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

Authors: Jonathon Mellor* , Maria L Tang*, Emilie Finch**, Rachel Christie*, Oliver Polhill*,
Christepher E Querten^{1,3}, Ann Ueben⁴, Amy Deurles^{4,} Sereb B Deepu¹, Themas Ward¹ Christopher E Overton^{1,}", Ann Hoban⁺, Amy Douglas⁺, Sarah R Deeny⁺, Thomas Ward⁺

- 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
- 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

Abstract

Background

Methods

Results

Abstract
Backgrou
Noroviru:
Diagnosti
time surv
Methods
To nowcz
mechanis
model ino
Nesults
Modellin_;
harnesse
techniqui
be indica
surveillar
be indica
Interpret
Analysis (understa
making. 1
the repor
producers
producers no to mean the real-time to acute the model of acute gastro in the model of the model of acute gastro in the real-time system of acute gastro in the real-time system of acute gastro in the real-time system of acute gas pro Methods
To nowcast the read
mechanistic Bayes
model including sy
over weekly nowcare
Results
Modelling current
harnessed a time (techniques. Hower
of temporally chare
surveillance data v
be indicative poor
Interpretation
A model including syndromic surveillance data were developed. These models were evaluative weekly nowcasts using a probabilistic scoring framework.
Nessults
inverselvely nowcasts using a probabilistic scoring framework.

Mes over weekly nowcasts using a probabilistic scoring framework.

Results

Modelling current cases clearly outperforms a simple heuristic approach. Models that

Modelling current cases clearly outperforms a simple heuristic a **Example 12**

Roadelling current cases clearly outperforms a simple heuristic

harmessed a time delay correction had higher skill, overall, relation

techniques. However, forecasting approaches were found to be

surveillan Internation at the delay correction had higher skill, overall, relative to forecasting the chiniques. However, forecasting approaches were found to be more reliable in the evolution of the more interaction of norovirus syn

Interpretation

techniques. However, forecasting approaches were found to be more reliable in to fermporally changeable reporting patterns. The incorporation of norovirus syndsurveillance data was not shown to improve model skill in this 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 ind surveillance data was not shown to improve model skill in this nowcasting task, which m
be indicative poor correlation between the indicator and norovirus incidence.

Interpretation

Analysis of surveillance data enhanced Surface indicative poor correlation between the indicator and norovirus incidence.

Surface indicative poor correlation between the indicator and norovirus incidence.

Analysis of surveillance data enhanced by nowcasting r Franklysis of surveillance data enhanced by nowcasting reporting delays improver
Analysis of surveillance data enhanced by nowcasting reporting delays improver
understanding over simple model assumptions, which is importan understanding over simple model assumptions, which is important for real-time making. The structure of the modelling approach needs to be informed by the pathe reporting delay and can have large impacts on operational perf 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
TE: This preprint reports new resea making. The structure of the modelling. The structure of the modelling approximate of the modelling delay and can have large impacts on operational performance and insights produced.
Produced.
TE: This preprint reports new produced.
1
TE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical

$\mathbf{1}$ **Introduction**

-
-
-
-
-
-
-
-

-
-
-
-
-
-
- 4 momentum considerate controllation and nurser in the system of the same on the same of a pandemic response, followed by resurrent streament person of the SARS CoV partents resumed to pre-pandemic levels [4]. Norovrus is 6 which pays the material of the second of the second server in the SARS-CoV-C and the second of the SARS-CoV-C is a constant we consider the pays of the station of the pays of the station of the station of the station of e 2 pandemic response, binduced by response a parameter and the second mixing pathogen and the station of the figlicating to station response to the station of the matrix policing in the time line is the matrix policing in
-
-
-
-
-
-
- 1 Introduction

2 Norovirus is a

3 diarrhoea. No

4 bomes and nu

15 winter [2] [3].

2 pandemic rr

patterns resu

16 transmission,

insights for ta

10 transmission,

insights for ta

10 Morovirus sur

These include

un 2 Norovirus is a gaso uncertainty in a substantial bureau and mean substantial and the mass and musselesting symptoms or moves and the mass and the mass mind and the mass of the SA virus in the mass mind and the system, an 3 ownered into the term of the methods in encoses acceleated in the same of the same of the same of the same in the same of the same that 9 results and network in the same of multiplarity. The control of the minimum is the specifical state in the signists for taking public health action.

2 Transmission, highlighting the importance of monitoring and improvin 11 Insights for taking public incelatinatedon.

12 Norovirus surveillance in England uses

13 These include norovirus positive laboration

14 Indergo molecular typing, as well as no

15 delay between diagnostic test admini 14 undergo molecular typing, as well as the more of the molecular controls of the molecular delay between diagnostic test administration and reporting to the motional survellance
the data, particuly attributable to norwism 26 each vector owing solar test administration and reporting to the ministric diagnostic term and reports truncal chata, partially attributable to norrolnts on beling a Schedule 2 not his hold causative agen time period sh 18 Norovirus san excellent candidate for the application of nowcasting moment delay in case reporting as a non-priority pathogen. Research has incredibited for the application of noverating momentation of normation of corr 22 on short term projections using somation intentions [12] principal methods and national methods intentions cases. Notowirus incidence is his stochastic, with a partially seasonal pattern and high heterogeneity between l 23 stochastic, with purchast year of the real-time may immediate material control of the real-time in the SARS-CoV-2 pandemic [12] [13] modelling on nowcasting
125 research applied during the SARS-CoV-2 pandemic [12] [13]
-
-
-
-
-
-

31 Methods

27 In this paper, we explored the reporting delay for norovirus a

26 2023/2024 winter. We evaluated a range of methods for now

28 2023/2024 winter. We evaluated a range of methods for now

37 structures, guide signals, a 27 In this paper, we explore the reporting extracted for norontal extension and the reporting suing different m
27 2023/2024 winter. We evaluated a range of methods for noroxasting using different m
27 structures, guide si 31 **Methods**

31 **Norovirus Cases**

33 **Data**

33 **Norovirus Cases**

34 **Individual test results were extracted from the Second Generation Surve

35 (SGSS) database in UKHSA (UK Health Security Agency) [14] for positive

3** 33
33
34 Indivi
35 (SGSS
36 condu
38 and s
39 episo
2 Norovirus cases
34 Individual test results we
35 (SGSS) database in UKHS
36 conducted in England. T
37 results uploaded by fron
38 and social care settings.
89 episode. Under the legis
2

-
-
- 7 patterns resume to the parameteric state in the several state in the seve eration for the might and shift 11 Institute the state in the state of monitoring the importance of monitoring the state increase in the state increase of the state increase of which state in the state increase of which state increase of which state incr 12 Norovirus surveillance in including an well as a reporting the surveillance in the surveillance in England unit

13 These include norovirus positive laboratory reports from confirmed cases, of which a surveillance in Eq 19 Norovirus Interaction convertion and proposition in the application in the application of the application 22 exploration of correcting for time delays in norovirus deach. Norovirus cases, the most simply a stochastic, with a partially seasonal pattern and high heterogeneity between localised
24 outbreaks and national trends, m reactions and notional tentors, memberging to procett. Columnic or reaction applied during the SARS-CoV-2 pandemic [12] [13] modelling can be explored timprove understanding of the real-time norovirus dynamics.

27 In this 25 research applied during to social control applied the real-time incovirus dynamics.

27 In this paper, we explored the reporting delay for norovirus cases in England over the

2023/2024 winter. We evaluated a range of m 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 on 33 conducted in England. The database only stores information on positive information to conducted in England. The database only stores information on positive laboratory test
33 conducted in England. The database only st 36 conducted in England. The database only stores information on positive laboratory test
results uploaded by frontline diagnostic laboratories, with a sampling bias towards healt
and social care settings. We deduplicated
-
- 14 These include norothomology in the matter of the matter of which a subset of which a subset of the matter distingent in the matter of t or or the period shown by the method state. Those to norovirus and the method of being the national difficial stats. Surveillance reports truncate the time period shown by one week to remove partially complete data [8].
 17 legislation [9]. Due to this legislation for the nation of the nation of the national of th communication of orientation and priority as a non-priority pathogen. In constraints of the simulation of correcting for time delays in norovirus cases. Norovirus incidence is highly stochastic, with a partially seasonal p 28 2022 extractions and assumptions about data completeness to consider the trade-
28 structures, guide signals, and assumptions about data completeness to consider the trade-
28 offs between different approaches applicabl 29 structures, guide signals, and assumptions about data completeness to consider the trade-
29 offs between different approaches applicable to norovirus and beyond.
29 Data
29 Data
29 Data
29 Norovirus Cases
29 Individual 37 results uploaded by from the diagnostic laboratories, with a sampling bias towards frequent and social care settings. We deduplicated tests to keep the first test per patient infection episode. Under the legislation pos 23 and social care settings. We deduplicated tests to keep the first test per patient infection
approach episode. Under the legislation positive norovirus diagnostic tests are required to be notifi-
2 39 episode. Under the legislation positive norovirus diagnostic tests are required to be notified

40 to the UKHSA and the UKHSA, but not required with the UKHSA, the spectrum of the UKHSA, the spectrum of the UKHSA 41 periodicity shown in Supplementary Figure 1.

42 In this analysis we focused on two main time the section of the Secondy, the report date t_r defines the date the Secondy, the report date t_r defines the date is the 42 In the conduction of the species of the specific method of the specific specifical from the indicted individual for testing.

42 In defines when the specifies the coronius case through motional surveillance. As symptom 43 commutes when the year of the specific model in the specific model in the specific model in the specific discondity in the specific disconding the model of the model of the specific disconding the inference of the spec notifying UKHSA of the noro
dates are not reported, the
dates are not reported, the
Despite being impacted by t
representation of the epider
each test. The difference be
reporting delay.
To model the epidemic and
and *d* 44 states are not to produce y when the record of the system with the specified of the record is defined into the specified, the specified into t 45 notify walked with the tots of the normal surveillance. The symptom of the symptom 46 does be the the epidemiological many experiences who we are not reported to the most presentation of the epidemic date are not reported to the epidemic and correspond compared with other walsolve ten event. The differe 47
48 Depresentation of the epidemic's progression compared with other available time events
not test The difference between report date and specimen date $d = t_r - t$ is the
reporting delay.
To model the epidemic and corresp 48 representation of the epidemic and comparison of the epidemic of the epidemic and specified between the epidemic and corresponding delay distributions, we aggregated the data by to roodel the epidemic and corresponding reporting delay.

To model the epidemic and corresponding delay distributions, we and

and d to construct a so-called data "reporting triangle" [16], illustrate

reporting triangle's an array with elements $n_{t,d}$, for te 49 each test. The difference between report date and specimen date $d = t_r - t$ is the 50 reporting delay. ited the day
ited the day
Figure 1. 1
In reportin
of the spectric alysis, we
e a shorte
ght trunca
on making
eporting t
ese daily of 51 To model the ep

52 and d to construe

reporting triangl

53 reporting triangl

55 The element $n_{t,a}$

55 The element $n_{t,a}$

56 date time series

57 in practice most

58 In real-time, case

59 In real-time, case

59

-
-
- is the
elay.
en
rever,
sume
alue.
n.
d
d
ints
-
-
-
-
-
-
- Therefore, for cases at $t = T$ only cases with $d = 0$ can be known, with other values
- 61 $1 \le d \le D$ 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
-
- 52 To model the epidemic was delay delay and corresponding to the $\epsilon = [1, 1]$ and $d = [0, 0]$, where $\epsilon = T$ is the epidemic and α is the epidemic and correcting the anary with elements, n_{ϵ} , for $\epsilon \in [1, T]$ and $d \in$ 52 and a report maxin The el date t in pranamax In rea
55 56 date t in pranamax In rea
59 There $1 \leq d$ proac theref are all 62 proac theref are all 3 ting triangle is an array with elements $n_{\rm r,d}$, for $t \in [1, T]$ and $d \in [0, D]$, where T is
numeleagh of the speciene data their series t and D is the maximum reporting de-
nement $n_{\rm rf}$ are presents the number of Starting transition of the specifiers that there is an array with centerion n_{eff} , or $v = (r_1, r_2)$, the element n_{ref} represents the number of tests collected or date time series t and D is the secifier days. In the maintimum of the t^{th}
) could the t^{th}
 ≥ 0 and ≥ 0 and t^{new} and ≥ 0 and N_t .
In all need in the ≥ 0 and ≥ 0 a From the properties of the specim
be very large. How
this analysis, we ass
ssume a shorter v:
s a right truncatio
ecision making an
The reporting trian
ds, these daily counts, these daily counts. 54 maximum length of this packimen date time series t and D is the maximum reporting delay.
The element n_{A} represents the number of tests collected on the tth day of the specimen date time series that were report date time series that were reported after *d* days. In theory, *D* could
the remest reporting delays are under 10 days. Therefore, for
a maximum possible reported delay of 50, though each model may
an inreal-time, cases 55 The element $n_{i,j}$ regresents the number of tests collected on the i^{th} day of the speciment $n_{i,j}$ required to the speciment $n_{i,j}$ for a data time series that were reported after d days. In theory, D could b So the series that were reported atter to provide a maximum possible reported delay of 50, the real-time, cases $n_{t,d}$ cannot exist when d Therefore, for cases at $t = T$ only cases with $1 \le d \le D$ unknown. The quantity of 10 days. Therefore, for this analysis, we assume

hough each model may assume a shorter value.
 $> T - t$ which introduces a right truncation.
 $h d = 0$ can be known, with other values
 $h d = 0$ inform decision making and

tie **57** in protocol notice most reporting schedular products and the most product in the Tatelline, cases n_{c} are alternations are product that $d = 0$ can be known, with other values.
Therefore, for cases n_{c} acc SB announce to provide the reported delay of the model may be the model model in the delay and the forcates at $t = T$ only cases with $d = D$ can be known, with other values of 56 . The forcates of $t = T$ only cases with \frac Therefore, for cases $v_{t,d}$ admitive cannot exist when α
50 Therefore, for cases at $t = T$ only cases with $1 \leq d \leq D$ unknown. The quantity of most
52 proactive communications was the total cas
53 therefore collapsed \ddot{v} 0 can be known, with other values
est used to inform decision making and
specimen date N_t . The reporting triangl
ort operational needs, these daily counts
therpretation. 61 $1 \le d \le D$ unknown. The question of $1 \le d \le D$ unknown. The question of 62 proactive communications we therefore collapsed into N_t are also aggregated to week! in antity of most internal cases
as the total cases
= $\sum_{d=0}^D n_{t,d}.$ To sup
y levels for ease of
 α rest used to inform decision making a
by specimen date N_t . The reporting tripport operational needs, these daily contempretation. we communications was the total cases by specimen date N_t . The reporting triang
ore collapsed into $N_t = \sum_{d=0}^n n_{t,d}$. To support operational needs, these daily coun
o aggregated to weekly levels for ease of interpreta provide the total product of $N_t = \sum_{d=0}^{n} n_{\text{f},d}$. To support operational needs, these daily counts are also aggregated to weekly levels for ease of interpretation. 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.
- therefore contains the collapsed into $\frac{u}{u} = 0.1$, u
are also aggregated to weekly levels for 64 are also aggregated to weekly levels for ease of interpretation.

65
66

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 spec 68 sum of each row across the reporting delays. Each cell represents the case count for a given specimen date and reporting 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. 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 Trapidly available. These data were be leveraged to inform case prediction. NHS 111 Online

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

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

-
-

91 approach to compare against. We assumed each predicted day will be equal to the o
92 count the previous week giving an autocorrelated prediction with day-of-week effect
93 The central estimate is set as $\overline{N}_t = \sum_{d=0$ 92 count the previous week giving an autocorrelated prediction with day-of-week effects.

93 The central estimate is set as $\overline{N}_t = \sum_{d=0}^{T-(t-7)} n_{t-7,d}$, which corresponds to the reported data

94 from the seven days p The central estimate is set as $\overline{N}_t = \sum_{d=0}^{T-(t-7)} n_{t-7,d}$, which corresponds to the reported
from the seven days prior – matching the weekly reporting cycle in surveillance. Most
norovirus cases were reported with d The central estimate is set as $N_f = \sum_{d=0}^{n} n_{t-7,d}$, which corresponds to the reported data

from the seven days prior – matching the weekly reporting cycle in surveillance. Most

norovirus cases were reported with $d \le$ $\lim_{d=0}^{t-(t-7)} n_{t-7,d}$

- 94 From the seven days prior matering the weekly reporting cycle in surveillance. Most
norovirus cases were reported with $d \leq 7$ and as such this method gives predictions o
complete case numbers. We did not consider u
-

France Thorovirus cases were reported with a
96 complete case numbers. We did not complication of the scoring methodolog
98 baseline model the prediction interva
99 **Generalised Additive Model**
00 We used a generalised add This are calculated within the baseline method. For
y, prediction intervals are required. Therefore, for the
s were assumed equal to the central estimate.
(GAM) utilising partially reported data, based on a
This estimated 97 application of the scoring methodology, prediction intervals are required. Therefore, for th
98 baseline model the prediction intervals were assumed equal to the central estimate.
99 **Generalised Additive Model**
99 **We** 98 baseline model the prediction intervals were assumed equal to the central estimate.

99 **Generalised Additive Model**

99 **Generalised Additive Model**

99 **Generalised Additive Model**

99 **We used a generalised additive** 99 **Saseline model the prediction intervals were assumed equal to the central estimate.**
99 **Generalised Additive Model**
90 We used a generalised additive model (GAM) utilising partially reported data, based a
92 nowcasti 999 Generalised Additive Model

99 Generalised additive model

99 Specimen date t, \overline{N}_t , as the sum of k

99 reporting delays $d \in [0, T - t]$, and

94 $\overline{n}_{t,d}$, for reporting delays $d \in [T - t]$ 101 nowcasting model for mpox [18] [19]. This estimated the total number of cases with

102 specimen date t, \overline{N}_t , as the sum of known data that has already been reported, $n_{t,d}$, for

103 reporting delays $d \in [0, T -$ 101 nowcasting model for mpox [18] [19]. This estimated the total number of cases with

102 specimen date t , \overline{N}_t , as the sum of known data that has already been reported, $n_{t,d}$,

103 reporting delays $d \in [0, T - t]$ specimen date t , n_t , as the sum of known data that has already been reported, $n_{t,d}$, for reporting delays $d \in [T-t+1, D]$, i.e.

103 reporting delays $d \in [T-t+1, D]$, i.e.

105 $\overline{N}_t = \sum_{d=0}^{T-t} n_{t,d} + \sum_{d=T-t+1}^{D} \overline{n$ ے۔
delavs ہ

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

 $d = 0$

103 reporting delays $d \in [0, 1 - t]$, and estimates for the drikhown data yet to be reported,

104 $\bar{n}_{t,d}$, for reporting delays $d \in [T - t + 1, D]$, i.e.

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

106 As $n_{t,d}$ is 104 $n_{t,d}$, for reporting delays $a \in [1 - t + 1, D]$, i.e.

105 $\overline{N}_t = \sum_{d=0}^{T-t} n_{t,d} + \sum_{d=0}^{T-t} n_{t,d}$

106 As $n_{t,d}$ is known, \overline{N}_t has a natural lower bound of

107 with a negative binomial distribution accounti

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

with a negative binomial distribution accounting for the non-negative integer values and
overdispersion. Using the mean and variance parameterisation,
 $n_{t,d} \sim \text{NegBin}(\mu_{t,d}, \mu_{t,d} + \mu_{t,d}^2 / k)$
with dispersion parameter k. W 107 **a** $n_{t,d}$ is known, n_t has a hatdal lower bound or $\Delta d = 0$ $n_{t,d}$. The dikknown data was modelled
107 with a negative binomial distribution accounting for the non-negative integer values and
108 overdispersion. 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 epidem

110 process, where 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 mode

110 process, where $\mu_{t,d}$ depends on both

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

where β_0 is a constant. We assumed that the number of cases vary smoothly over specimen
date t and number of days delay d as $s_1(t)$ and $s_2(d)$, with random day-of-week effects 110 process, where $\mu_{t,d}$ depends on both t and d according to

110 process, where $\mu_{t,d}$ depends on both t and d according to

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

112 where β_0 is a consta 110 process, where $\mu_{t,d}$ depends on both t and a

111 log($\mu_{t,d}$) = β_0 + $s_1(t)$ + $s_2(d)$ +

112 where β_0 is a constant. We assumed that the

113 date *t* and number of days delay *d* as $s_1(t)$ a

114 ω_1 ω_1 (wday(t)

number of ca

nd s_2 (d), with

vely. The mod

urn-in and po)).
'er
ek e
mg
''mp
se : date t an
 ω_1 (wday

function if

the mode

aggregate

Models w

for s. (t) where p_0 is a constant. We assumed that the number or cases vary smoothly over speciment

date t and number of days delay d as $s_1(t)$ and $s_2(d)$, with random day-of-week effects
 $\omega_1(\text{wday}(t))$ and $\omega_2(\text{wday}(t + d))$ r date *t* and number of days delay *d* as $s_1(t)$ and $s_2(d)$, with random day-of-week effects $\omega_1(\text{wday}(t))$ and $\omega_2(\text{wday}(t + d))$ respectively. The model was fitted in *R* using the *ga*, function from the *mgcv* package 113 date *t* and number of days delay *d* as $s_1(t)$ and $s_2(d$

114 $\omega_1(\text{wday}(t))$ and $\omega_2(\text{wday}(t + d))$ respectively. Th

115 function from the *mgcv* package [20]. 1000 burn-in a

116 the model using the *gratia* packag 114 $\omega_1(\text{wday}(t))$ and $\omega_2(\text{wday}(t + d$

115 function from the *mgcv* package [2

116 the model using the *gratia* package

117 aggregated to \overline{N}_t (eqn. 1), with pre

118 Models were fit to the past 56 day

119 for s)) respectively. The model was fitted in *R* using the *gam*
'0]. 1000 burn-in and posterior samples were drawn from
'e [21] with a Metropolis-Hastings sampler. Samples were
diction intervals taken using quantiles of thes 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 \overline{N}_t (e 117 aggregated to \overline{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 $s_1(t)$ an aggregated to N_f (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

120 Supplementary Section 2 119 Models were fit to the past 56 days, with cable regression basis functions every $t = 7$ 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 119 for $s_1(t)$ and $s_2(d)$

120 Supplementary Se

121 **Epinowcas**

122 We also used a Ba

123 package [22], with

124 nowcasting approa), and a maximum reporting delay $D = 14$. Model tuning is outlined in
ction 2.
t
yesian hierarchical nowcasting framework implemented in the *epinowc*
the implementation described below. This approach builds on earlier

- 120 Supplementary Section 2.

121 **Epinowcast**

122 We also used a Bayesian h

123 package [22], with the imp

124 nowcasting approaches [2 121 Epinowcast

122 We also used a Bay

123 package [22], with t

124 nowcasting approac

5 we also used a Bayesian inerarchical nowcasting maniework implemented in the epinowcast
package [22], with the implementation described below. This approach builds on earlier
nowcasting approaches [23] [24]. As with the "G
- package $[22]$, with the implementation described below. This approach builds on earlier
nowcasting approaches $[23]$ $[24]$. As with the "GAM" model, the estimate for the total
5 124 nowcasting approaches $[25]$ $[24]$. As with the "GAM" model, the estimate for the total 5

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 dif Here n_t and is a multimomial distribution with a probability vector $(\mu_t$ and is

a estimated jointly with the expected number of final reported cases. This differs from t

"GAM" model approach, where each $n_{t,d}$ is in

-
- The unknown data, $\bar{n}_{t,d}$ (eqn. 1).

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 n 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 fin 129 GAM model approach, where each $h_{t,d}$ is independent. We used the default

130 implementation of modelling expected final reported cases as a first order rand
 $E[N_t] = \lambda_t$
 $log(\lambda_t) \sim \text{Normal}(log(\lambda_{t-1}, \sigma^{\lambda}))$
 $log(\lambda_0) \sim \text{Normal}($

$$
E[N_t] = \lambda_t
$$

\n
$$
\log(\lambda_t) \sim \text{Normal}(\log(\lambda_{t-1}, \sigma^2))
$$

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

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

\n
$$
\text{er}_t \text{ is defined as the log of the}
$$

131

\n132

\n133

\nThe instantaneous growth rate
$$
r_t
$$
 is defined as the log of the expected number of final reported (log(λ_t) ~ Normal(log($N_0 + 1$), 1)

\n134

\n135

\n136

\n137

\n138

\n139

\n130

\n131

\n132

\n133

\n134

\n135

\n136

\n137

\n138

\n139

\n130

\n131

\n132

\n133

\n134

\n135

\n136

\n137

\n138

\n139

\n130

\n131

\n132

\n133

\n134

\n135

\n136

\n137

\n138

\n139

\n130

\n131

\n132

\n133

\n134

\n135

\n136

\n137

\n138

\n139

\n130

\n131

\n132

\n133

\n134

\n135

\n136

\n137

\n138

\n139

\n130

\n131

\n132

\n133

\n134

\n135

\n136

\n137

\n138

\n139

\n130

\n131

\n130

\n131

\n132

\n133

\n134

\n135

\n136

\n137

\

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

 $+ 1$),1)

).
 \vdots the exp

elled on

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

133 The instantaneous growth rate r_t is defined as the log of th

134 reported tests between time *t* and *t* – 1. r_t is then modell

135 random effect $\omega_1(t)$ and a random effect for the d

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

The instantaneous growth rate r_t is defined as the log of the expected number of final

134 is 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 reported tests between time t and t \mathbf{r} . t_f is then modelled on the log scale by a daily

random effect $\omega_1(t)$ and a random effect for the day of the week $\omega_2(\text{wday}(t))$, to acco

for weekly periodicity in the un 135 andom 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 distribu 136 For weekly periodicity in the underlying data.
 $log(r_t) = \omega_1(t)$

137 Within *epinowcast* the delay distribution $(p_{t,c}$

138 model where:
 $h_{t,d} = P(\text{delay} =$ $\begin{aligned} &\log(r_t) = \omega_1(t) + \omega_2(\mathrm{wday}(t)) \ &\text{distribution}\left(p_{t,d}\right) \text{is then defined} \ &\text{if} \ &\text{if$ 137 Within epinowcast the delay distribution (pt_t, d) is then defined as a discrete time hazard

138 model where:

139 $h_{t,d} = P(\text{delay} = d | \text{day} \ge d, \text{W}_{t,d})$.

140 Here, the hazard is determined by a design matrix W_{t,d

139
$$
h_{t,d} = P\left(\text{delay} = d \,|\text{day} \ge d, \text{W}_{t,d}\right)
$$

138 model where:
139
140 Here, the haza
141 distribution an
143 the log mean a real
cluding
affect
sed log-
th a dail 140 Here, the hazard is determined by a designmatrix $w_{t,d}$ including a baseline delay

141 distribution and time- and delay- specific covariates which affect the reporting de

142 assume the probability of reporting $p_{$ 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 when

143 the log mean and log the log mean and log standard deviation are modelled with a daily random effect (the model
default).
 $p_{t,d}^i$ ~LogNormal (μ_t, ν_t)
where the parametric logit hazard γ_{td} is given by 142 assume the probability of reporting $p_{t,d}^t$ follows a discretised log-normal distribution where

143 the log mean and log standard deviation are modelled with a daily random effect (the mode

144 default).

145 wher 143 the log mean and log standard deviation are modelled with a daily random effect (the model

144 default).

145 where the parametric logit hazard γ_{td} is given by
 $\gamma_{td} = \logit\left(\frac{p'_{t,d}}{1 - \frac{1}{2}}\right)$

1 i,u $p_{t,d}^i$ ~LogNormal (μ_t, ν_t)

144 default).
145 where th

$$
p_{t,d}^{1} \sim \text{LogNormal}(\mu_{t}, \nu_{t})
$$
\n145 where the parametric logit hazard γ_{td} is given by

\n
$$
\gamma_{td} = \logit\left(\frac{p_{t,d}^{t}}{\left(1 - \sum_{d'=0}^{d-1} p_{t,d'}^{t}\right)}\right)
$$
\n146 We also use a constant non-parametric logit hazard such that:

\n
$$
\epsilon_{t,d} = \beta_{0}
$$
\nThe overall hazard is then modelled as $\logit(h_{t,d}) = \gamma_{t,d} + \epsilon_{t,d}$

\nTo estimate final observed reported cases a negative binomial.

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

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 $logit(h_{t,d}) = \gamma_{t,d} + \epsilon_{t,d}$

148 To estimate final observed reported cases a negative binomial

149

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

148 To estimate final observed reported cases a negative binomial c

where:
 $\bar{n}_{t,d} | \lambda_t, p_{t,d} \sim NegBin(\lambda_t \times p_{t,d}, \varphi), t =$

6

 $\frac{1}{2}$ where:

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

\nTo estimate final observed reported cases a negative binomial observation model is used
\nwhere:
\n
$$
\bar{n}_{t,d} |\lambda_t, p_{t,d} \sim NegBin(\lambda_t \times p_{t,d}, \phi), t = 1, ... T
$$

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

150 and the overdispersion parameter ϕ 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, d
153 distributions f Unlike the "GAM" model, this approach introduces parametric, discrete, and truncated
distributions for the reporting delay, better reflecting the reporting measurements. Models
are fit in *stan* with c*mdstan* [25] using t 151 and N_t is given by (eqn. 1).

152 Unlike the "GAM" model, the distributions for the report

154 are fit in *stan* with *cmdstan*

155 U-Turn Sampler). We ran 10

156 maximum reporting delay of

157 and prior specifi 152 Unike the "GAM" model, this approach introduces parametric, discrete, and truncated
153 distributions for the reporting delay, better reflecting the reporting measurements. Mo
154 are fit in *stan* with *cmdstan* [25] distributions for the reporting delay, better reflecting the reporting measurements. Models

are fit in *stan* with *cmdstan* [25] using the Hamiltonian Monte Carlo (HMC) with NUTS (No-

U-Turn Sampler). We ran 1000 iterat are fit in stan with cmdstan [25] using the Hamiltonian Worke Carlo (HMC) with NOTS (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, wit

158 Bayesian Structural Time Series

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 out 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 We employed a flexible Bayesian structural time ser 157 and prior specification are outlined in Supplementary Section 3.

158 **Bayesian Structural Time Series**

160 produce a nowcast without harnessing partial reported case cou

161 truncated by 7 days, with the unknown dai 159 We employed a flexible Bayesian structural time series (BSTS) inodemig approach to
160 produce a nowcast without harnessing partial reported case counts. The time series *I*
161 truncated by 7 days, with the unknown da produce a nowcast without harnessing partial reported case counts. The time series N_t is

truncated by 7 days, with the unknown daily counts estimated in a traditional forecasting

approach. The BSTS allows for a state 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 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 t
164 models using the *bsts* R package [27 263 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 onlin

165 indicators.

166 The first 164 models using the bass 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 m

165 mulcators.
166 The first moment
168 capture the
169 The equation 167 We have mean μ_t , slope δ_t and seasonal component τ_t , with a season as $S = 7$ days to

168 capture the day-of-week effects.

169 $log(\lambda_t) = \mu_t + \tau_t$ where $N_t \sim Poisson(\lambda_t)$

170 The equation for the mean μ_t is given

168 capture the day-of-week effects.
\n169
$$
\log(\lambda_t) = \mu_t + \tau_t
$$
 where $N_t \sim Poisson(\lambda_t)$ (2)
\n170 The equation for the mean μ_t is given by
\n
$$
\mu_{t+1} = \mu_t + \delta_t + \eta_{0,t}
$$
 with $\eta_{0,t} \sim \mathcal{N}(0, \sigma_\mu)$ (3)
\n172 and the slope, (4)

169
\n
$$
\log(\lambda_t) = \mu_t + \tau_t \quad \text{where } N_t \sim Poisson(\lambda_t) \tag{2}
$$
\n170 The equation for the mean μ_t is given by\n
$$
\mu_{t+1} = \mu_t + \delta_t + \eta_{0,t} \text{ with } \eta_{0,t} \sim \mathcal{N}(0, \sigma_\mu) \tag{3}
$$
\n172 and the slope,\n
$$
\delta_{t+1} = \delta_t + \eta_{1,t} \quad \text{and} \quad \eta_{1,t} \sim \mathcal{N}(0, \sigma_\delta). \tag{4}
$$
\n1731
\n174 Lastly the seasonality component is determined via dummy regression variables,

171
$$
\mu_{t+1} = \mu_t + \delta_t + \eta_{0,t} \text{ with } \eta_{0,t} \sim \mathcal{N}(0, \sigma_{\mu})
$$
(3)
172 and the slope,
173
$$
\delta_{t+1} = \delta_t + \eta_{1,t} \text{ and } \eta_{1,t} \sim \mathcal{N}(0, \sigma_{\delta}).
$$
(4)
174 Lastly the seasonality component is determined via dummy regression variables,

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

Lastly the seasonality component is determined via dummy regression variables,

173
$$
\delta_{t+1} = \delta_t + \eta_{1,t} \text{ and } \eta_{1,t} \sim \mathcal{N}(0, \sigma_{\delta}).
$$
\n(4)
\n174 Lastly the seasonality component is determined via dummy regression variables,
\n
$$
\tau_{t+1} = -\sum_{s=1}^{S-1} \tau_{t-s+1} + \eta_{2,t}
$$
\n175 with $\eta_{2,t} \sim \mathcal{N}(0, \sigma_{\tau}).$ (5)
\n176 This ensures that the seasonal component τ_t accounts for the cumulative seasonal effects over the specified period *S* in our case one week Therefore log(λ_t) follows a local linear

175 with $\eta_{2,t} \sim \mathcal{N}(0, \sigma_{\tau})$. (5)
176 This ensures that the seasonal component τ_t accounts for the cumulative seasonal effects
177 over the specified period *S*, in our case one week. Therefore, $\log(\lambda_t)$ follows a This ensures that the seasonal component τ_t accounts for the cumulative seasonal effects
over the specified period *S*, in our case one week. Therefore, $\log(\lambda_t)$ follows a local linear
trend with seasonality, where the 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" 179 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 up

180 limits of $\exp(\sigma_{\mu})$ and $\exp(\sigma_{\delta})$ eq 180 limits of $exp(\sigma_{\mu})$ and $exp(\sigma_{\delta})$ equal to 1.1. Model tuning is outlined in Supplementary
7 limits of $\exp(\sigma_\mu)$ and $\exp(\sigma_\delta)$ equal to 1.1. Model tuning is outlined in Supplementary
7

- 182 2,000 burn in.

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 x_t

18
- 182 2,000 burn in:

183 To produce the

184 equation (1) to

185 The β_i values a

187 sensible variab
-

185
$$
\log(\lambda_t) = \mu_t + \tau_t + \beta^T x_t, \text{ where } N_t \sim Poisson(\lambda_t)
$$

184 equation (1) to include the *i* regressor symptom category scaled counts $x_{i,t}$ in x_t

185 $log(\lambda_t) = \mu_t + \tau_t + \beta^T x_t$, where $N_t \sim Poisson(\lambda_t)$.

186 The β_i values are estimated using spike and slab priors [28] centred on

equation (1) to include the R regressor symptom category scaled counts $x_{i,t}$ in x_t

185 $log(\lambda_t) = \mu_t + \tau_t + \beta^T x_t$, where $N_t \sim Poisson(\lambda_t)$.

186 The β_i values are estimated using spike and slab priors [28] centred on zero 185 $log(\lambda_t) = \mu_t + \tau_t + \beta^T x_t$, where $N_t \sim Poisson(\lambda_t)$.

186 The β_i values are estimated using spike and slab priors [28] centred on zero

187 sensible variable selection. For the "BSTS + NHS 111 online" model we choos

188 le The p_i values are estimated using spike and slab priors [28] centred on zero to allow for

sensible variable selection. For the "BSTS + NHS 111 online" model we choose a training

length of 150 days, 5 expected regressi

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

193 Evaluation

To compare the ϵ Posterior Samples
Posterior Summary of key
Fraluation
To compare the difference
probabilistic frame
telling us the prope *Table 1. Sum
Table 1. Sum
Evaluatio
To compa
probabilis
telling us
our case !* and a structures, assumption (2,000)

and the structures, assumption

examption of the interval condition

tion of observation (1)

assumptic

sting ap

erval contraction

rinterval (2,000)

and chard

caches

rage is a

rat are and chard
and chard
oaches
rage is a
nat are v (1,000
to comp
ploy m
ure of
given p to compat
to compat
ploy mu
ure of p
given pr
lculate (1,000
(1,000
^{odel.}
c calib
terval (*1,000)*

ing met

c calibra

iterval r

ige devi 194 To compare the different nowcasting approaches we employ mattiple scoring methods in a
195 probabilistic framework. The interval coverage is a measure of probabilistic calibration,
196 elling us the proportion of obser 195 probabilistic framework. The interval coverage is a measure or probabilistic calibration,

196 telling us the proportion of observations that are within given prediction interval ranges

197 our case 50% and 90%. From 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 ne 197 our case 50% and 50%. 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 208 average difference between the measured interval coverage and the specified interval
199 value, with a coverage deviation nearer zero being preferred. The (weighted) interval so
200 (WIS) is a proper scoring rule compo 1999 value, with a coverage deviation nearer 2010 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 perform 201 (WIS) is a proper scoring rule composed or sharpness and under/overprediction, giving an overall measure of performance where low values are better. The weighted interval skill
8 201 overall measure of performance where low values are better. The weighted interval skill
8

- score is calculated as WISS $_{\text{model}} = 1 \frac{w_{\text{IS}} \text{model}}{\text{WIS}_{\text{baseline}}}$
model better than the "baseline" model. The bias
under/overprediction telling us if the models syst
bias models having a score nearer zero. The med
absolut www.pasem 202 score is calculated as WISS_{model} – 1
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 low
205
- model better than the "baseline" model. The bias is a relative measure of
-
-
-
- 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
205 bias models having a score nearer zero. The median absolute
- 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 n 207 using the scoringuins 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 c

209 day-of-week predicted. Since the data is upload
- 208 aggregations, as well as explored by noweast horizon it, where $h = 7$
209 day-of-week predicted. Since the data is uploaded weekly, the nowcas
210 corresponds to a unique day-of-week where Sunday will be a nowcas
211
- 204 under/overprediction telling us in 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 betwe 205 bias models having a secre 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* packa aggregations, as well as explored by nowcast horizon h , where $h = T - t$ in our case is the
day-of-week predicted. Since the data is uploaded weekly, the nowcast horizon h
corresponds to a unique day-of-week where *Sunda*
-

212 Results

-
- 209 day-of-week predicted. Since the data is uploaded weekly, the nowcast horizon of C
210 corresponds to a unique day-of-week where Sunday will be a nowcast horizon of C
211 Monday will have a nowcast horizon of 6 days.
2 210 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 211 Monday will have a nowcast horizon of 6 days.
212 **Results**
213 Winter 2023/2024 followed the seasonal trend
214 onwards, reaching a stable trend from Decemb
215 final and initial cases is largest in the most rece
216 213 Winter 2023/2024 followed the seasonal field of increasing cases from September
214 onwards, reaching a stable trend from December 2023 onwards. The difference bet
215 final and initial cases is largest in the most re
-
- 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). Acros mear zero (Figure 2a). Across each week approximately 20% of the data are revisions (cases) added the following week). These revisions can change the narrative of the real-time tre without correction (Figure 2b). The dist 215 final and initial cases is largest in the most recent days each week, as expected, with $n_{t,0}$
216 aer 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 218 without correction (Figure 2b). The distribution of d shows few reports on $d = 0$, a peak at 1-2 days and most reports within 7 days (Figure 3). The time varying reporting delay is given in 1-2 days and most reports w
- 218 without correction (Figure 2b). The distribution of α
219 1-2 days and most reports within 7 days (Figure 3). The time varying reporting delay is given

The time varying reporting delay is given
 $\frac{1}{2}$ 219 1-2 days and most reports within 7 days (Figure 3). The time varying reporting delay is given

221
222

 223 sinapshots of reporting, showing the most recent observed counts are substantially lower than the final revised data. (b.)
221 week become of tests at each inaset and final revisions. The end date for each week was 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 nowed of data from the previous week 224 weekly counts of tests at each ingest and
225 nowcast of data from the previous week.

226

227
228

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. 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 partially complete data underpredict the co 232 (Figure 4 & 5). Both the "GAM" and "epinoweast" models show increasing directiantly
232 towards the most recent date where data is more incomplete. The models using the
233 partially complete data underpredict the com 232 towards the most recent date where data is more incomplete. The models dsing the

233 partially complete data underpredict the complete cases in the week ending 14 January

234 2024, which we also see in the weekly est 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 i 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 r
236 though the "baseline" model has no 233 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 236 though the "baseline" modernas 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 value 237 prediction intervals compared to the "GAM", with the "epinowcast" model prediction
238 intervals being skewed towards higher values. 238 intervals being skewed towards higher values.

242 Figure 4. Daily predictions from all models with 50% and 90% prediction intervals against initial and final reported count of 243 tests.

245

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" outperf 249 has high WIS, expected given its small interval width. The partial reporting delay inodels
250 "epinowcast" and "GAM" outperform other models across WIS and MAE, generally
251 overpredicting, when other models are unde 250 "epinoweast" and "GAM" outperform other models across WIS and MAE, generally
251 overpredicting, when other models are underpredicting. The "BSTS" model perform
252 than the baseline across all daily metrics, whereas t 251 overpredicting, when other models are underpredicting. The 19513 inoderperforms better
252 than the baseline across all daily metrics, whereas the "BSTS + NHS 111 online" performs
253 broadly worst. Across daily and we 252 than the baseline across all daily metrics, whereas the "BSTS" NMS 111 online "periorms"
253 broadly worst. Across daily and weekly scoring the "BSTS" model has the best calibration
254 "GAM" and "epinowcast" models ov 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 254 with lowest coverage deviation, though other models have similar values. Notably, the
255 "GAM" and "epinowcast" models over and underpredict respectively.
256
257 255 "GAM" and "epinoweast" models over and underpredict respectively.
256
13 256

257

258

259

Table 2. Breakdown of overall model scores by temporal granularity. The daily granularity shows the average daily score 200 Table 2. Breakdown of overall model scores by temporal gramularity. The daily gramularity shows the average daily score
261 Awer the time ceries The weekly gramularity shows the average weekly score over the time cerie

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 "GAM" and "epinowcast" mo 264 SKIN OVER THE BASEINE INDEED IN MOST BUT NITE OF ALL WEENS (Figure 0c). 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 263 series, the "BSTS+NHS 111 online" modernias inglier wis than the baseline moder (Figure 6b). The "GAM" and "epinowcast" models have bias > 0 during the epidemic growth phas indicating overprediction (Figure 6c). The we 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 o

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 a 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 Supplemer
270 Figures 5 & 6. 269 particularly low. Further scoring at daily and weekly levels are given in Supplementary
270 Figures 5 & 6.

 270 Figures 5 & 6.

271

272
273

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.

by breaking down by the day-of-week (and therefore nowcast horizon, in our case) we can
explore how varying data completeness affects model performance. Relative to "baseline"
the "BSTS" model exhibits a flat skill across 277 explore now varying data completeness ancees moder performance. Relative to "baseline"
278 the "BSTS" model exhibits a flat skill across days (Figure 7a), whereas the relative skill of the
280 "GAM" and "epinowcast" ge 279 "CAM" and "epinowcast" gets deteriorates towards the end of the week (Figure 7b). The "baseline" consistently underpredicts, while "epinowcast" underpredicts at the start of the week but becomes less biased toward Sund 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 tow 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 com 281 week but becomes less biased toward Sunday (Figure 7c). Compared to the 19513 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 comp 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.
15 283 prediction week when data is more complete.

284

285
286

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 nearth service winter pressures through hospital
293 outbreaks, reduced bed availability and staff absences. As such, timely surveillance is cr
293 for situational awareness, part 292 outbreaks, reduced bed availability and start 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 repo 293 for situational awareness, particularly to understand enanges 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 a 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 show
296 that harnessing partially compl 295 Individual 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 sens 297 that harnessing partially complete data outperforms a transact and forceast approach, but
297 the performance can be sensitive to the consistency of case reporting, which is challenging
298 in frontline health protecti 297 the performance can be sensitive to the consistency of case reporting, which is changing
298 in frontline health protection. The delay in reporting impacts the analysis of trends in
299 national surveillance, so it is 298 in frontline health protection. The delay in reporting impacts the analysis of trends in
299 in ational surveillance, so it is important official reporting exclude these partially report
200 days, though nowcasting can national surveillance, so it is important official reporting exclude these partially reported
days, though nowcasting can support decision making in real-time. The nowcasting problem
16 300 days, though nowcasting can support decision making in real-time. The nowcasting problem

302 delay, a single test type, and without considering regional or age-related variation. This n
303 partially explain the strong performance of approximate methods in the scoring.
304 Nowcasting approaches are increasingl 303 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 increasin Bartially explain the strong performance of approximate inethods in the scoring.
304 Nowcasting approaches are increasingly used to predict case counts by accountin
305 delays in reporting, and have been crucial in the rec 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 fro
307 the epidemic literature to this prob 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 fro
307 the epidemic literature to this prob 307 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 join 308 implementation, *epinowcast*, which jointly models a reporting delay distribution
309 underlying process model, and a more approximate but highly flexible and comp
310 efficient GAM-based model. We also consider a Baye 309 implementation, epinowcast, which jointly models a reporting delay distribution with an
310 underlying process model, and a more approximate but highly flexible and computationa
311 testing the utility of incorporating 310 underlying process model, and a more approximate but highly hexible and computationally
310 efficient GAM-based model. We also consider a Bayesian structural time series approach,
311 knowledge this is the first study 311 testing the utility of incorporating leading indicators into the modelling framework. To our knowledge this is the first study to apply time delay nowcasting methods to norovirus case
313 which may be more challenging 312 knowledge this is the first study to apply time delay nowcasting methods to norovirus case.
313 which may be more challenging to nowcast than other infectious diseases due to high level
314 of underreporting, regional ST2 knowledge this is the first study to apply time delay noweasting inctitious to norovirus cases,
313 which may be more challenging to nowcast than other infectious diseases due to high levels
315 of underreporting, regi 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 generated operationally useful p 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 substant
317 improvement over using truncated da 315 settings such as care homes, senools and hospitals [31]. Despite this, several models
316 generated operationally useful predictions of norovirus test counts, offering a substa
317 improvement over using truncated data Site and the current standard) or a naïve seasonal baselin
318 However, when reporting delay data is unavailable, time series forecasting presents an
319 adaptive alternative with good coverage and performance compared to 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 n 319 adaptive alternative with good coverage and performance compared to the baseline. In contrast to previous studies, we did not find including leading indicators improved our predictions [32]. This could perhaps be expla 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 323 contrast to previous studies, we did not find including reading indicators improved our
321 predictions [32]. This could perhaps be explained by lower signal in the indicators
322 considered, related to confounding eff predictions [32]. This could perhaps be explained by lower signal in the indicators
322 considered, related to confounding effects from other winter pathogens. Finally, o
323 findings that several models perform well with

323 considered, related to confounding enects from other winter pathogens. Finally, our
323 findings that several models perform well with different accuracy and biases over tim
324 day of the week suggests the potential b 323 Finally strat 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] 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
327 325 demonstrated in other contexts [12].
326 Models incorporating reporting delay
327 models that do not, showing the utilit
328 erformance is driven by reduced unc
330 time delay approximation method in t
331 complex "epi 327 Models incorporating reporting delays consistently performed better than forecasting
327 models that do not, showing the utility of leveraging this data when available. This imp
328 early in the nowcast window. Among o 328 performance is driven by reduced uncertainty when there is more complete reported data,
329 erformance is driven by reduced uncertainty when there is more complete reported data,
329 early in the nowcast window. Among Berival and the nowcast window. Among our models using reporting delays, we found that the

time delay approximation method in the "GAM" scored slightly better than the more

complex "epinowcast" model's full joint distrib Early in the nowcast window. Among our models using reporting delays, we found that the
time delay approximation method in the "GAM" scored slightly better than the more
complex "epinowcast" model's full joint distribution 331 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 d S31 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. Wid 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 hat modelling based on recent dis 333 and underlying process model. Wide intervals are penalised in scoring interies like the WIS,
334 however, this larger uncertainty may better reflect the uncertainty in the system. We saw
335 that modelling based on rec 335 that modelling based on recent distributions of reporting delays can perform poorly if thes
336 distributions change rapidly, although in these cases, the "epinowcast" model's optional
337 time-varying delay may be adv 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 ad 337 time-varying delay may be advantageous compared to a fixed distribution approach, such the one in the "GAM". Speed is key in a real-time modelling context, with some models being substantially faster than others, howev 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. 339 the one in the "GAM". Speed is key in a real-time infodentig context, with some models
340 being substantially faster than others, however, all approaches ran in a reasonable time
340 (Table 1) for real-time inference. 3339 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
341 other models, however, was impactf 340 (Table 1) for real-time inference. The computational expense of "epinowcast" compared to
341 other models, however, was impactful during model development.
17 341 other models, however, was impactful during model development.
17

-
-
-
- 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 incidenc 343 Inverparameter 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 incidenc 343 simulating a plausible real-time sechario – which may bias selection toward
345 hyperparameters good at flat periods of incidence. There are reporting chan
346 healthcare delivery which can impact the performance of ti
-
-
- Thealthcare 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 modellin
348 which should be explored further. Futu
- 343 healthcare delivery which can impact the performance of time delay informed models –
347 hese local practices are challenging to understand in real-time and adjust for in modelling,
348 which should be explored further
-
-
-
- 247 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
350 can be incorporated into mo Which should be explored further. Future work should explore how local testing practices

as can be incorporated into modelling directly. Understanding testing pathways and real-time

350 modelling of norovirus will be cru 249 can be incorporated into modelling directly. Onderstanding testing pathways and real-time
350 modelling of norovirus will be crucial for the next strain replacement event highlighting the
351 importance of developing o 351 importance of developing or understanding and preparedness.
351 importance of developing our understanding and preparedness.
353 Strain and an unpleasant infection for the individual, increasing associated opportunity 352 Importance of developing our understanding and preparedness.
352 While not a high priority pandemic potential pathogen, norovirus
353 Israin and an unpleasant infection for the individual, increasing a
354 Israel by bl SEREGAN STREET THOTHY PAINTERT PROTHER PATHOGEN, INCOVITATS CAUSES HEATIGHTER SYSTEM STRAIN AND A strain and an unpleasant infection for the individual, increasing associated opportunity cost
354 by blocking beds and elong
-
-
- by blocking beds and elongating patient length of stay [3]. Estimating the current case
burden when accounting for delayed reporting can be an important tool for supporting
effective public health response. In this work we 353 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, highl
-
- 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 d 255 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 will underpin situational awareness 357 correct for delayed reporting, highlighting their strengths and inhibited ons – notably
358 demonstrating the importance of explicitly modelling the partially complete data. T
360 **Contributions**
361 **JM** – Conceptuali
-

360 Contributions

-
-
- 359 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 Conceptua 359 Will underpin situational awareness should the next strain replacement event occur.
360 **Contributions**
361 **JM** – Conceptualisation, Methodology, Software, Validation, Formal Analysis, Data Cu
362 Writing – Original D
-
- 362 JM Conceptualisation, Methodology, Software, Validation, Formal Analysis, Data Curation,
363 MT Methodology, Software, Validation, Formal Analysis, Data Curation, Writing Original
364 Draft, Writing Review & Edit 362 Writing Original Draft, Writing Review & Editing, Visualisation, Troject Administration
363 MT - Methodology, Software, Validation, Formal Analysis, Data Curation, Writing – Origina
364 Draft, Writing – Review & Editin 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, 364 Draft, Writing Heview & Editing, Visualisation
365 EF - Methodology, Software, Formal Analysis, W
366 Editing, Visualisation
367 RC – Conceptualisation, Data Curation, Writing
369 CEO – Methodology, Writing – Review & 263 Editing, Visualisation

366 Editing, Visualisation

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

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

369 CEO – Methodology, Writin
- 367 RC Conceptualisation
368 OP Software, Forma
369 CEO Methodology,
370 AH Investigation, W
371 AD Conceptualisation
372 SRD Conceptualisation
-
-
-
- 367 RC Conceptualisation, Data Curation, Writing Review and Editing

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

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

371 **AD** Conceptualisation, 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 – 369 CEO Methodology, Writing Review & Editing
370 AH – Investigation, Writing – Review & Editing
371 AD – Conceptualisation, Investigation, Writing –
372 SRD – Conceptualisation, Writing – Review & Edit
373 TW – Writing – 370 AH Investigation, Writing Review & Editing
371 **AD** – Conceptualisation, Investigation, Writing
372 **SRD** – Conceptualisation, Writing – Review & E
373 **Ethical Approval**
375 UKHSA have an exemption under regulation 3
-
-
-
-
- 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 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 th
376 Act (2006) 373 TW Willing Review & Editing, Supervision
374 Ethical Approval
375 UKHSA have an exemption under regulation 3
376 Act (2006) to allow identifiable patient inform
377 prevent, or recognise trends in, communicable
Confli 374 Ethical Approval
375 UKHSA have an e.
376 Act (2006) to allo
377 prevent, or recog
378 **Conflict of Intere**
379 The authors have 373 UKHSA have an exemption under regulation 3 of section 251 of the National Health Service

Act (2006) to allow identifiable patient information to be processed to diagnose, control,

377 prevent, or recognise trends in,
- 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 **Data Availability Statement**
382

- 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 Data Availability Statement
382
- 380
381
- 378 Conflict of Interest
379 The authors have de
380 **Data Availability St**
382
18 379 The authors have declared that no competing interests exist.
380
381 Data Availability Statement
382
18 Data Availability Statement
- 382

-
-
-
-
- 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 ena
- 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 mode
387 is available at https://github.com/jo
-
- 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 inf 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
399 identifiability. An applicat 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 make to the UK Health Security Agency.
390 UKHSA application for dat 389 on the reporting delay used to inform initial exploration are not available due to patient
389 identifiability. An application for data access can be make to the UK Health Security Ager
390 UKHSA applicates a robust go 389 Identifiability. An application for data access can be make to the OK Health Security Agency.
390 UKHSA ?
391 considers:
392 • the benefits and risks of how the data will be used
393 • compliance with policy, regulator 390 UKHSA Boperates 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 oblig
-
-
- 391 considers:
- -
- 392 the benefits and risks of how the data will be used
393 compliance with policy, regulatory and ethical oblig
394 data minimisation
395 how the confidentiality, integrity, and availability
396 retention, archi 393 • compliance with policy, regulatory and ethical obligations
394 • data minimisation
395 • how the confidentiality, integrity, and availability will be m
396 • retention, archival, and disposal requirements
397 • best 394 • data minimisation
395 • how the confidenti
396 • retention, archival,
397 • best practice for pr
398 by default', emergi
399 Access to protected data i
400 contracts. • 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
398 by default', 396 • retention, archival, and disposal requirements
397 • best practice for protecting data, including the
398 by default', emerging privacy conserving techn
399 Access to protected data is always strictly controlled u
40
- 399 Access to protected data is always strictly controlled using legally binding data sharing
399 Access to protected data is always strictly controlled using legally binding data sharing
301 UKHSA Nelcomes data applicatio

• 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 co SEE Access to protected data is always strictly controlled using legally binding data sharing

400 contracts.

401 UKHSA Nelomes data applications from organisations looking to use protected data f

402 public health purpo 400 contracts.
401 UKHSA
500 public hea
403 To request
404 submit, co
405

- 401 UKHSAEWEICOMES 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
405 References 402 public health purposes.
403 To request an applicatio
404 submit, contact <u>DataAcd</u>
405 References 403 To request an application pack or discuss a request for OKHSA data you would like to
404 Submit, contact <u>DataAccess@ukhsa.gov.uk</u>.
405 References
-

405
406
407

- 404 submit, contact <u>DataAccess@ukhsa.gov.uk</u>.
405
407 [1] J. Xerry, C. I. Gallimore, M. Iturriza-Gómar 406 References

407

[1] J. Xerry, C. I.

within outbre

the P2 domai

947-953, 200 within outbreaks of gastroenteritis determined through analysis of nucleotide sequences c
the P2 domain of genogroup II noroviruses," Journal of clinical microbiology, vol. 46, no. 3,
947-953, 2008.
[2] S. M. Bartsch, B. A 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.
S. M. Bartsch, B.
	- the P2 domain of genogroup II noroviruses," *Journal by Elimear IIII cobiology, vol.* 46, no. 3, pp.
947-953, 2008.
S. M. Bartsch, B. A. Lopman, S. Ozawa, A. J. Hall and B. Y. Lee, "Global economic burden of
norovirus gast
	- 947-953
S. M. Bartsch, B
norovirus gastre
F. G. Sandmann
Larkin, F. Wurie 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. Alle patients," *Clinical Infectious Diseases,* vol. 67, no. 5, pp. 693-700, 2018.
K. M. O'Reilly, F. Sandman, D. Allen, C. I. Jarvis, A. Gimma, A. Douglas, L. Larkin, K. L. Wong, M. 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 Disease*
	- norovirus-associated gastroenteritis in England and its opportunity costs for nonadmitted
patients," *Clinical Infectious Diseases,* vol. 67, no. 5, pp. 693-700, 2018.
K. M. O'Reilly, F. Sandman, D. Allen, C. I. Jarvis, A. northelians," *Clinical Infectious Diseases, vol.* 67, no. 5, pp. 693-700, 2018.
K. M. O'Reilly, F. Sandman, D. Allen, C. I. Jarvis, A. Gimma, A. Douglas, L. Larkin, K. L. Wong
Baguelin, R. S. Baric, L. C. Lindesmith, R. A patients, "*Chincal Infectious Diseases, vol. 07, no. 5, pp. 693-700, 2018.*
K. M. O'Reilly, F. Sandman, D. Allen, C. I. Jarvis, A. Gimma, A. Douglas, L.
Baguelin, R. S. Baric, L. C. Lindesmith, R. A. Goldstein, J. Breuer Example 1. M. 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 restric 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.
P. Whi associated with COVID-19 restrictions in England: a mathematical modeling study," *BMC*
Medicine, vol. 19, pp. 1-10, 2021.
P. White, "Evolution of norovirus," Clinical Microbiology and Infection, vol. 20, no. 8, p.
74107
	- associated with COVID-19 restrictions in England: a mathematical modeling study, BMC
Medicine, vol. 19, pp. 1-10, 2021.
P. White, "Evolution of norovirus," Clinical Microbiology and Infection, vol. 20, no. 8, p.
7410745, 2 Medicine, vol. 19, pp. 1-10, 2021.
P. White, "Evolution of norovirus,
7410745, 2014. $[5]$ P. White, "Evolution of norovirus," *Clinical Microbiology and Injection*, vol. 20, no. 8, p.
7410745, 2014. 741
741 - 2014
741 - 2014

-
- 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. P.

GII. 4 Sydney 2012 norovirus lineage is circulating worldwi Norovirus strains, "Pios one, vol. 7, no. 7, p. 108, 2012.
C. Ruis, S. Roy, J. R. Brown, D. J. Allen, R. A. Goldstein ar
GII. 4 Sydney 2012 norovirus lineage is circulating world
polymerase changes that may increase virus GII. 4 Sydney 2012 norovirus lineage is circulating worldwide, arose by late-2014 and contain polymerase changes that may increase virus transmission," PloS one, vol. 12, no. 6, pp. 1-9, 2017.

2017.

[8] UK Health Securit
- polymerase changes that may increase virus transmission," *PloS one, vol.* 12, no. 6, pp. 1-9, 2017.
UK Health Security Agency, "National norovirus and rotavirus surveillance reports: 2023 to
2024 season," 9 May 2024. [Onl UK Health Security Agency, "National norovirus and rotavirus surveillance reports: 2023 to 2024 season," 9 May 2024. [Online]. Available: [8] UK Government, "The Health Protection (Notification) Regulations 2010," 2010. [Online].

2024 season," 9 May 2024. [Online]. Available:

reports-2023-to-2024-season.

[9] UK Government, "The Health Protection (Notifica 2022 season, " 9 May 2024 season, https://www.gov.uk/government/statistics/natreports-2023-to-2024-season.
PUK Government, "The Health Protection (Notification).
Available: https://www.legislation.gov.uk/uksi/
-
- Harris, "Comparison of statistical approaches to predicting norovirus laboratory reports before reports-2023-
UK Government, "The Health F
Available: https://www.legislat
N. Ondrikova, H. Clough, A. Do
Harris, "Comparison of statistic [9] UK GOVERNMENT, THE HEALTH PROVIDING (NOTIFICATION, NCGLESS) 2010-19.

2010 N. Ondrikova, H. Clough, A. Douglas, R. Vivancos, M. Itturiza-Gomara, N. Cunliffe and J. P.

10] N. Ondrikova, H. Clough, A. Douglas, R. Vivanc M. Ondrikova, H. Clough, A. Douglas, R. Vivancos, M. Itturiza-Gomara, N. Cu
Harris, "Comparison of statistical approaches to predicting norovirus laborand during COVID-19: insights to inform public health surveillance," Sc Harris, "Comparison of statistical approaches to predicting norovirus laboratory reports be
and during COVID-19: insights to inform public health surveillance," *Scientific reports,* vol.
no. 1, 2023.
[11] S. Lee, E. Cho, 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 learnin
- and during COVID-19: insights to inform public health surveillance, "Scientific reports, vol. 15, no. 1, 2023.
S. Lee, E. Cho, G. Jang, S. Kim and G. Cho, "Early detection of norovirus outbreak using
machine learning metho
- s.
S. Lee, E. Che
machine lea
D. Wolffram
van de Kasst machine learning methods in South Korea," *PLoS One,* vol. 17, no. 11, 2022.

[12] D. Wolffram, S. Abbott, M. An der Heiden, S. Funk, F. Günther, D. Hailer, S. Heyder, T. I

van de Kassteele, H. Küchenhoff, S. Muller-Hanse machine learning methods in South Korea, "PLOS One, vol. 17, no. 11, 2022.
D. Wolffram, S. Abbott, M. An der Heiden, S. Funk, F. Günther, D. Hailer, S. F
van de Kassteele, H. Küchenhoff, S. Muller-Hansen, D. Syliqi, A. Ull (113) U. 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.
- "Collaborative nowcasting of COVID-19 hospitalization incidences in Germany," PLOS
Computational Biology, vol. 19, no. 8, 2023.
J. T. Wu, K. Leung, T. T. Lam, M. Y. Ni, C. K. Wong, J. M. Peiris and G. M. Leung, "Nowcas
epi Computational Biology, vol. 19, no. 8, 2023.

J. T. Wu, K. Leung, T. T. Lam, M. Y. Ni, C. K. Wong, J. M. Peiris and G. M. Leung, "Nowe

epidemics of novel pathogens: lessons from COVID-19," Nature Medicine, vol. 27, no.

3 Computational Biology, vol. 15, no. 8, 2023.
J. T. Wu, K. Leung, T. T. Lam, M. Y. Ni, C. K. W.
epidemics of novel pathogens: lessons from
388-395, 2021.
UK Health Security Agency, "Guidance: Notif
- 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 20 epidemics of novel pathogens: lessons from COVID-19," Nature Medicine, vol. 27, no. 9, pp.
388-395, 2021.
UK Health Security Agency, "Guidance: Notifiable diseases and causative organisms: how to
report," 1 January 2024. [Servery Server
UK Health Secu
report," 1 Janua
diseases-and-ca
UK Health Secu
- metalth Security Agency 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 ata/file/1159953/UKHSA_Laboratory_reporting_guidelines_May_2023.pdf. METHER THERM PRAILED TO SAMPLE AT THE PERSON
UK Health Security Agency, "Laboratory reporting to
laboratories," May 2023. [Online]. Available:
https://assets.publishing.service.gov.uk/governme
ata/file/1159953/UKHSA_Labora [15] Ukrain Security May 2023. [Online]. Available:
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/atta
ata/file/1159953/UKHSA_Laboratory_reporting_guidelines_May_2023.pdf.
[16] S. F. McGough, M
- May 2023. [Chinagon Line Promance.
https://assets.publishing.service.gov.uk/gove
ata/file/1159953/UKHSA_Laboratory_reporti
S. F. McGough, M. A. Johansson, M. Lipsitch a
Smoothing: A flexible, generalizable model fo
computa https://assets.publishing.com/government/government/uploads/file/1159953/UKHSA_Laboratory_reporting_guidelines_May_2023.pdf.
S. F. McGough, M. A. Johansson, M. Lipsitch and N. A. Menzies, "Nowcasting by Bayesian
Smoothing: ata. F. McGough, M. A. Johansson, M. Lipsitch and N. A. Menzies, "Nowcasting
Smoothing: A flexible, generalizable model for real-time epidemic tracking,"
computational biology, vol. 16, no. 4, p. e1007735, 2020.
NHS, "111 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:/
-
- Smoothing: A flexible, generalizable model for real-time epidemic tracking, PLOS
computational biology, vol. 16, no. 4, p. e1007735, 2020.
NHS, "111 online, Get help for your symptoms," [Online]. Available: https://111.n
C computational biology, vol. 16, no. 4, p. e1007735, 2020.
NHS, "111 online, Get help for your symptoms," [Online].
C. E. Overton, S. Abbott, R. Christie, F. Cumming, J. Day, O
Ward, "Nowcasting the 2022 mpox outbreak in En [18] C. E. Overton, S. Abbott, R. Christie, F. Cumming, J. Day, O. Jones, R. Paton, C. Turner and
Ward, "Nowcasting the 2022 mpox outbreak in England," *PLoS computational biology,* vo
20 Ward, "Nowcasting the 2022 mpox outbreak in England," *PLoS computational biology*, vol. 1
20 Ward, "Nowcasting the 2022 mpox outbreak in England," PLos computational biology, vol. 19,

- no. 9, p. e1011463, 2023.

[19] J. van de Kassteele, P. H. Eilers and J. Wallinga, "Nowcasting the number of new symptomatic

cases during infectious disease outbreaks using constrained P-spline smoothing,"
 Epidemiology, 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.
-
- Epidemiology, vol. 30, no. 5, pp. 737-745, 2019.
S. Wood, "Package `mgcv`," *R package version,* vol. 1, no. 29, p. 729, 2015.
G. L. Simpson, "gratia: graceful ggplot-based graphics and other functions for GAN
using mgcv," Epidemiology, vol. 30, no. 3, pp. 737-743, 2013.
S. Wood, "Package `mgcv`," *R package version,* \
G. L. Simpson, "gratia: graceful ggplot-based gra
using mgcv," 2024. [Online]. Available: https://g
S. Abbot. A. Lison. S. [20] S. Wood, Tackage migev, *N puckuge version, vol.* 1, no. 29, p. 729, 2015.
[21] G. L. Simpson, "gratia: graceful ggplot-based graphics and other functions fit using mgcv," 2024. [Online]. Available: https://gavinsimps
- Engine of the University of the SIBO, The Sibbot, A. Lison, S. Funk, C. Pearson, H. Gruson, F. Guenther and M. DeWitt, epinowcast:
F using mga mandalakan in 1924. [Canada Sanda S
Flexible Hierarchical Nowcasting, 10.5281/zenodo.5637165.
M. Höhle and M. an der Heiden, "Bayesian nowcasting
- Flexible Hierarchical Nowcasting, 10.5281/zenodo.5637165.
[23] M. Höhle and M. an der Heiden, "Bayesian nowcasting during the STEC 0104: H4 outbreak in
Germany, 2011," *Biometrics,* vol. 70, no. 4, pp. 993-1002, 2014.
[24] Germany, 2011," Biometrics, vol. 70, no. 4, pp. 993-1002, 2014.
- Example and M. and M. Höhler, V. and A. app. 993-1002, 2014.

[24] F. Günther, A. Bender, K. Katz, H. Küchenhoff and M. Höhle, "Nowcasting the COVID-19

pandemic in Bayaria," *Biometrical Journal*, vol. 63, no. 3, pp. 490-
- pandemic in Bavaria," *Biometrical Journal,* vol. 63, no. 3, pp. 490-502, 2021.
B. Carpenter, A. Gelman, M. D. Hoffman, D. Lee, B. Goodrich, M. Betancourt, M. A. Brubake
Guo and P. Li, "Stan: A probabilistic programming la material formulation and M. B. S. Goodrich, M. Betancourt, M. A. Brub Guo and P. Li, "Stan: A probabilistic programming language," Journal of statistical softwards and M. Höhle, "Now the Covid-19" and M. Höhle, "Now the Co pandemic in Bavaria, "*Biometrical Journal*, vol. 63, no. 3, pp. 450-502, 2021.
B. Carpenter, A. Gelman, M. D. Hoffman, D. Lee, B. Goodrich, M. Betancourt
Guo and P. Li, "Stan: A probabilistic programming language," *Journ*
- [25] B. 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. Guo and P. Li, "Stan: A probabilistic programming language," Journal of Statistical software,
vol. 76, 2017.
S. L. Scott and H. R. Varian, "Predicting the present with Bayesian structural time series,"
International Journa vol.
S. L. Scott and
International .
pp. 4-23, 2014
S. L. Scott, *R P* International Journal of Mathematical Modelling and Numerical Optimisation, vol. 5, no.

pp. 4-23, 2014.

[27] S. L. Scott, *R Package 'bsts'*, 2016.

[28] H. Ishwaran and J. S. Rao, "Spike and slab variable selection: Fre
-
- International Journal of Mathematical Modelling and Numerical Optimisation, vol. 5, no. 1-2,
pp. 4-23, 2014.
B. L. Scott, *R Package 'bsts',* 2016.
H. Ishwaran and J. S. Rao, "Spike and slab variable selection: Frequentist 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. 730-7
- [27] S. L. Scott, *N Puckage 'bsts'*, 2016.

[28] H. Ishwaran and J. S. Rao, "Spike a

strategies," *The Annals of Statistic*.

[29] N. I. Bosse, H. Gruson, A. Cori, E. vi

with scoringutils in R," 2022. [Onlin 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 fored

with scoringutils in R," 2022. [Online]. Available:
- strategies," *The Annuls of Statistics, vol. 33, no. 2, pp. 730-773, 2003.*
N. I. Bosse, H. Gruson, A. Cori, E. van Leeuwen, S. Funk and S. Abbott,
with scoringutils in R," 2022. [Online]. Available: https://arxiv.org/abs
 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 Mpo 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 Healt
Making and Lessons Learned," *Epidemics,* vol. 47, no.
- "Nowcasting and Forecasting the 2022 US Mpox Outbreak: Support for Public Healt
Making and Lessons Learned," *Epidemics,* vol. 47, no. 1755-4365, p. 100755, 2024.
[31] N. Ondrikova, H. Clough, N. Cunliffe, M. Iturriza-Goma Making and Lessons Learned," *Epidemics,* vol. 47, no. 1755-4365, p. 100755, 2024.
N. Ondrikova, H. Clough, N. Cunliffe, M. Iturriza-Gomara, R. Vivancos and J. Harris,
"Understanding norovirus reporting patterns in England Making and Lessons Learned," *Epidemics, Vol. 47, no. 1755-4365, p. 100755, 2024.*
N. Ondrikova, H. Clough, N. Cunliffe, M. Iturriza-Gomara, R. Vivancos and J. Harris,
"Understanding norovirus reporting patterns in England
- "Understanding norovirus reporting patterns in England: a mixed model approach,

Public Health, vol. 21, pp. 1-9, 2021.

[32] F. Bergström, F. Günther, M. Höhle and T. Britton, "Bayesian nowcasting with leadi

indicators a Public Health, vol. 21, pp. 1-9, 2021.
Public Health, vol. 21, pp. 1-9, 2021.
F. Bergström, F. Günther, M. Höhle and T. Britton, "Bayesian nowcasting with leading
indicators applied to COVID-19 fatalities in Sweden," *PLOS* Public Health, vol. 21, pp. 1-9, 2021.
F. Bergström, F. Günther, M. Höhle a
indicators applied to COVID-19 fatali
12, p. e1010767, 2-22. Indicators applied to COVID-19 fatalities in Sweden," PLOS Computational Biology, vol
12, p. e1010767, 2-22.
T. Britan now casting with leading indicators applied to COVID-19 fatalities in Sweden," PLOS Computational Biology, vol. 18, no.
12, p. e1010767, 2-22. 12, p. encouraged the control of th
12, p. e.g. control of the control o

408

...
409