testing the utility of incorporating leading indicators into the modelling framework. To our  
312 knowledge this is the first study to apply time delay nowcasting methods to norovirus cases,  
313 which may be more challenging to nowcast than other infectious diseases due to high levels  
314 of underreporting, regional heterogeneity and its association with outbreaks in closed  
315 settings such as care homes, schools and hospitals [31]. Despite this, several models  
316 generated operationally useful predictions of norovirus test counts, offering a substantial  
317 improvement over using truncated data (the current standard) or a naïve seasonal baseline.  
318 However, when reporting delay data is unavailable, time series forecasting presents an  
319 adaptive alternative with good coverage and performance compared to the baseline. In  
320 contrast to previous studies, we did not find including leading indicators improved our  
321 predictions [32]. This could perhaps be explained by lower signal in the indicators  
322 considered, related to confounding effects from other winter pathogens. Finally, our  
323 findings that several models perform well with different accuracy and biases over time and  
324 day of the week suggests the potential benefit of an ensemble approach, as has been  
325 demonstrated in other contexts [12].

326 Models incorporating reporting delays consistently performed better than forecasting  
327 models that do not, showing the utility of leveraging this data when available. This improved  
328 performance is driven by reduced uncertainty when there is more complete reported data,  
329 early in the nowcast window. Among our models using reporting delays, we found that the  
330 time delay approximation method in the “GAM” scored slightly better than the more  
331 complex “epinowcast” model’s full joint distribution approach, in this application. The  
332 “epinowcast” has increased uncertainty due to modelling the reporting delay distribution  
333 and underlying process model. Wide intervals are penalised in scoring metrics like the WIS,  
334 however, this larger uncertainty may better reflect the uncertainty in the system. We saw  
335 that modelling based on recent distributions of reporting delays can perform poorly if these  
336 distributions change rapidly, although in these cases, the “epinowcast” model’s optional  
337 time-varying delay may be advantageous compared to a fixed distribution approach, such as  
338 the one in the “GAM”. Speed is key in a real-time modelling context, with some models  
339 being substantially faster than others, however, all approaches ran in a reasonable time  
340 (Table 1) for real-time inference. The computational expense of “epinowcast” compared to  
341 other models, however, was impactful during model development.

342 The performance of some models may have been limited due to the tuning approach taken.  
343 Hyperparameter optimisation was performed on a time before the epidemic wave started,  
344 simulating a plausible real-time scenario – which may bias selection toward  
345 hyperparameters good at flat periods of incidence. There are reporting changes in frontline  
346 healthcare delivery which can impact the performance of time delay informed models –  
347 these local practices are challenging to understand in real-time and adjust for in modelling,  
348 which should be explored further. Future work should explore how local testing practices  
349 can be incorporated into modelling directly. Understanding testing pathways and real-time  
350 modelling of norovirus will be crucial for the next strain replacement event highlighting the  
351 importance of developing our understanding and preparedness.

352 While not a high priority pandemic potential pathogen, norovirus causes healthcare system  
353 strain and an unpleasant infection for the individual, increasing associated opportunity cost  
354 by blocking beds and elongating patient length of stay [3]. Estimating the current case  
355 burden when accounting for delayed reporting can be an important tool for supporting  
356 effective public health response. In this work we have compared the options available to  
357 correct for delayed reporting, highlighting their strengths and limitations – notably  
358 demonstrating the importance of explicitly modelling the partially complete data. This work  
359 will underpin situational awareness should the next strain replacement event occur.

## 360 **Contributions**

361 **JM** – Conceptualisation, Methodology, Software, Validation, Formal Analysis, Data Curation,  
362 Writing – Original Draft, Writing – Review & Editing, Visualisation, Project Administration

363 **MT** - Methodology, Software, Validation, Formal Analysis, Data Curation, Writing – Original  
364 Draft, Writing – Review & Editing, Visualisation

365 **EF** - Methodology, Software, Formal Analysis, Writing – Original Draft, Writing – Review &  
366 Editing, Visualisation

367 **RC** – Conceptualisation, Data Curation, Writing – Review and Editing

368 **OP** - Software, Formal Analysis, Writing – Review & Editing

369 **CEO** – Methodology, Writing – Review & Editing

370 **AH** – Investigation, Writing – Review & Editing

371 **AD** – Conceptualisation, Investigation, Writing – Review & Editing

372 **SRD** – Conceptualisation, Writing – Review & Editing, Supervision

373 **TW** – Writing – Review & Editing, Supervision

## 374 **Ethical Approval**

375 UKHSA have an exemption under regulation 3 of section 251 of the National Health Service  
376 Act (2006) to allow identifiable patient information to be processed to diagnose, control,  
377 prevent, or recognise trends in, communicable diseases and other risks to public health.

## 378 **Conflict of Interest**

379 The authors have declared that no competing interests exist.

380

## 381 **Data Availability Statement**

382

383 Training data for the models explored in this manuscript is available at  
384 <https://github.com/jonathonmellor/norovirus-nowcast>. This data is aggregate with  
385 statistical noise added to preserve anonymity. This data enables each model to be fit and  
386 can be used for the future development of nowcasting models. Code for running all models  
387 is available at <https://github.com/jonathonmellor/norovirus-nowcast>. Individual-level data  
388 on the reporting delay used to inform initial exploration are not available due to patient  
389 identifiability. An application for data access can be made to the UK Health Security Agency.  
390 UKHSA operates a robust governance process for applying to access protected data that  
391 considers:

- 392 • the benefits and risks of how the data will be used
- 393 • compliance with policy, regulatory and ethical obligations
- 394 • data minimisation
- 395 • how the confidentiality, integrity, and availability will be maintained
- 396 • retention, archival, and disposal requirements
- 397 • best practice for protecting data, including the application of ‘privacy by design and  
398 by default’, emerging privacy conserving technologies and contractual controls

399 Access to protected data is always strictly controlled using legally binding data sharing  
400 contracts.

401 UKHSA welcomes data applications from organisations looking to use protected data for  
402 public health purposes.

403 To request an application pack or discuss a request for UKHSA data you would like to  
404 submit, contact [DataAccess@ukhsa.gov.uk](mailto:DataAccess@ukhsa.gov.uk).

405

## 406 References

407

- [1] J. Xerry, C. I. Gallimore, M. Iturriza-Gómara, D. J. Allen and J. J. Gray, "Transmission events within outbreaks of gastroenteritis determined through analysis of nucleotide sequences of the P2 domain of genogroup II noroviruses," *Journal of clinical microbiology*, vol. 46, no. 3, pp. 947-953, 2008.
- [2] S. M. Bartsch, B. A. Lopman, S. Ozawa, A. J. Hall and B. Y. Lee, "Global economic burden of norovirus gastroenteritis," *PloS one*, vol. 11, no. 4, p. e0151219, 2016.
- [3] F. G. Sandmann, L. Shallcross, N. Adams, D. J. Allen, P. G. Coen, A. Jeanes, Z. Kozlakidis, L. Larkin, F. Wurie, J. V. Robotham, M. Jit and S. R. Deeny, "Estimating the hospital burden of norovirus-associated gastroenteritis in England and its opportunity costs for nonadmitted patients," *Clinical Infectious Diseases*, vol. 67, no. 5, pp. 693-700, 2018.
- [4] K. M. O'Reilly, F. Sandman, D. Allen, C. I. Jarvis, A. Gimma, A. Douglas, L. Larkin, K. L. Wong, M. Baguelin, R. S. Baric, L. C. Lindesmith, R. A. Goldstein, J. Breuer and J. W. Edmunds, "Predicted norovirus resurgence in 2021–2022 due to the relaxation of nonpharmaceutical interventions associated with COVID-19 restrictions in England: a mathematical modeling study," *BMC Medicine*, vol. 19, pp. 1-10, 2021.
- [5] P. White, "Evolution of norovirus," *Clinical Microbiology and Infection*, vol. 20, no. 8, p. 7410745, 2014.

- [6] K. Zakikhany, D. J. Allen, D. Brown and M. Iturriza-Gómara, "Molecular evolution of GII-4 Norovirus strains," *PloS one*, vol. 7, no. 7, p. 108, 2012.
- [7] C. Ruis, S. Roy, J. R. Brown, D. J. Allen, R. A. Goldstein and J. Breuer, "The emerging GII. P16-GII. 4 Sydney 2012 norovirus lineage is circulating worldwide, arose by late-2014 and contains polymerase changes that may increase virus transmission," *PloS one*, vol. 12, no. 6, pp. 1-9, 2017.
- [8] UK Health Security Agency, "National norovirus and rotavirus surveillance reports: 2023 to 2024 season," 9 May 2024. [Online]. Available: <https://www.gov.uk/government/statistics/national-norovirus-and-rotavirus-surveillance-reports-2023-to-2024-season>.
- [9] UK Government, "The Health Protection (Notification) Regulations 2010," 2010. [Online]. Available: <https://www.legislation.gov.uk/uksi/2010/659/contents/made>.
- [10] N. Ondrikova, H. Clough, A. Douglas, R. Vivancos, M. Iturriza-Gomara, N. Cunliffe and J. P. Harris, "Comparison of statistical approaches to predicting norovirus laboratory reports before and during COVID-19: insights to inform public health surveillance," *Scientific reports*, vol. 13, no. 1, 2023.
- [11] S. Lee, E. Cho, G. Jang, S. Kim and G. Cho, "Early detection of norovirus outbreak using machine learning methods in South Korea," *PLoS One*, vol. 17, no. 11, 2022.
- [12] D. Wolfram, S. Abbott, M. An der Heiden, S. Funk, F. Günther, D. Hailer, S. Heyder, T. Hotz, J. van de Kasstele, H. Küchenhoff, S. Muller-Hansen, D. Syliqi, A. Ullrich and M. Weigert, "Collaborative nowcasting of COVID-19 hospitalization incidences in Germany," *PLOS Computational Biology*, vol. 19, no. 8, 2023.
- [13] J. T. Wu, K. Leung, T. T. Lam, M. Y. Ni, C. K. Wong, J. M. Peiris and G. M. Leung, "Nowcasting epidemics of novel pathogens: lessons from COVID-19," *Nature Medicine*, vol. 27, no. 3, pp. 388-395, 2021.
- [14] UK Health Security Agency, "Guidance: Notifiable diseases and causative organisms: how to report," 1 January 2024. [Online]. Available: <https://www.gov.uk/guidance/notifiable-diseases-and-causative-organisms-how-to-report>.
- [15] UK Health Security Agency, "Laboratory reporting to UKHSA, A guide for diagnostic laboratories," May 2023. [Online]. Available: [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/1159953/UKHSA\\_Laboratory\\_reporting\\_guidelines\\_May\\_2023.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1159953/UKHSA_Laboratory_reporting_guidelines_May_2023.pdf).
- [16] S. F. McGough, M. A. Johansson, M. Lipsitch and N. A. Menzies, "Nowcasting by Bayesian Smoothing: A flexible, generalizable model for real-time epidemic tracking," *PLoS computational biology*, vol. 16, no. 4, p. e1007735, 2020.
- [17] NHS, "111 online, Get help for your symptoms," [Online]. Available: <https://111.nhs.uk/>.
- [18] C. E. Overton, S. Abbott, R. Christie, F. Cumming, J. Day, O. Jones, R. Paton, C. Turner and T. Ward, "Nowcasting the 2022 mpox outbreak in England," *PLoS computational biology*, vol. 19,

no. 9, p. e1011463, 2023.

- [19] J. van de Kasstelee, P. H. Eilers and J. Wallinga, "Nowcasting the number of new symptomatic cases during infectious disease outbreaks using constrained P-spline smoothing," *Epidemiology*, vol. 30, no. 5, pp. 737-745, 2019.
- [20] S. Wood, "Package `mgcv`," *R package version*, vol. 1, no. 29, p. 729, 2015.
- [21] G. L. Simpson, "gratia: graceful ggplot-based graphics and other functions for GAMs fitted using mgcv," 2024. [Online]. Available: <https://gavinsimpson.github.io/gratia/>.
- [22] S. Abbot, A. Lison, S. Funk, C. Pearson, H. Gruson, F. Guenther and M. DeWitt, *epinowcast: Flexible Hierarchical Nowcasting*, 10.5281/zenodo.5637165.
- [23] M. Höhle and M. an der Heiden, "Bayesian nowcasting during the STEC O104: H4 outbreak in Germany, 2011," *Biometrics*, vol. 70, no. 4, pp. 993-1002, 2014.
- [24] F. Günther, A. Bender, K. Katz, H. Küchenhoff and M. Höhle, "Nowcasting the COVID-19 pandemic in Bavaria," *Biometrical Journal*, vol. 63, no. 3, pp. 490-502, 2021.
- [25] B. Carpenter, A. Gelman, M. D. Hoffman, D. Lee, B. Goodrich, M. Betancourt, M. A. Brubaker, J. Guo and P. Li, "Stan: A probabilistic programming language," *Journal of statistical software*, vol. 76, 2017.
- [26] S. L. Scott and H. R. Varian, "Predicting the present with Bayesian structural time series," *International Journal of Mathematical Modelling and Numerical Optimisation*, vol. 5, no. 1-2, pp. 4-23, 2014.
- [27] S. L. Scott, *R Package 'bsts'*, 2016.
- [28] H. Ishwaran and J. S. Rao, "Spike and slab variable selection: Frequentist and Bayesian strategies," *The Annals of Statistics*, vol. 33, no. 2, pp. 730-773, 2005.
- [29] N. I. Bosse, H. Gruson, A. Cori, E. van Leeuwen, S. Funk and S. Abbott, "Evaluating forecasts with scoringutils in R," 2022. [Online]. Available: <https://arxiv.org/abs/2205.07090>.
- [30] K. Charniga, Z. J. Madewell, N. B. Masters, J. Asher, Y. Nakazawa and I. H. Spicknall, "Nowcasting and Forecasting the 2022 US Mpox Outbreak: Support for Public Health Decision Making and Lessons Learned," *Epidemics*, vol. 47, no. 1755-4365, p. 100755, 2024.
- [31] N. Ondrikova, H. Clough, N. Cunliffe, M. Iturriza-Gomara, R. Vivancos and J. Harris, "Understanding norovirus reporting patterns in England: a mixed model approach," *BMC Public Health*, vol. 21, pp. 1-9, 2021.
- [32] F. Bergström, F. Günther, M. Höhle and T. Britton, "Bayesian nowcasting with leading indicators applied to COVID-19 fatalities in Sweden," *PLOS Computational Biology*, vol. 18, no. 12, p. e1010767, 2-22.

408

409