



MRC Centre for
Global Infectious
Disease Analysis

Imperial College
London

Phylogenomic early warning signals for SARS-CoV-2 epidemic waves

Kieran Drake

MRC Centre for Global Infectious Disease Analysis

WHO Collaborating Centre for Infectious Disease modelling

*School of Public Health
Imperial College London*

NORDITA program: Unifying the epidemiological and evolutionary dynamics of pathogens
21 June 2023



GLOBALLY,
access to timely
genomic sequencing
is improving

77% of countries
can now sequence
SARS-CoV-2 virus in-country

22% have access
to timely sequencing
through an **international
referral mechanism**

Only **1%**
of countries have
no capability or
timely access yet

Source: WHO, as of July 2022. Countries refers to WHO Member States. Data not available for 4 of 194 countries.

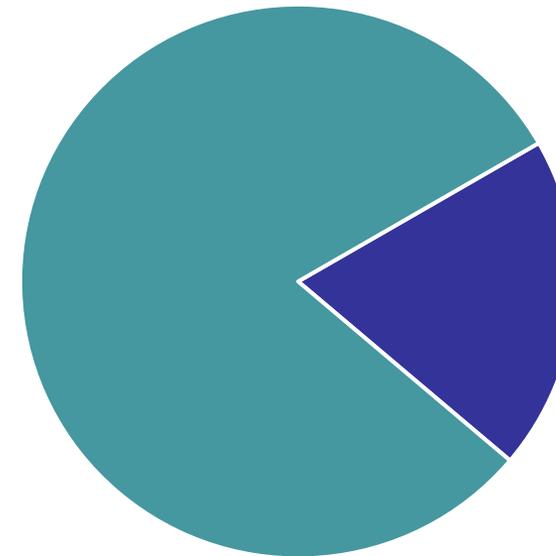
The Global Genomic Surveillance Strategy
for Pathogens with Pandemic and Epidemic Potential 2022 - 2032



<https://www.who.int/initiatives/genomic-surveillance-strategy>



15.21 million
hCoV-19 genome sequences



**COVID-19
GENOMICS
UK CONSORTIUM**

2.97 million
19.5%

GISAID data <https://gisaid.org/>

COG UK data https://cog-uk.s3.climb.ac.uk/phylogenetics/latest/cog_metadata.csv.gz
17 March 2023

Early warning signals (EWS)



MRC Centre for
Global Infectious
Disease Analysis

Imperial College
London



Typical approach

- Theory - 'Critical slowing down' with critical transition from $R_t < 1$ to $R_t > 1$
- Leading indicators - changes in statistics (e.g. variance, skewness, autocorrelation) of incidence or prevalence data



Typical approach

- Theory - 'Critical slowing down' with critical transition from $R_t < 1$ to $R_t > 1$
- Leading indicators - changes in statistics (e.g. variance, skewness, autocorrelation) of incidence or prevalence data

Our approach

- Theory - rapid detection of growing genomic variants is a predictor of future epidemic waves
- Leading indicators - derived from SARS-CoV-2 pathogen genome via phylodynamic analysis



3



Phylogenetic analysis



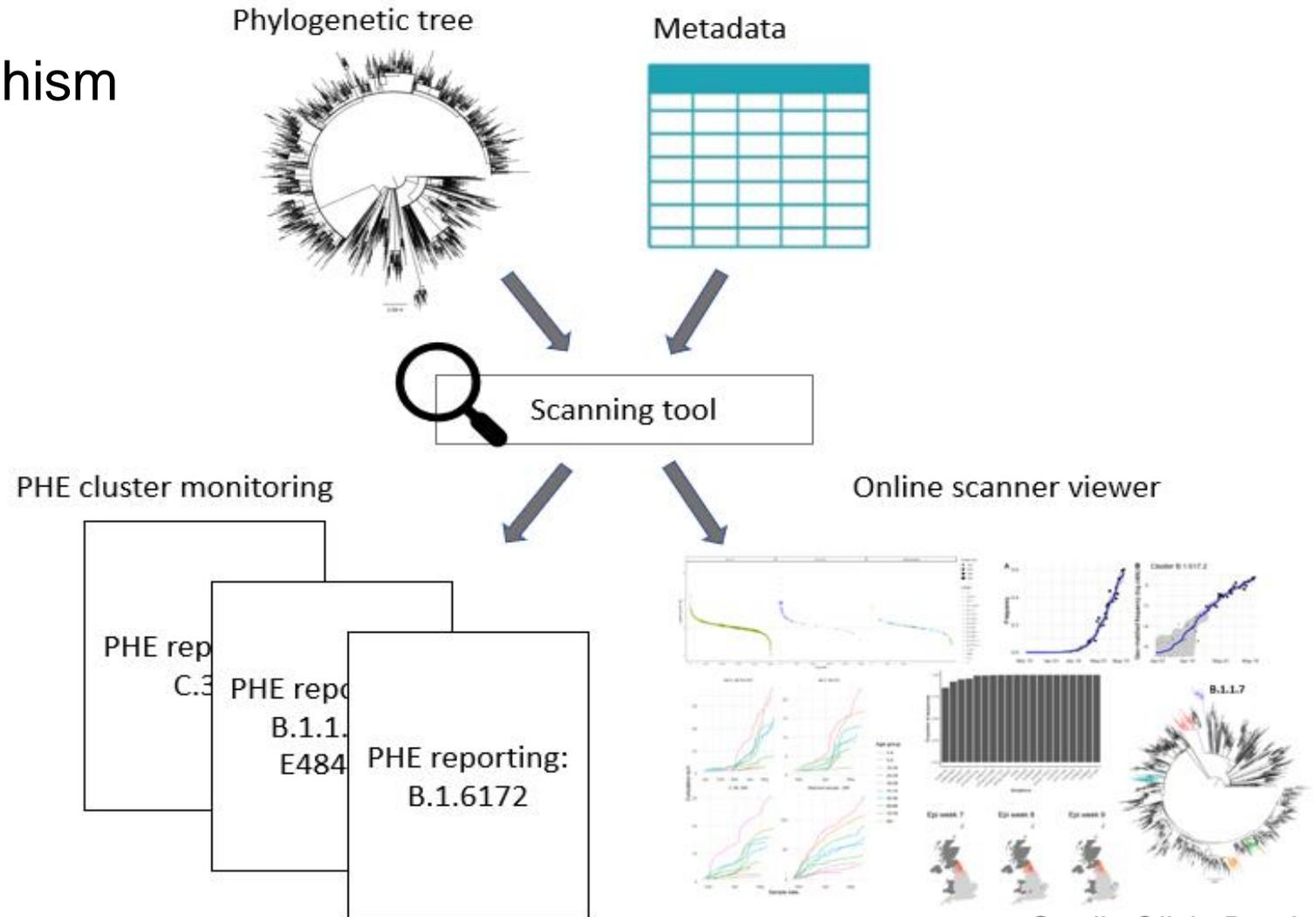
Generate potential leading indicators



Assess performance as early warning signal (EWS)

Transmission Fitness Polymorphism (TFP) Scanner

- Designated in an R package *mrc-ide/TFPScanner*
- Developed by Erik Volz, Olivia Boyd and Manon Ragonnet-Cronin



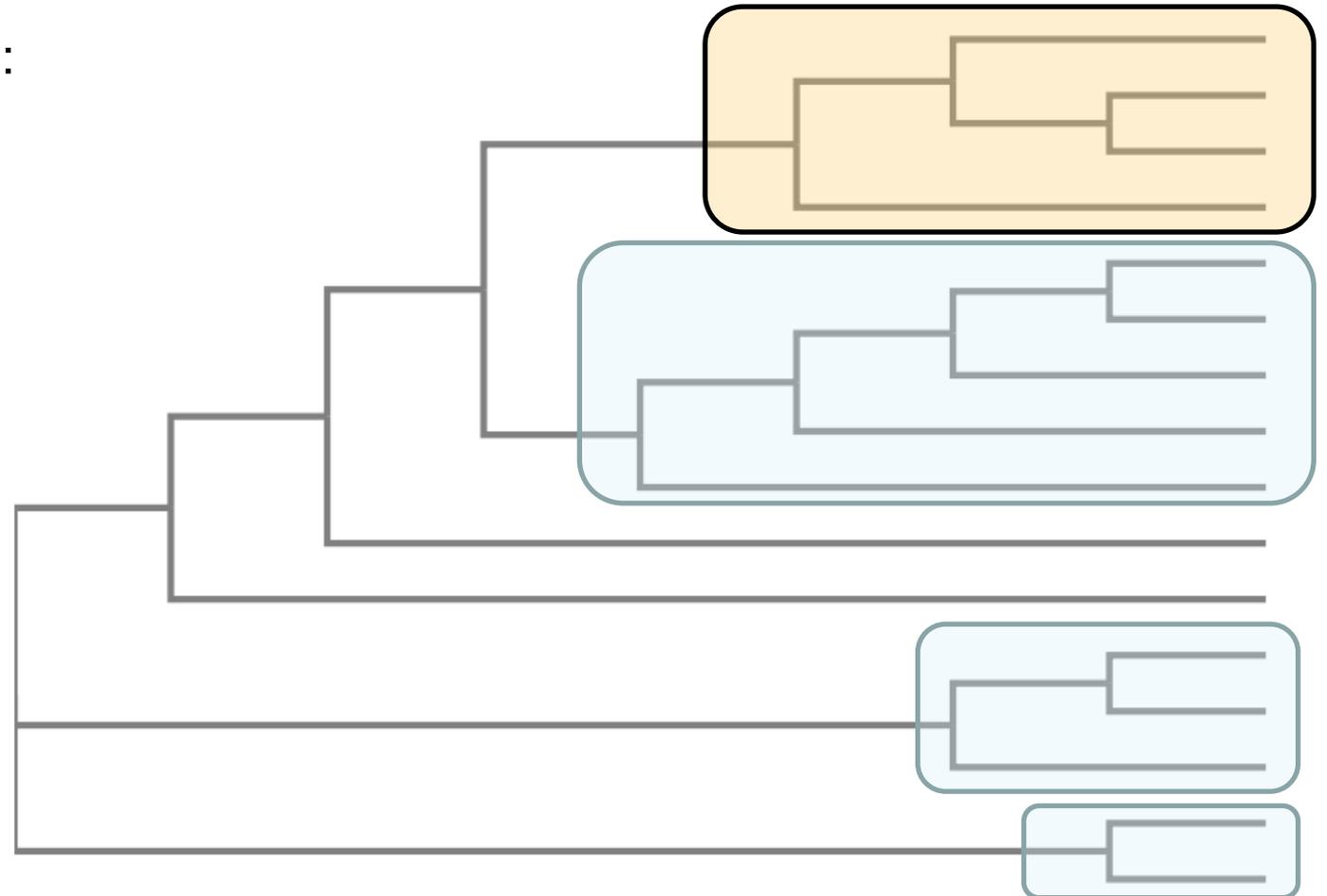
5

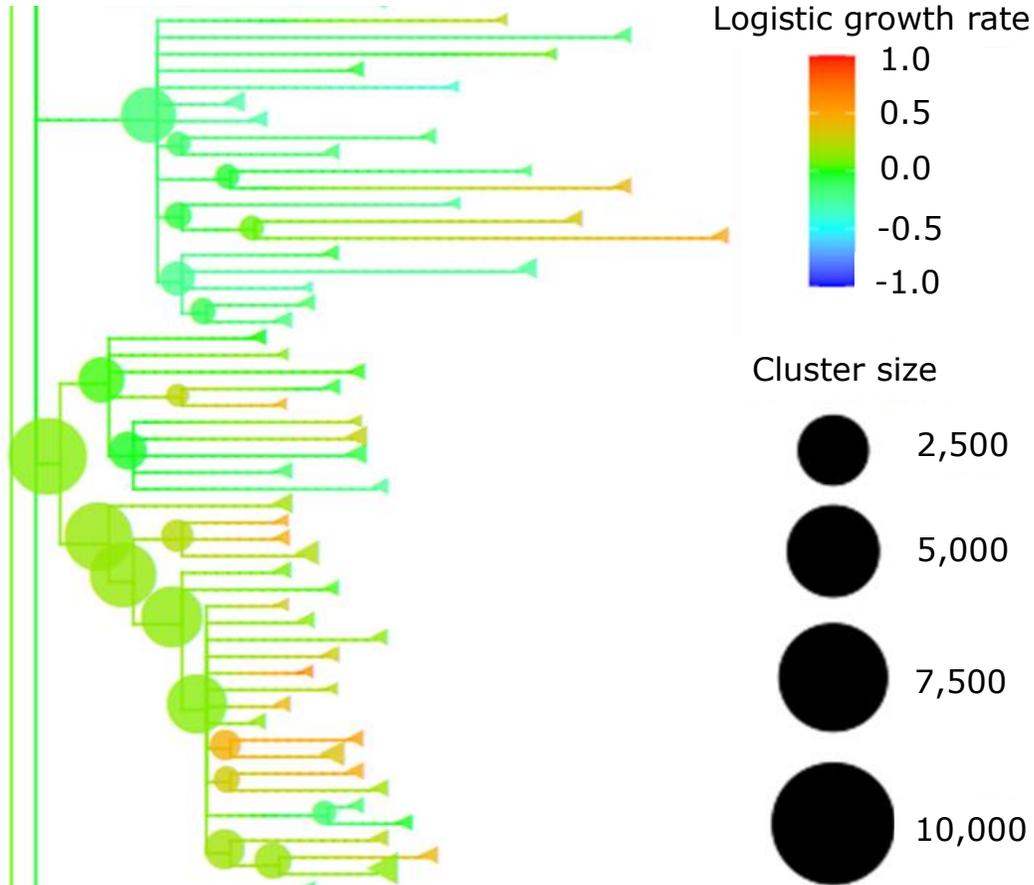
For every node + descendants (cluster):

- Min/max cluster size cut-off
- Min/max sample dates cut-off
- Matched clusters on region and time

Run statistical analyses + support

- Relative evolutionary estimates
- Relative growth rate estimates
- Treeviewer and report system





1. Logistic growth rate (LGR):

- GLM to calculate logistic odds of sample being from cluster vs matched sample (time, geographic, prevalence).
- Logistic odds x mean generation time = relative growth rate per generation time period for each cluster of interest

2. Growth over time:

- GAM + Gaussian process model.

3. Growth over time and space:

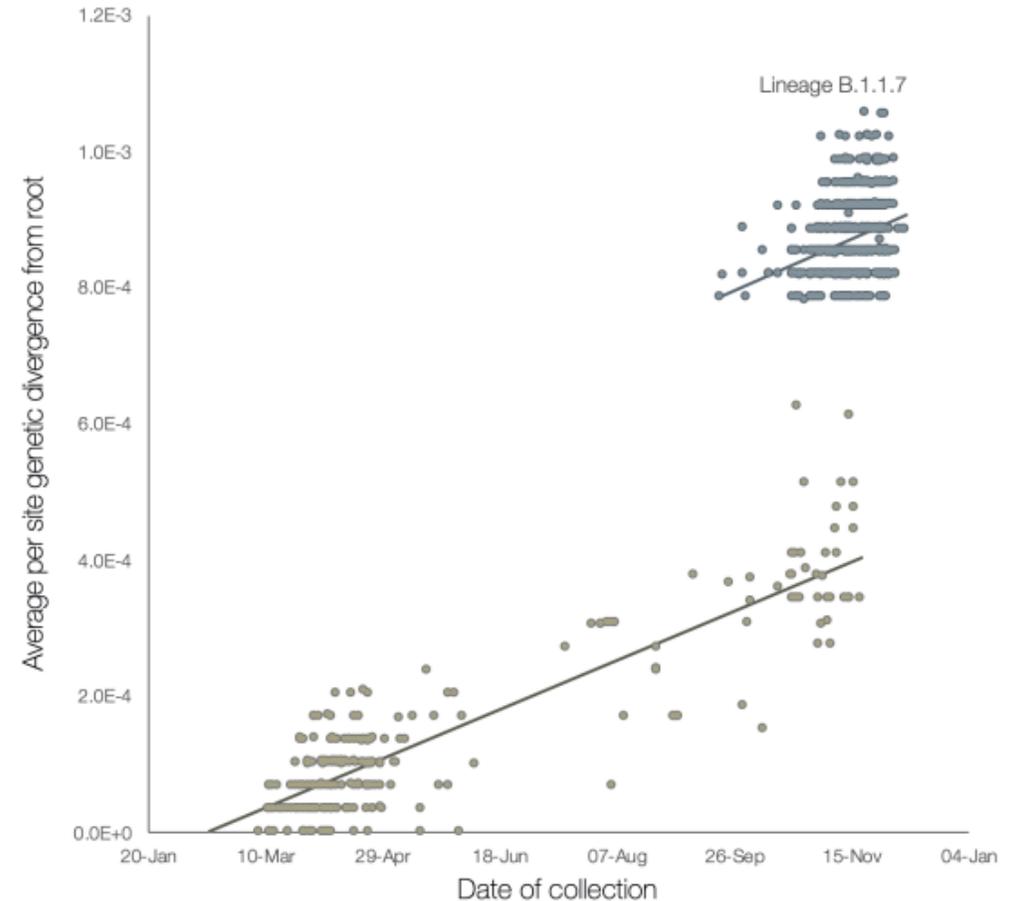
- GAM + Gaussian Markov random model
- Spatial correlation between neighbouring lower-tier local authorities (LTLAs)

Molecular clock outlier statistic



7

- *Actual vs expected evolutionary rates*
- *Root-to-tip regression*

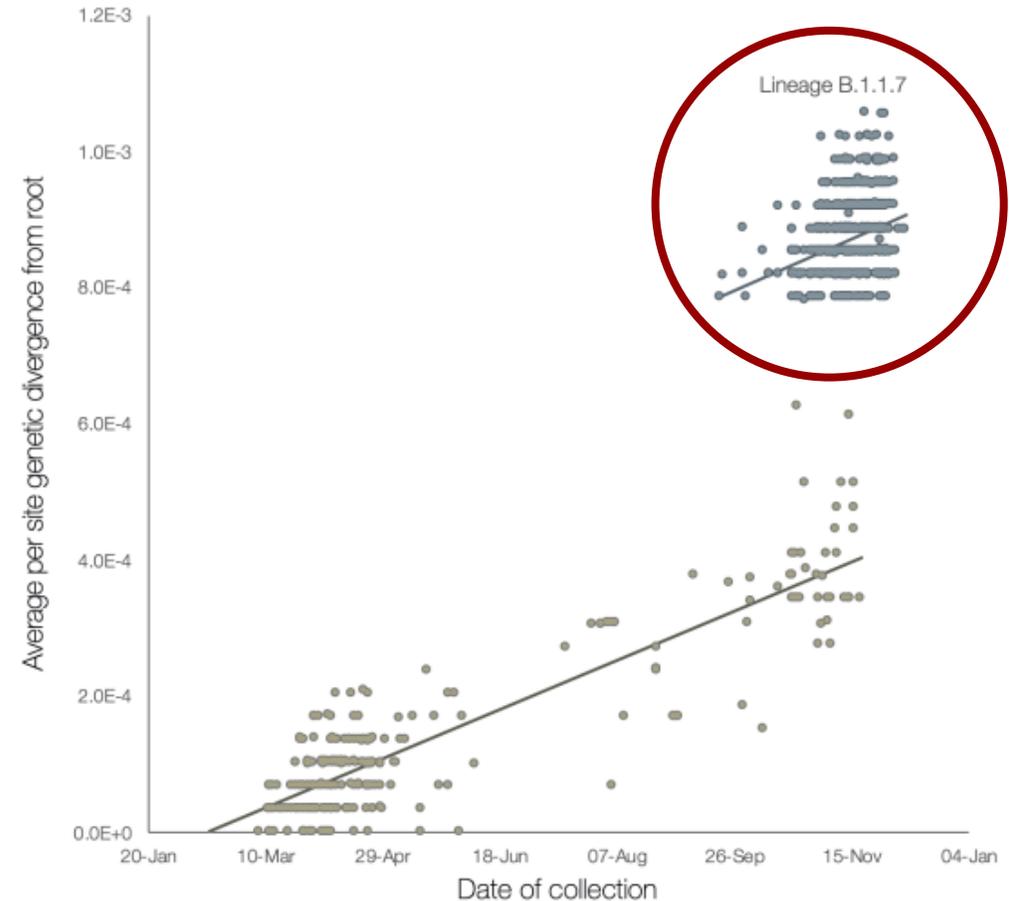


Rambaut, Loman, Pybus, Barclay, Barrett, Carabelli, Connor, Peacock, Robertson, Volz, CoG-UK (2020) Preliminary genomic characterisation of an emergent SARS-CoV-2 lineage in the UK defined by a novel set of spike mutations. virological.org

Molecular clock outlier statistic

7

- *Actual vs expected evolutionary rates*
- *Root-to-tip regression*



Rambaut, Loman, Pybus, Barclay, Barrett, Carabelli, Connor, Peacock, Robertson, Volz, CoG-UK (2020) Preliminary genomic characterisation of an emergent SARS-CoV-2 lineage in the UK defined by a novel set of spike mutations. virological.org

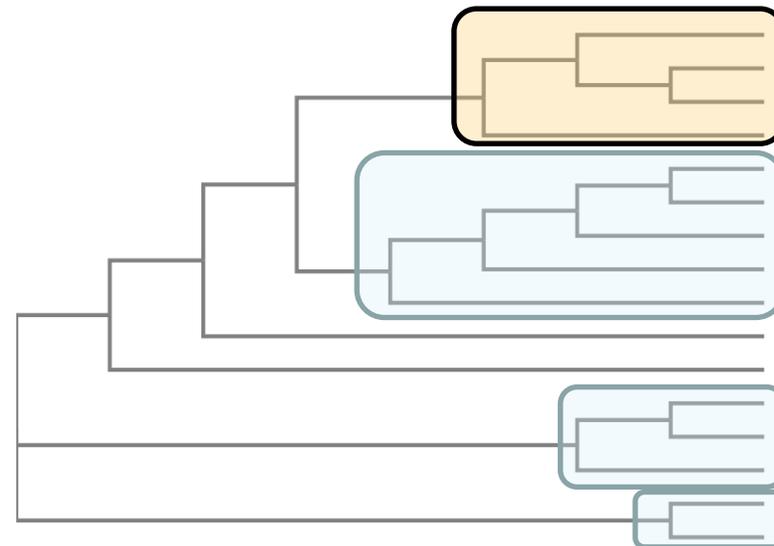
- Sequences
- TFP Scanner parameters
- Cluster filters
- Leading indicator types
- EWS thresholds

- UK pillar 2 only

Parameters investigated

- Sequences **x 1**
- **TFP Scanner parameters**
- Cluster filters
- Leading indicator types
- EWS thresholds

- Min cluster age (7, 14, 28 days)
- Max cluster age (56, 84 days)
- Min descendants (20, 50, 100, % of samples)



- Sequences **x 1**
- TFP Scanner parameters **x24**
- **Cluster filters**
- Leading indicator types
- EWS thresholds

- Only extant clusters (latest sample within 14 days of tree date)
- No overlapping clusters
- Only external clusters / replacement by parent cluster, based on parent cluster LGR being at least 60%-100% (5% increments) of max sub-cluster LGR
- LGR p-values (All, <0.05, <0.01)

Parameters investigated



8

- Sequences **x 1**
- TFP Scanner parameters **x24**
- Cluster filters **x30**
- **Leading indicator types**
- EWS thresholds

- LGR max, mean, wtd mean
- Simple LGR max, mean, wtd mean
- GAM LGR max, mean, wtd mean
- Molecular clock outlier max and mean
- Dominant Pango lineage max LGR
- Variance of LGR, simple LGR and GAM LGR
- $\text{Var}(\text{LGR wtd}) / \text{mean cluster size}$

- Sequences **x 1**
- TFP Scanner parameters **x24**
- Cluster filters **x30**
- **Leading indicator types**
- EWS thresholds

Fisher's Fundamental Theorem of Natural Selection

“The rate of increase in fitness of any organism at any time is equal to its genetic variance in fitness at that time.”

- Sequences **x 1**
- TFP Scanner parameters **x24**
- Cluster filters **x30**
- **Leading indicator types**
- EWS thresholds

Applied to COVID-19 pandemic

The rate of change in SC2

transmissibility

$$(e.g. \frac{dR_t}{dt} \sim \frac{d^2 \text{hospitalisations}}{dt^2})$$

could be linked to the

variance in logistic growth rates

among SC2 phylogenetic clades

Parameters investigated



8

- Sequences **x 1**
- TFP Scanner parameters **x24**
- Cluster filters **x30**
- Leading indicator types **x19**
- **EWS thresholds**

- 'Robust' Z score = 0 to 5 in increments of 0.05

Parameters investigated

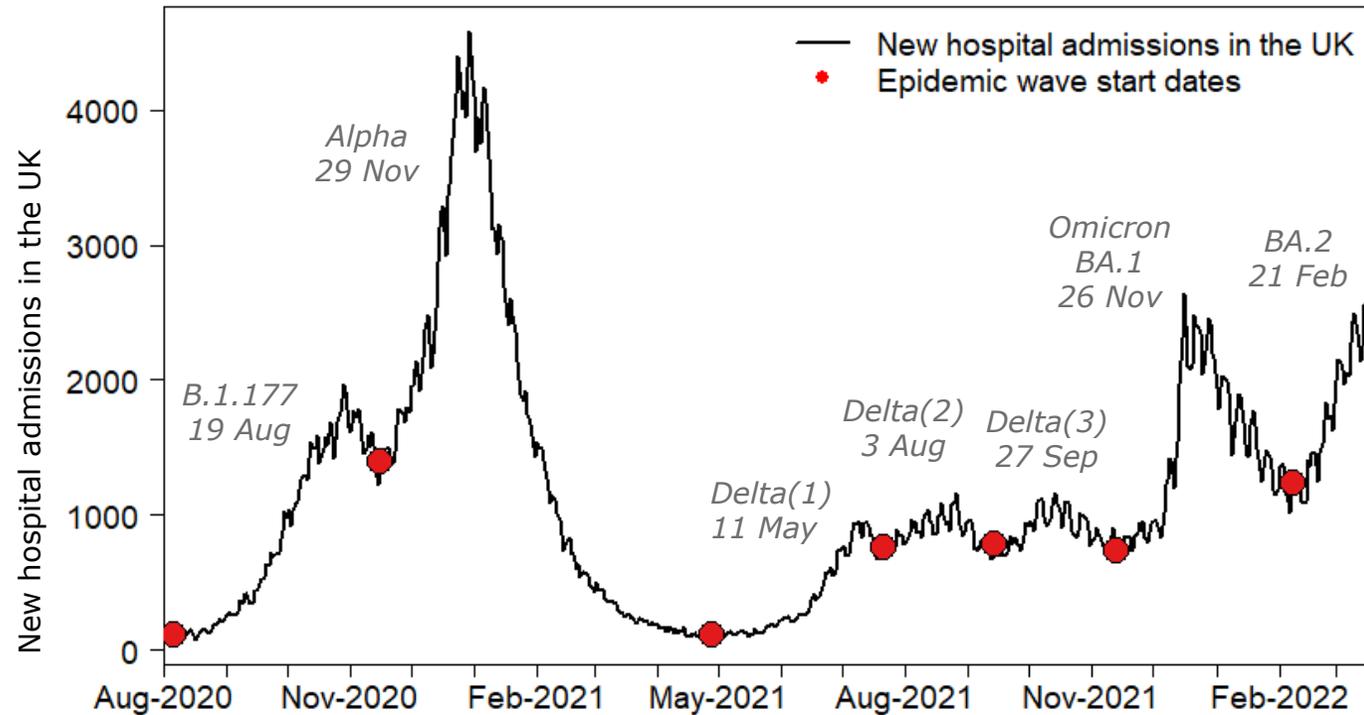


8

- Sequences **x 1**
- TFP Scanner parameters **x24**
- Cluster filters **x30**
- Leading indicator types **x19**
- EWS thresholds **x101**

1.38 million unique parameter sets

EWS Lead Time - defining wave start dates



- GAM optimised for smoothing function and basis dimension
- Low resolution filter applied so focused on new variant driven waves

Classifying EWS as True or False Positive

True Positive



	Dominant variant in cluster	Number of clusters
1	Variant X	49
2	Variant Y	22
3	Variant Z	10
4	Variant X2	7
5	Variant Y2	5
6	Variant Z2	3
...

False Positive

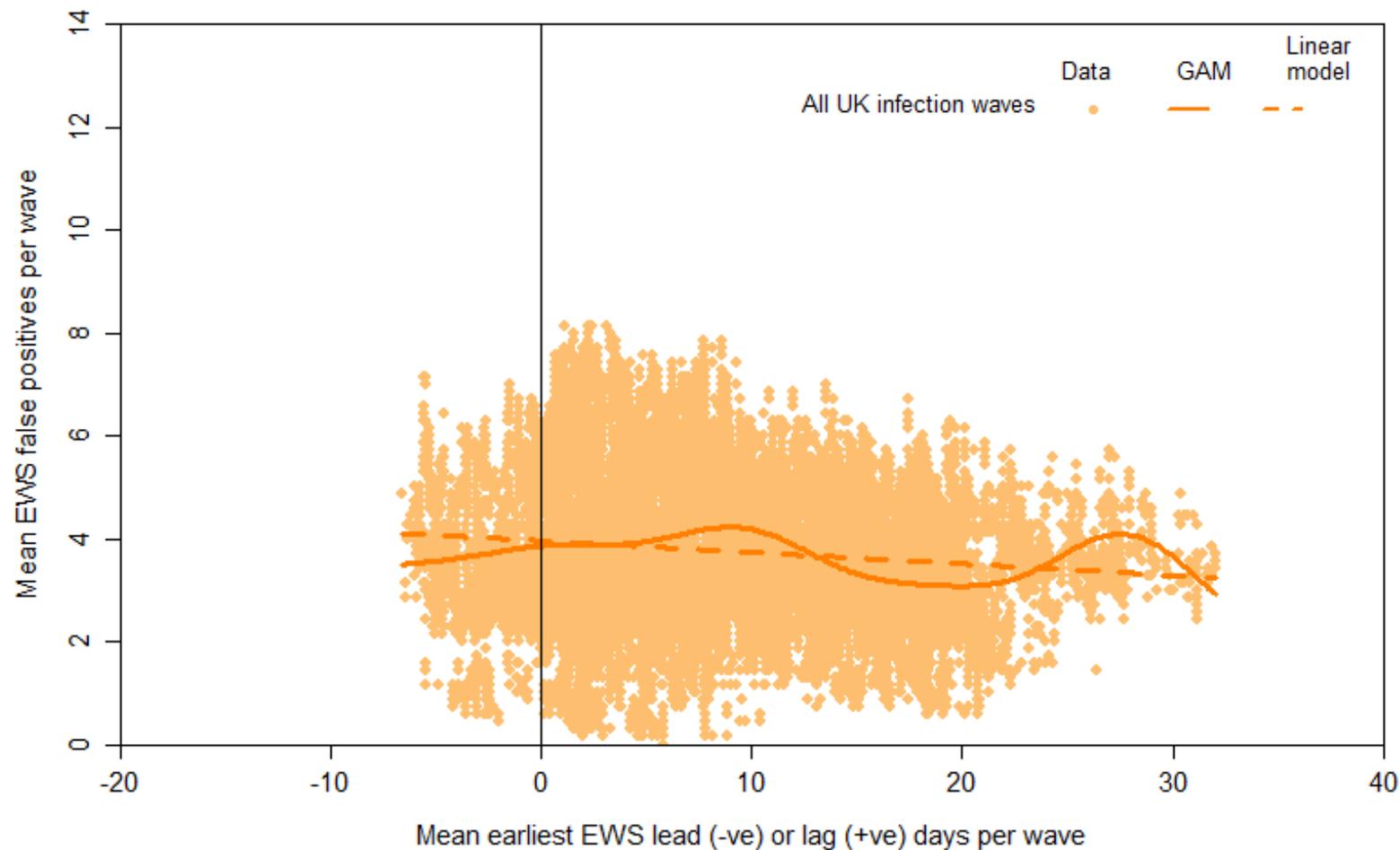


	Dominant variant in cluster	Growth rate
1	Variant X	2.1
2	Variant Y	1.5
3	Variant Z	1.4
4	Variant X2	1.2
5	Variant Y2	1.0
6	Variant Z2	0.8
...



The trade off

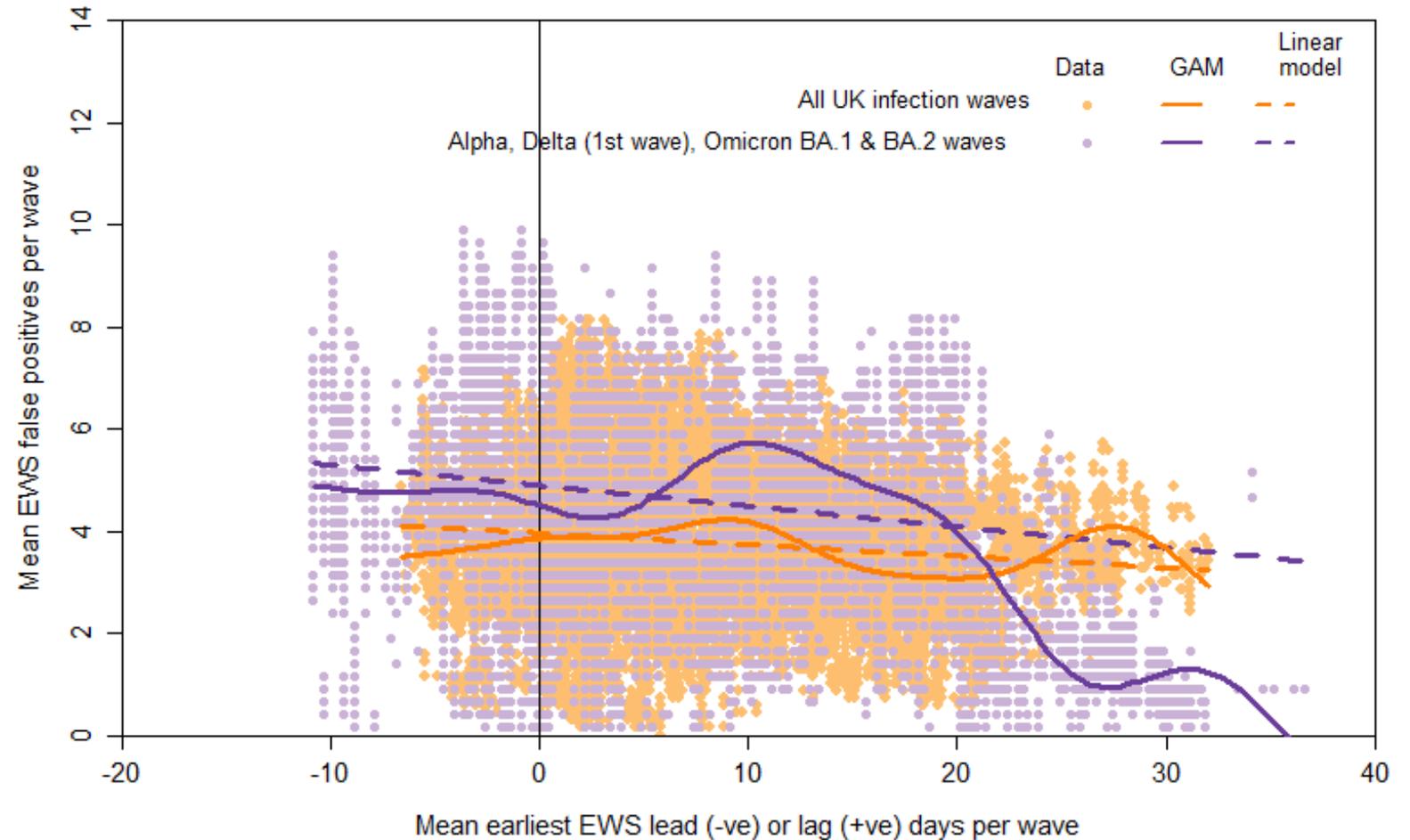
Earlier EWS,
more false positives



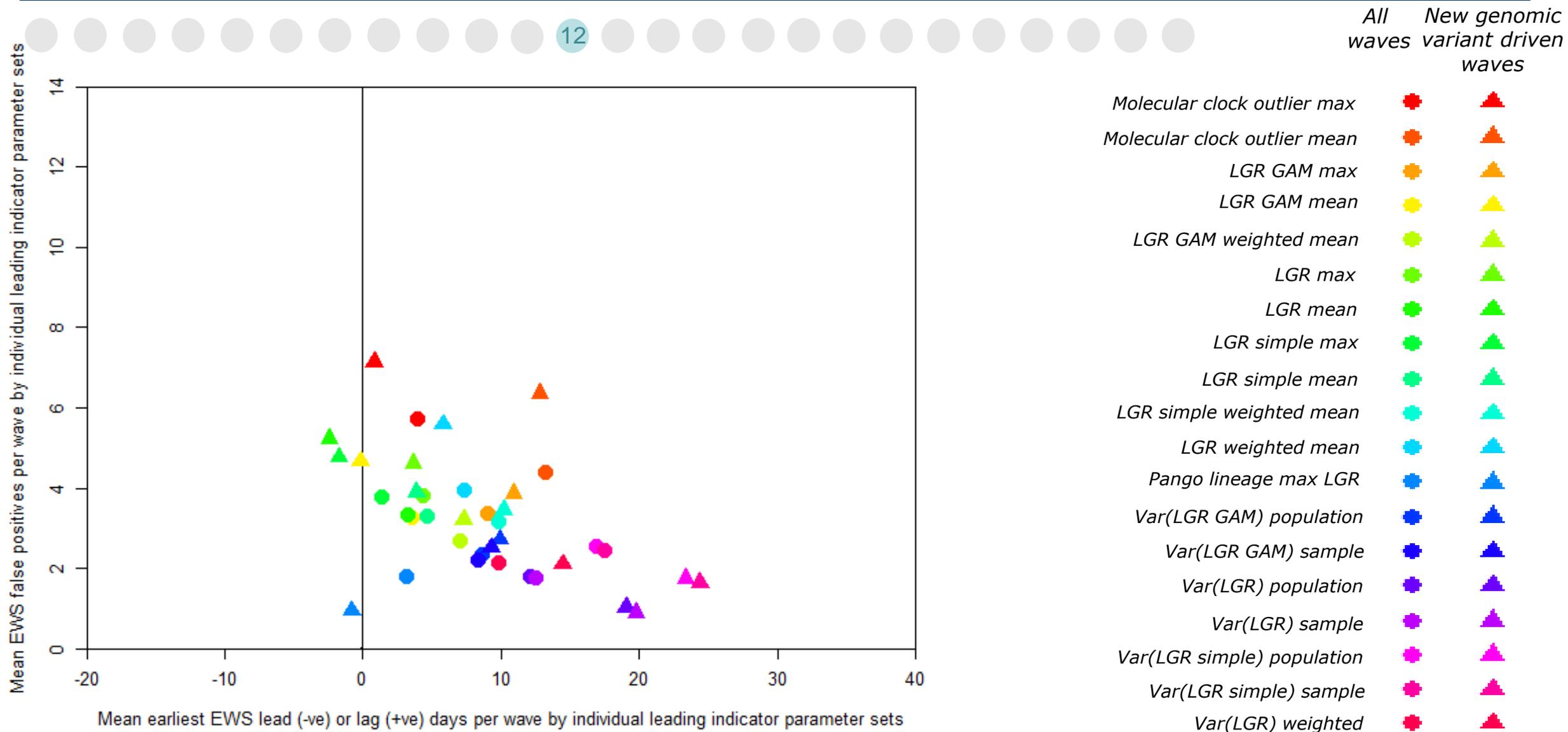
The trade off

Earlier EWS,
more false positives

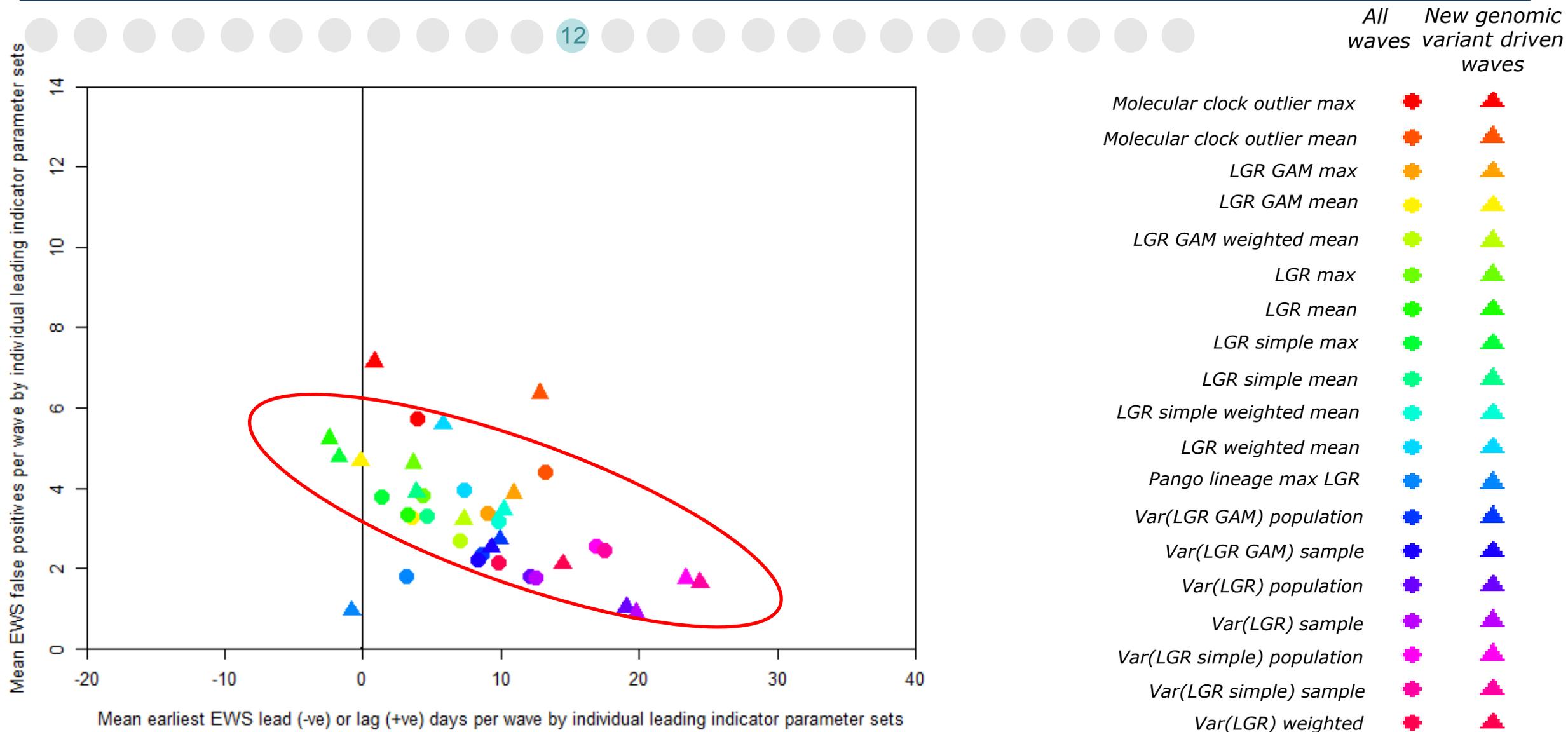
Better lead times for
waves driven by new
genomic variants



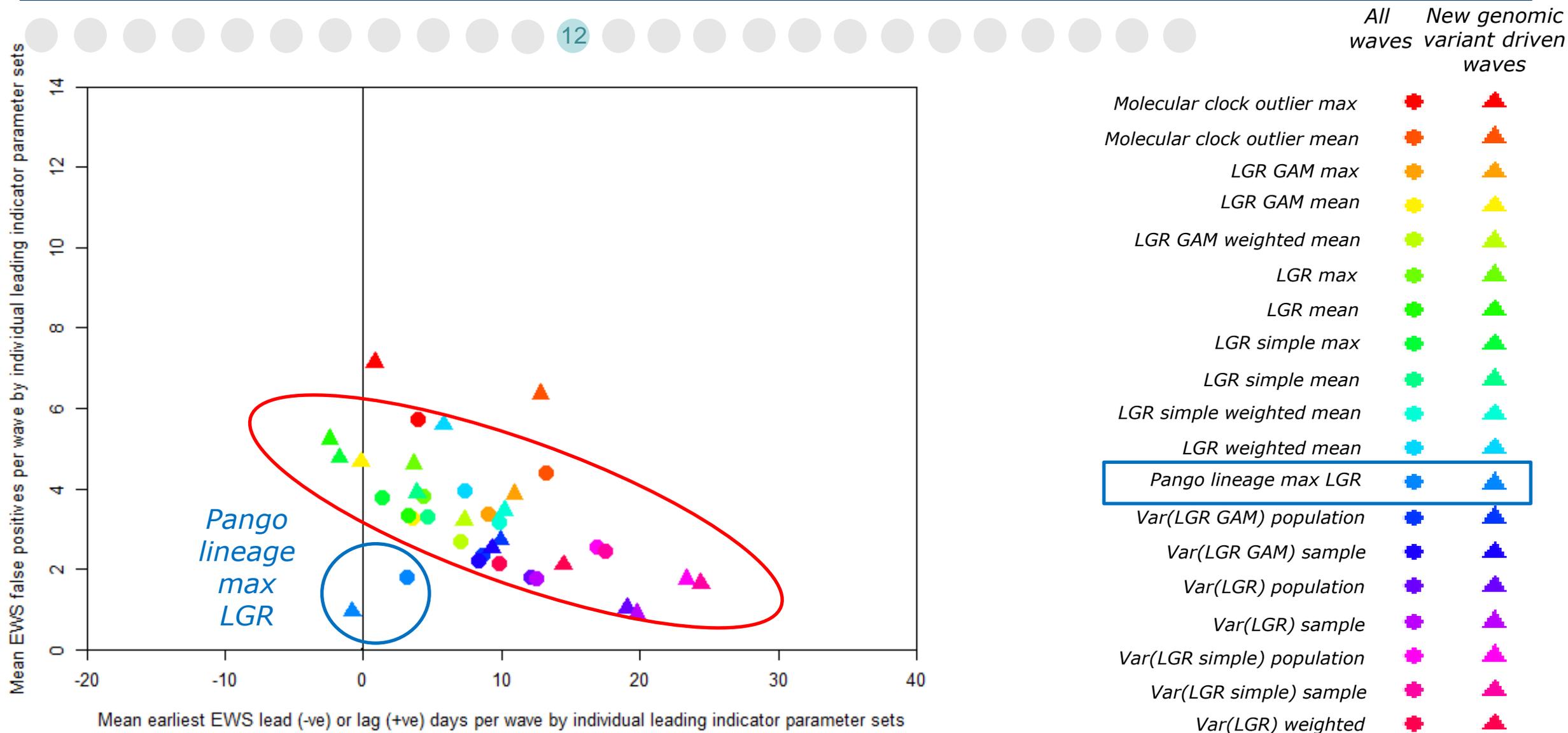
Separate leading indicators



Separate leading indicators

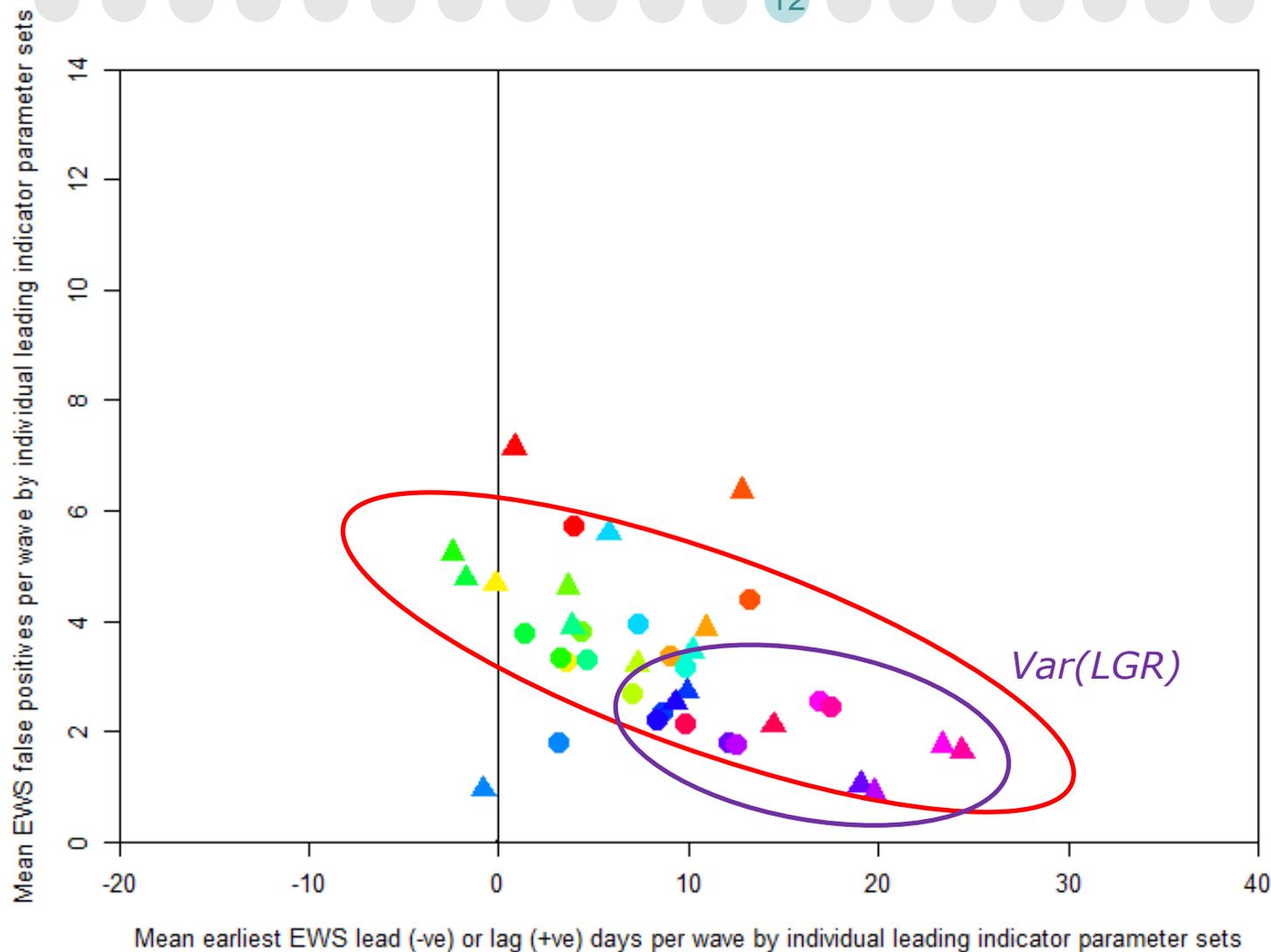


Separate leading indicators



Separate leading indicators

All waves New genomic
variant driven
waves

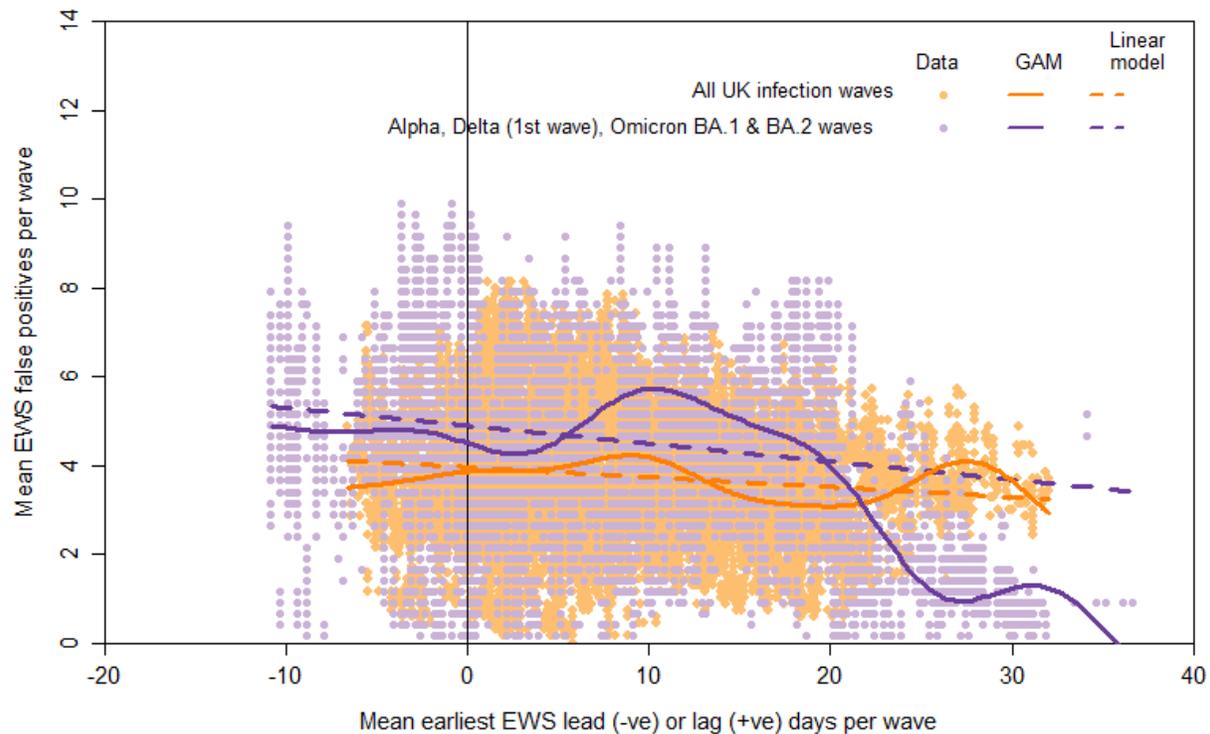


<i>Molecular clock outlier max</i>		
<i>Molecular clock outlier mean</i>		
<i>LGR GAM max</i>		
<i>LGR GAM mean</i>		
<i>LGR GAM weighted mean</i>		
<i>LGR max</i>		
<i>LGR mean</i>		
<i>LGR simple max</i>		
<i>LGR simple mean</i>		
<i>LGR simple weighted mean</i>		
<i>LGR weighted mean</i>		
<i>Pango lineage max LGR</i>		
<i>Var(LGR GAM) population</i>		
<i>Var(LGR GAM) sample</i>		
<i>Var(LGR) population</i>		
<i>Var(LGR) sample</i>		
<i>Var(LGR simple) population</i>		
<i>Var(LGR simple) sample</i>		
<i>Var(LGR) weighted</i>		

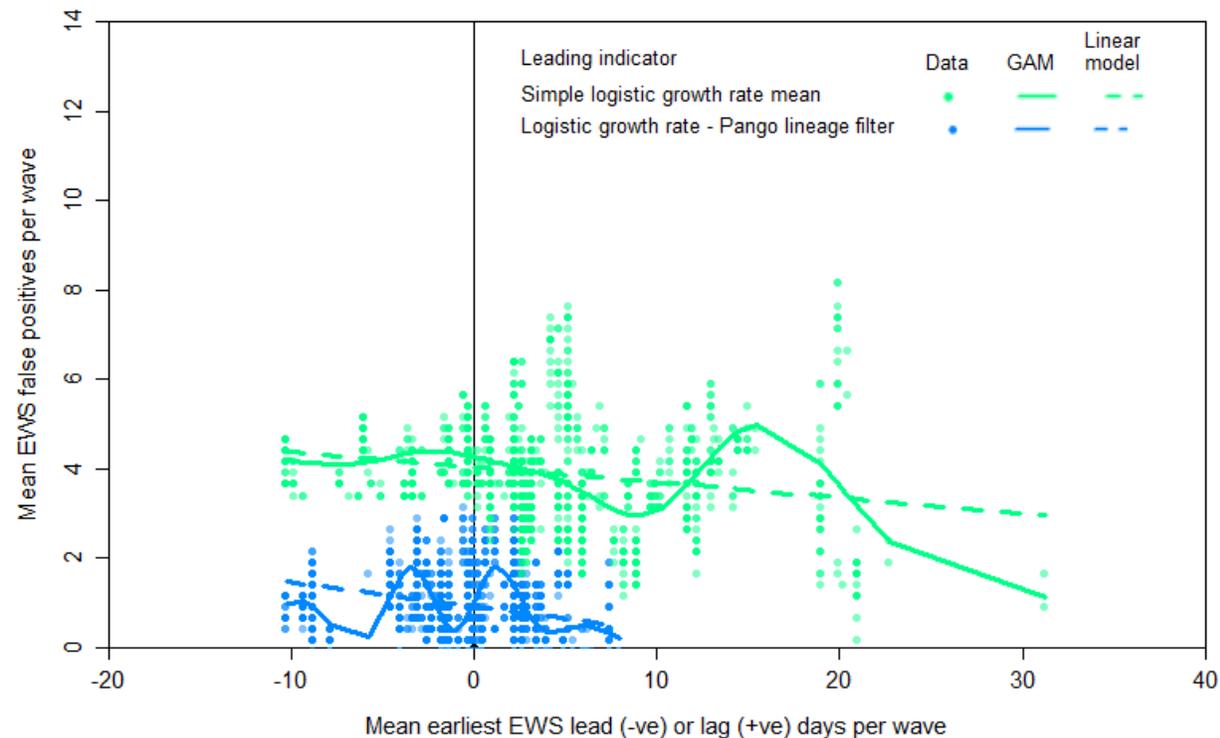
'Best' leading indicators



All leading indicators investigated



'Best' leading indicators



Better EWS for waves driven by new genomic variants?

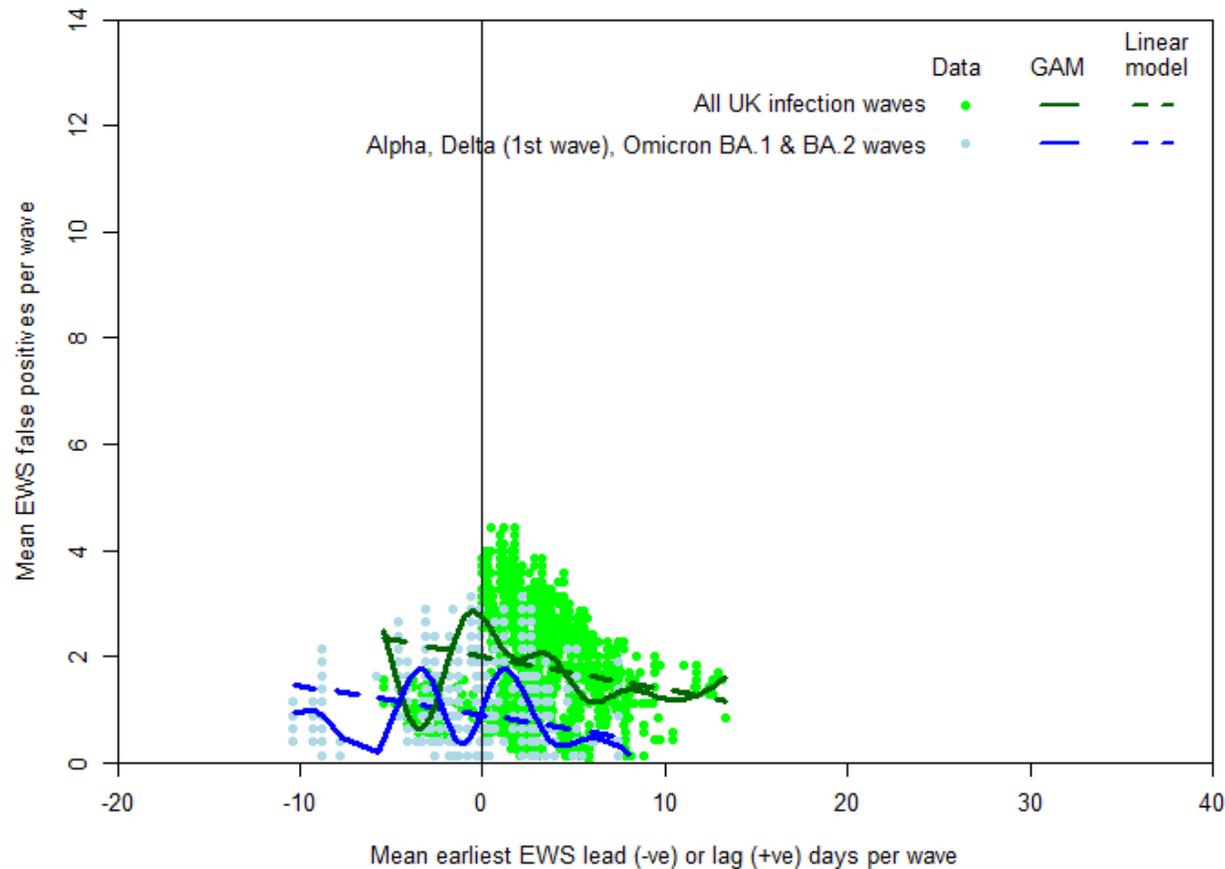


MRC Centre for
Global Infectious
Disease Analysis

Imperial College
London

14

Dominant Pango lineage max LGR



Ranking parameter sets



MRC Centre for
Global Infectious
Disease Analysis

Imperial College
London



B.1.177

Alpha

Delta(1)

Delta(2)

Delta(3)

Omicron

BA.1

Omicron

BA.2

True Positive EWS



Ranking parameter sets



MRC Centre for
Global Infectious
Disease Analysis

Imperial College
London



B.1.177

Alpha

Delta(1)

Delta(2)

Delta(3)

Omicron

BA.1

Omicron

BA.2

True Positive EWS



40,720

Ranking parameter sets



MRC Centre for
Global Infectious
Disease Analysis

Imperial College
London



B.1.177

Alpha

Delta(1)

Delta(2)

Delta(3)

Omicron

BA.1

Omicron

BA.2

True Positive EWS



43,059

Ranking parameter sets



MRC Centre for
Global Infectious
Disease Analysis

Imperial College
London



B.1.177

Alpha

Delta(1)

Delta(2)

Delta(3)

Omicron

BA.1

Omicron

BA.2

True Positive EWS



Limit on False Positive EWS (total and before earliest True Positive):

$\in \{ =0, \leq 2, \leq 5, \leq 10, \text{Any} \}$ per wave

Ranking parameter sets



B.1.177

Alpha

Delta(1)

Delta(2)

Delta(3)

Omicron

BA.1

Omicron

BA.2

True Positive EWS



Limit on False Positive EWS (total and before earliest True Positive):

$\in \{ =0, \leq 2, \leq 5, \leq 10, \text{Any} \}$ per wave

Constraint on lead time: Before wave start date or equal to best lead time

Ranking parameter sets



MRC Centre for
Global Infectious
Disease Analysis

Imperial College
London



B.1.177

Alpha

Delta(1)

Delta(2)

Delta(3)

Omicron

BA.1

Omicron

BA.2

True Positive EWS



Limit on False Positive EWS (total and before earliest True Positive):

$\in \{ =0, \leq 2, \leq 5, \leq 10, \text{Any} \}$ per wave

Constraint on lead time: Before wave start date or equal to best lead time

- Sort by total lead time:
- All waves
 - Only waves driven by new genomic variants

Subjective example



	B.1.177	Alpha	Delta(1)	Delta(2)	Delta(3)	Omicron BA.1	Omicron BA.2
Lead time	+7	-6	-20	+7	+2	+6	-17
	+9	-6	-20	+5	+2	+6	-17

Subjective example



	B.1.177	Alpha	Delta(1)	Delta(2)	Delta(3)	Omicron BA.1	Omicron BA.2
Lead time	+7	-6	-20	+7	+2	+6	-17
	+9	-6	-20	+5	+2	+6	-17
False positives	2	0	2	0	0	2	0
	2	0	0	0	0	2	0

Subjective example



		B.1.177	Alpha	Delta(1)	Delta(2)	Delta(3)	Omicron BA.1	Omicron BA.2
Lead time		+7	-6	-20	+7	+2	+6	-17
		+9	-6	-20	+5	+2	+6	-17
False positives		2	0	2	0	0	2	0
		2	0	0	0	0	2	0

'Best' parameter sets



MRC Centre for
Global Infectious
Disease Analysis

Imperial College
London



- 36 unique filtering and ranking criteria

'Best' parameter sets



- 36 unique filtering and ranking criteria
- 20 yielded results (with leading indicator parameter sets ranging 19 to 40,720)

'Best' parameter sets

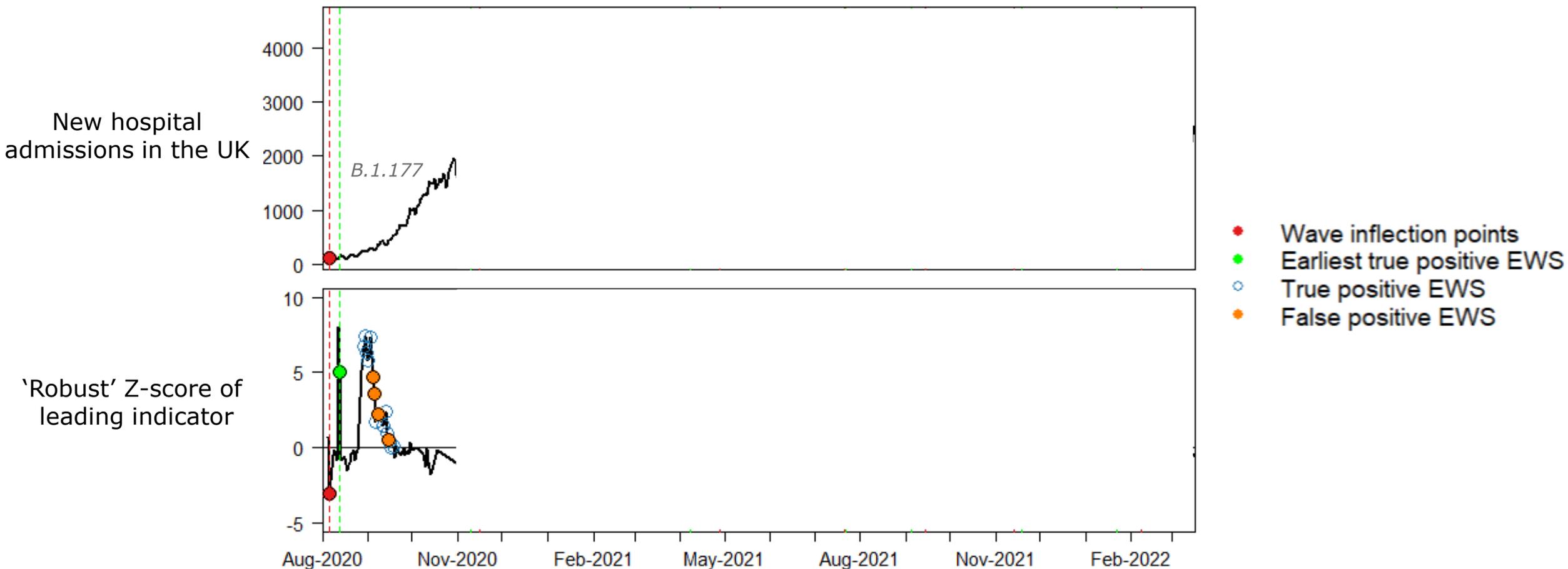


- 36 unique filtering and ranking criteria
- 20 yielded results (with leading indicator parameter sets ranging 19 to 40,720)
- 14 where Dominant Pango lineage max LGR was the 'best'

- 36 unique filtering and ranking criteria
- 20 yielded results (with leading indicator parameter sets ranging 19 to 40,720)
- 14 where Dominant Pango lineage max LGR was the 'best'
- 9 with the same TFP scanner parameters and cluster filters:
 - Cluster age min 7 days, max 56 days
 - Min descendants 20
 - Cluster LGR p-value < 0.01
 - Parent/sub-cluster replacement LGR threshold 85%
 - EWS threshold 0.00

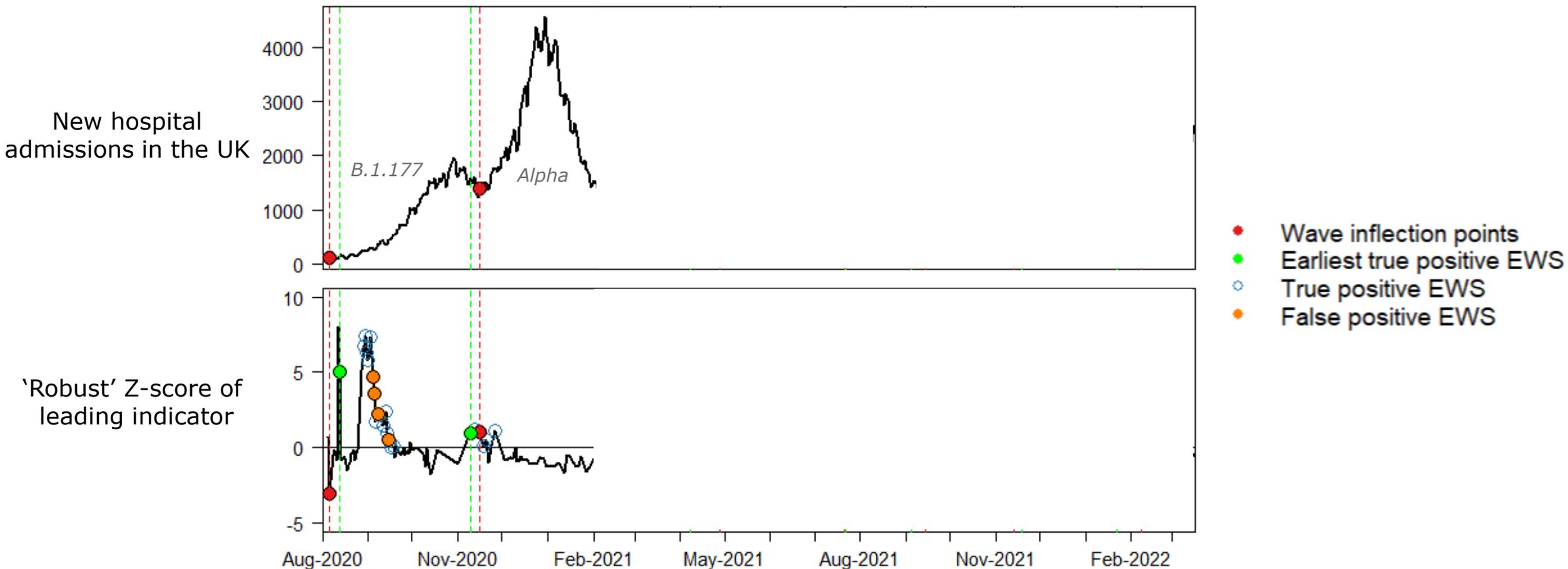
- 36 unique filtering and ranking criteria
- 20 yielded results (with leading indicator parameter sets ranging 19 to 40,720)
- 14 where Dominant Pango lineage max LGR was the 'best'
- 9 with the same TFP scanner parameters and cluster filters:
 - Cluster age min 7 days, max 56 days
 - Min descendants 20
 - Cluster LGR p-value < 0.01
 - Parent/sub-cluster replacement LGR threshold 85%
 - EWS threshold 0.00
- Low number of false positive EWS: 8 in total, only 2 before earliest TP EWS

'Best' leading indicator



Leading indicator shown is the maximum logistic growth rate filtered by Pango lineage, with min cluster age of 7 days, max cluster age of 56 days, min descendants of 20, LGR p-value <0.01, parent/sub-cluster replacement LGR threshold is 85% and EWS threshold is 0.

'Best' leading indicator

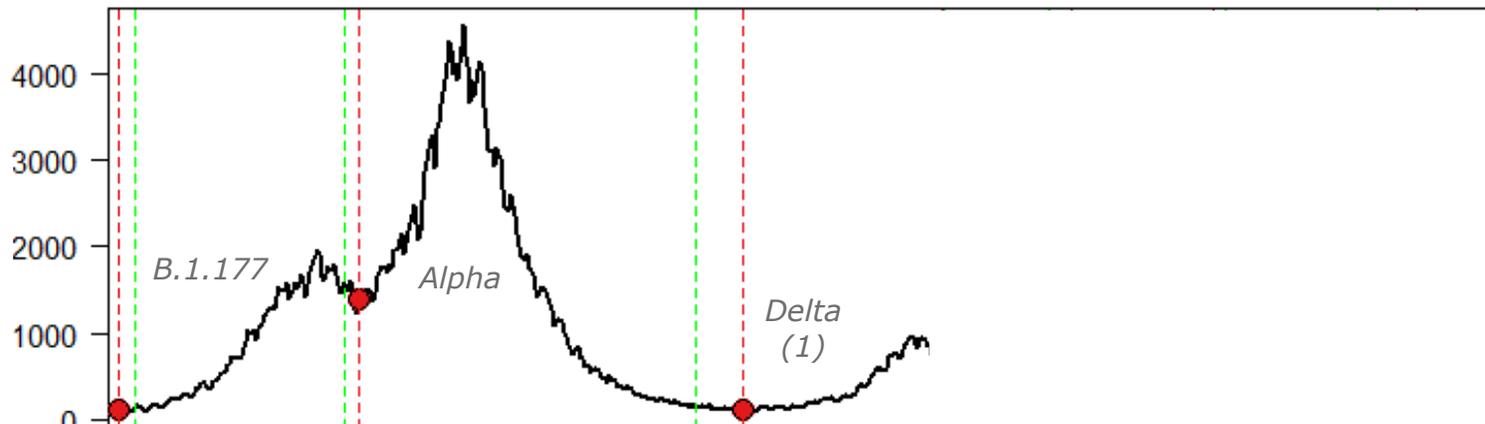


Leading indicator shown is the maximum logistic growth rate filtered by Pango lineage, with min cluster age of 7 days, max cluster age of 56 days, min descendants of 20, LGR p-value <0.01, parent/sub-cluster replacement LGR threshold is 85% and EWS threshold is 0.

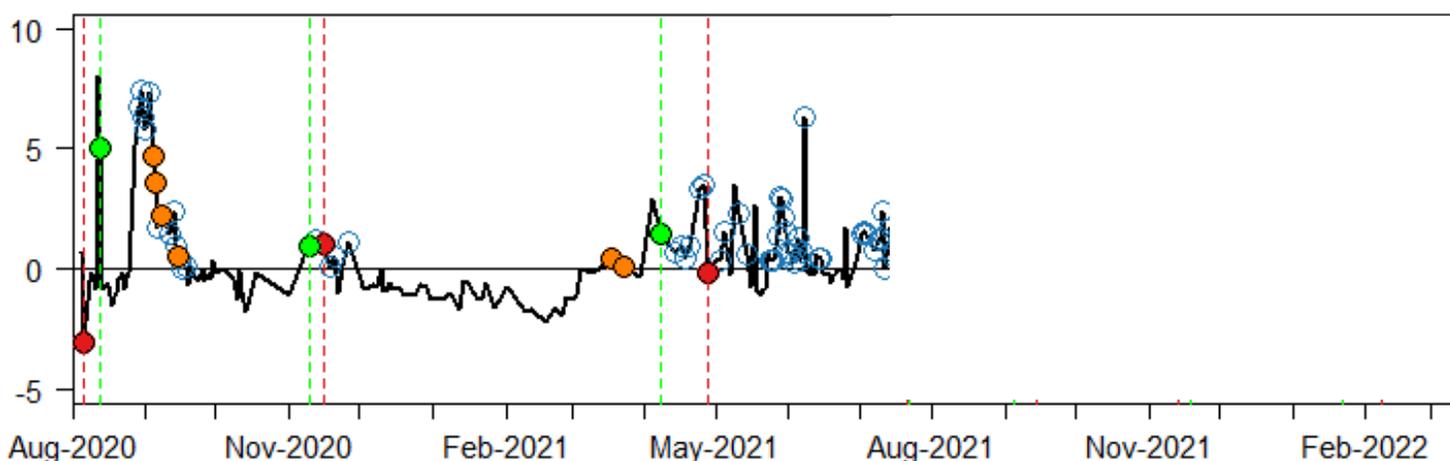
'Best' leading indicator



New hospital
admissions in the UK



'Robust' Z-score of
leading indicator



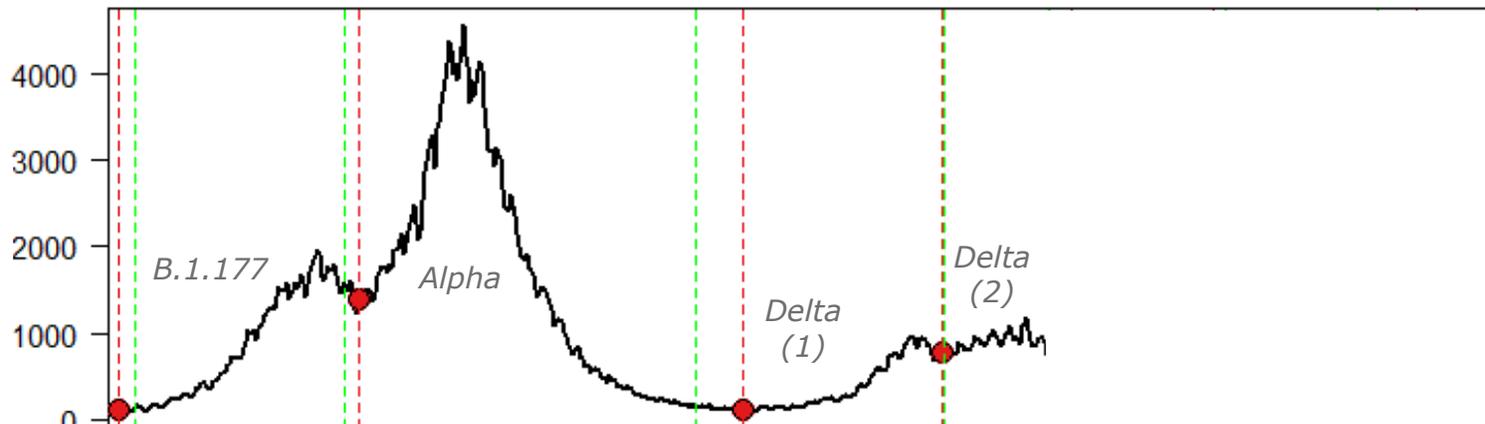
- Wave inflection points
- Earliest true positive EWS
- True positive EWS
- False positive EWS

Leading indicator shown is the maximum logistic growth rate filtered by Pango lineage, with min cluster age of 7 days, max cluster age of 56 days, min descendants of 20, LGR p-value <0.01, parent/sub-cluster replacement LGR threshold is 85% and EWS threshold is 0.

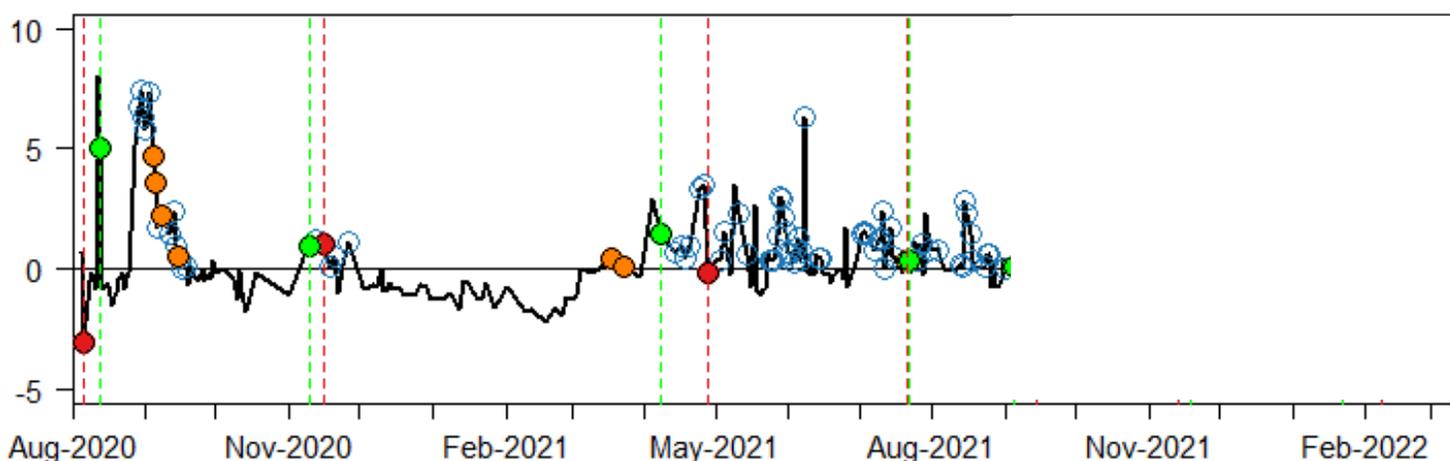
'Best' leading indicator



New hospital
admissions in the UK



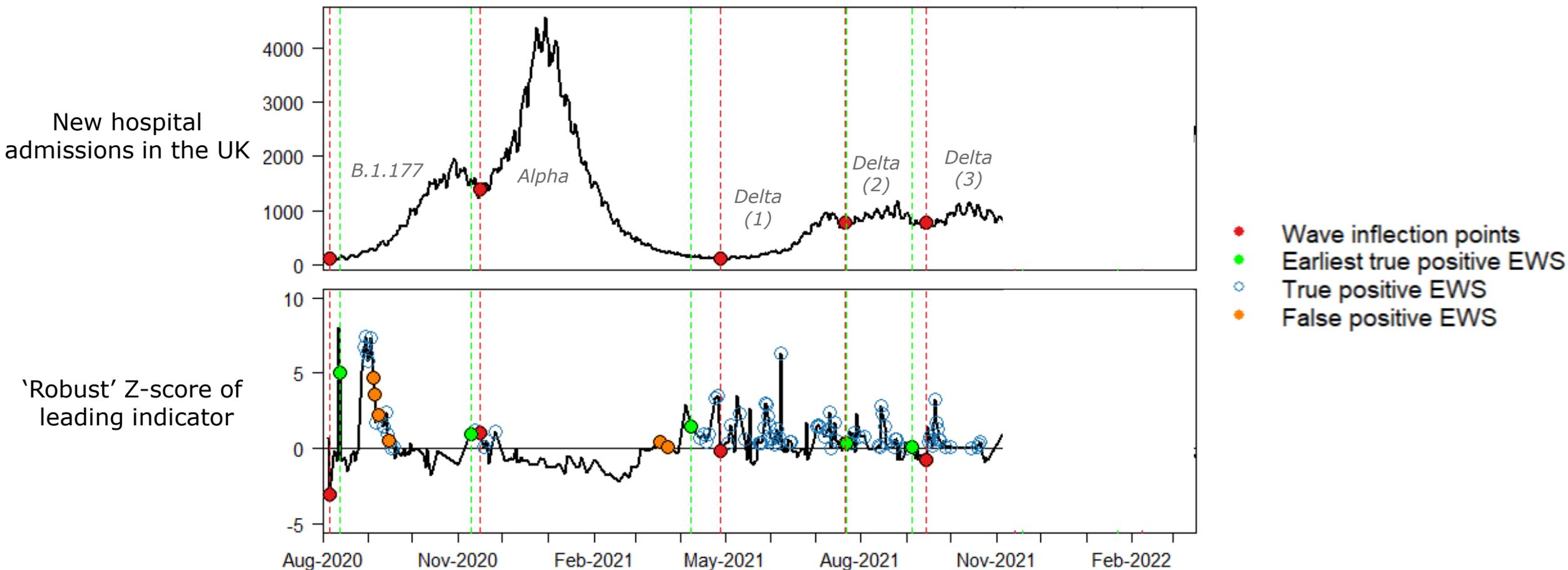
'Robust' Z-score of
leading indicator



- Wave inflection points
- Earliest true positive EWS
- True positive EWS
- False positive EWS

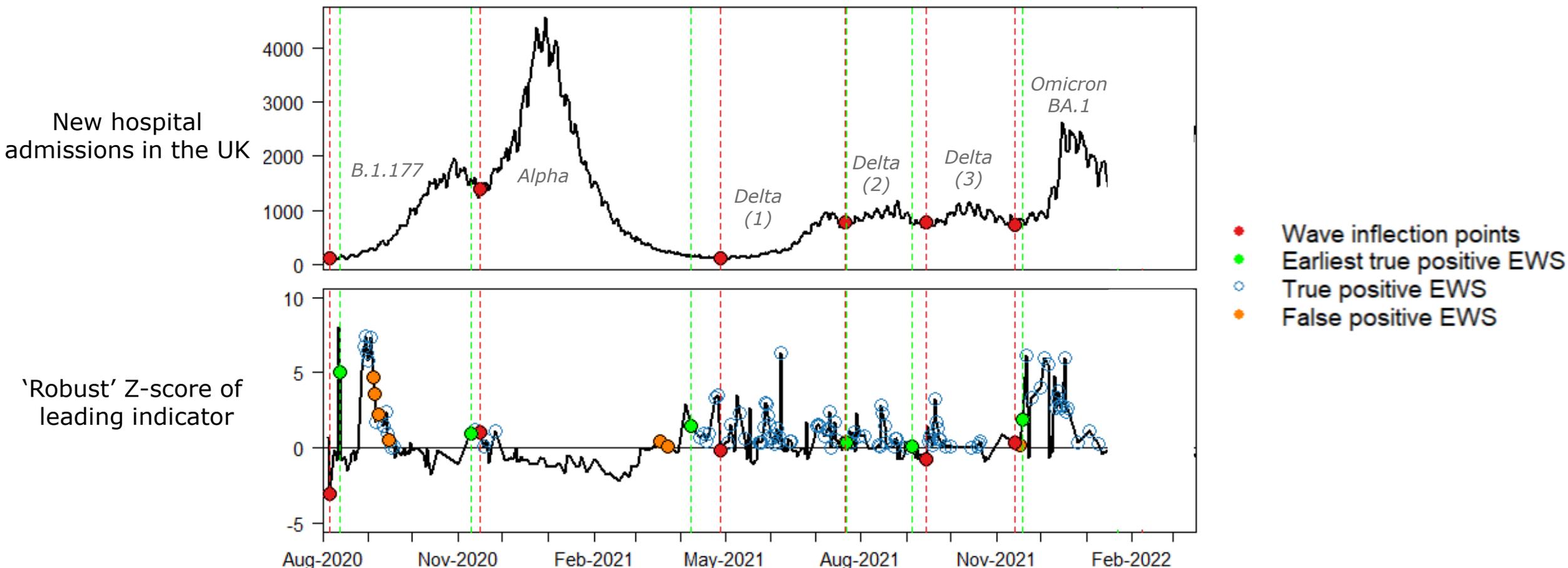
Leading indicator shown is the maximum logistic growth rate filtered by Pango lineage, with min cluster age of 7 days, max cluster age of 56 days, min descendants of 20, LGR p-value <0.01, parent/sub-cluster replacement LGR threshold is 85% and EWS threshold is 0.

'Best' leading indicator



Leading indicator shown is the maximum logistic growth rate filtered by Pango lineage, with min cluster age of 7 days, max cluster age of 56 days, min descendants of 20, LGR p-value <0.01, parent/sub-cluster replacement LGR threshold is 85% and EWS threshold is 0.

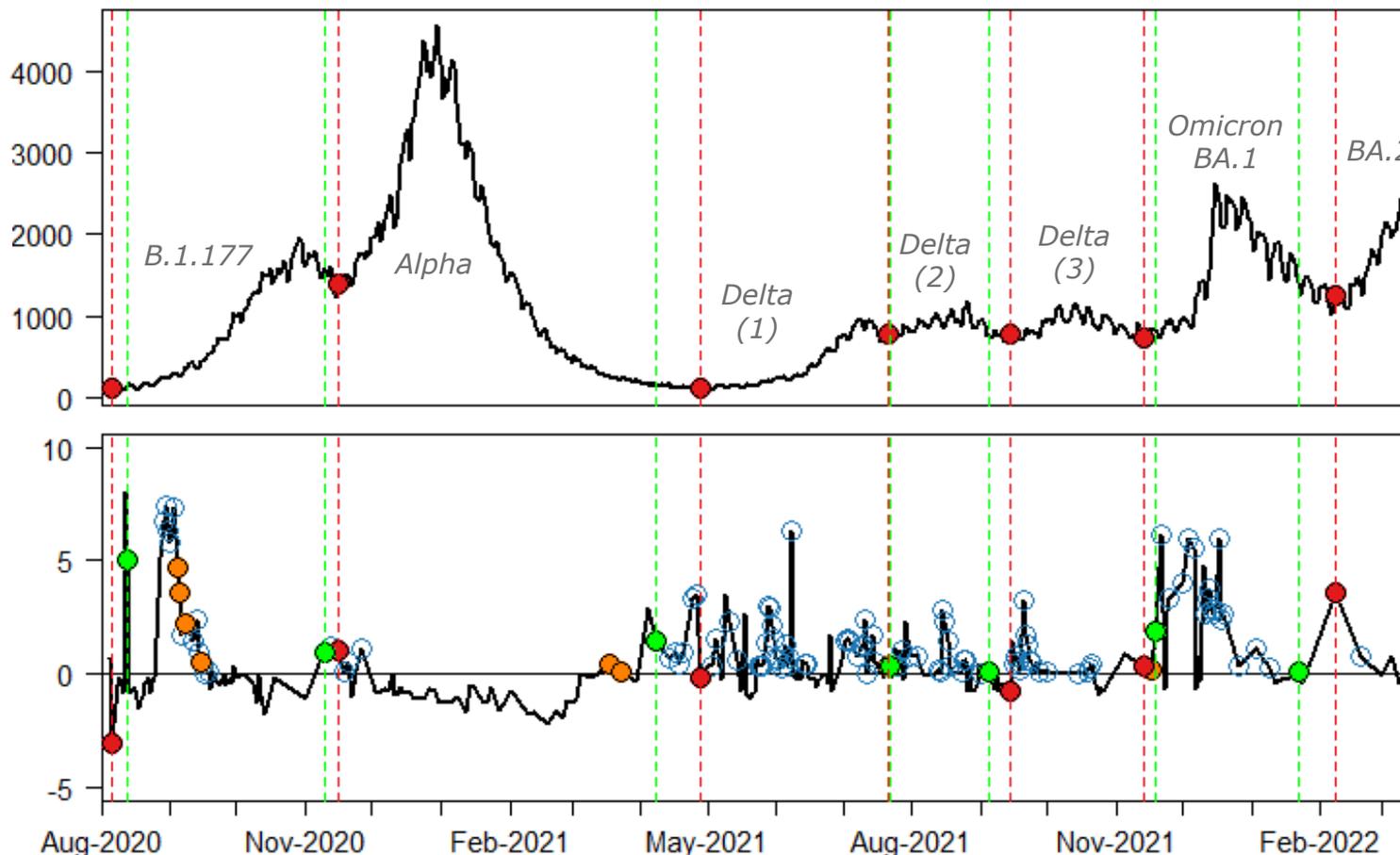
'Best' leading indicator



Leading indicator shown is the maximum logistic growth rate filtered by Pango lineage, with min cluster age of 7 days, max cluster age of 56 days, min descendants of 20, LGR p-value <0.01, parent/sub-cluster replacement LGR threshold is 85% and EWS threshold is 0.

'Best' leading indicator

New hospital
admissions in the UK



- Wave inflection points
- Earliest true positive EWS
- True positive EWS
- False positive EWS

'Robust' Z-score of
leading indicator

Leading indicator shown is the maximum logistic growth rate filtered by Pango lineage, with min cluster age of 7 days, max cluster age of 56 days, min descendants of 20, LGR p-value <0.01, parent/sub-cluster replacement LGR threshold is 85% and EWS threshold is 0.

'Best' leading indicator - summary



Wave	Wave start (inflection) date	EWS date	EWS lead time (days) Lead (-ve) and Lag (+ve)
B.1.177	19 Aug 2020	26 Aug 2020	+7
Alpha	29 Nov 2020	23 Nov 2020	-6
Delta (1)	11 May 2021	21 Apr 2021	-20
Delta (2)	3 Aug 2021	4 Aug 2021	+1
Delta (3)	27 Sep 2021	18 Sep 2021	-9
Omicron BA.1	26 Nov 2021	2 Dec 2021	+6
Omicron BA.2	21 Feb 2022	4 Feb 2022	-17

Leading indicator shown is the maximum logistic growth rate filtered by Pango lineage, with min cluster age of 7 days, max cluster age of 56 days, min descendants of 20, LGR p-value <0.01, parent/sub-cluster replacement LGR threshold is 85% and EWS threshold is 0.

'Best' leading indicator - summary



Wave	Wave start (inflection) date	EWS date	EWS lead time (days) Lead (-ve) and Lag (+ve)	Different parameter sets for Individual waves
B.1.177	19 Aug 2020	26 Aug 2020	+7	+6
Alpha	29 Nov 2020	23 Nov 2020	-6	-6
Delta (1)	11 May 2021	21 Apr 2021	-20	-24
Delta (2)	3 Aug 2021	4 Aug 2021	+1	-1
Delta (3)	27 Sep 2021	18 Sep 2021	-9	-9
Omicron BA.1	26 Nov 2021	2 Dec 2021	+6	+6
Omicron BA.2	21 Feb 2022	4 Feb 2022	-17	-19

Leading indicator shown is the maximum logistic growth rate filtered by Pango lineage, with min cluster age of 7 days, max cluster age of 56 days, min descendants of 20, LGR p-value <0.01, parent/sub-cluster replacement LGR threshold is 85% and EWS threshold is 0.

'Best' leading indicator - summary



Wave	Wave start (inflection) date	EWS date	EWS lead time (days) Lead (-ve) and Lag (+ve)	Different parameter sets for Individual waves	Simple LGR mean
B.1.177	19 Aug 2020	26 Aug 2020	+7	+6	+6
Alpha	29 Nov 2020	23 Nov 2020	-6	-6	-6
Delta (1)	11 May 2021	21 Apr 2021	-20	-24	-24
Delta (2)	3 Aug 2021	4 Aug 2021	+1	-1	-1
Delta (3)	27 Sep 2021	18 Sep 2021	-9	-9	-9
Omicron BA.1	26 Nov 2021	2 Dec 2021	+6	+6	+6
Omicron BA.2	21 Feb 2022	4 Feb 2022	-17	-19	-17

Mean no. FP EWS

1.1

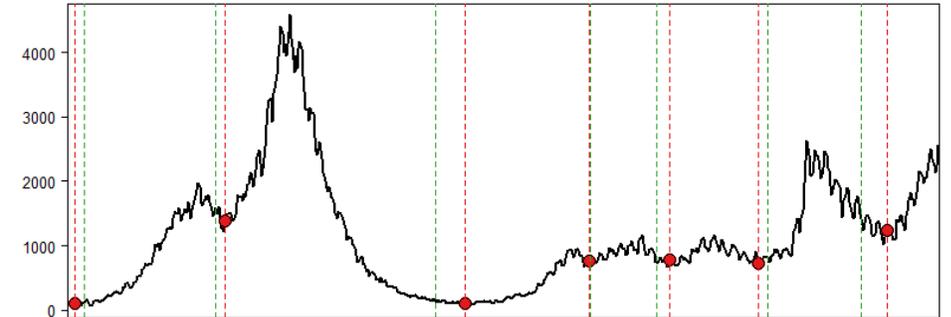
2.9

Leading indicator shown is the maximum logistic growth rate filtered by Dominant Pango lineage, with min cluster age of 7 days, max cluster age of 56 days, min descendants of 20, LGR p-value <0.01, parent/sub-cluster replacement LGR threshold is 85% and EWS threshold is 0.

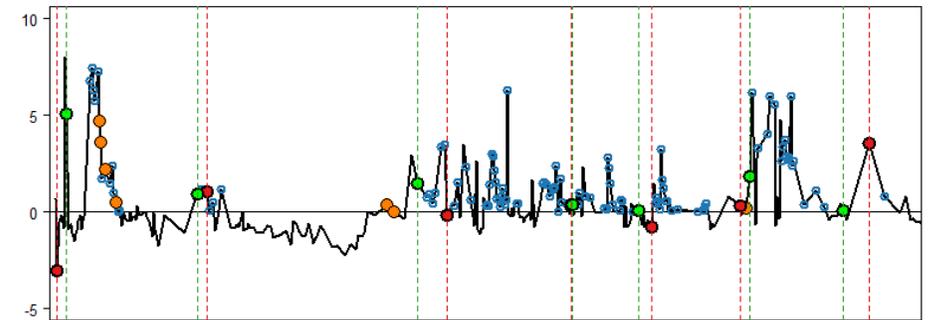


How is the method impacted by sampling density?

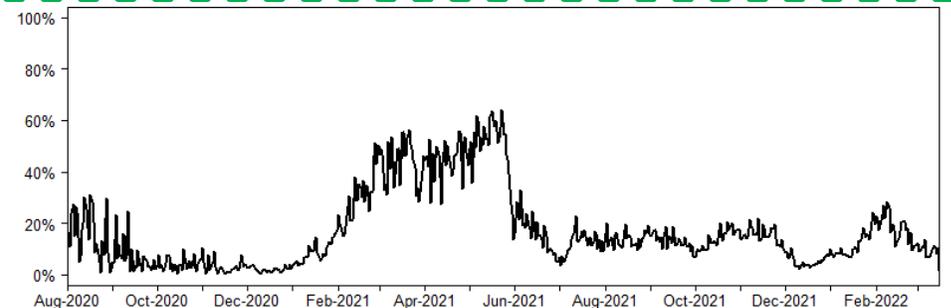
New hospital
admissions in the UK



'Robust' Z-score of
leading indicator



% of recorded UK cases
sequenced



Further work



MRC Centre for
Global Infectious
Disease Analysis

Imperial College
London



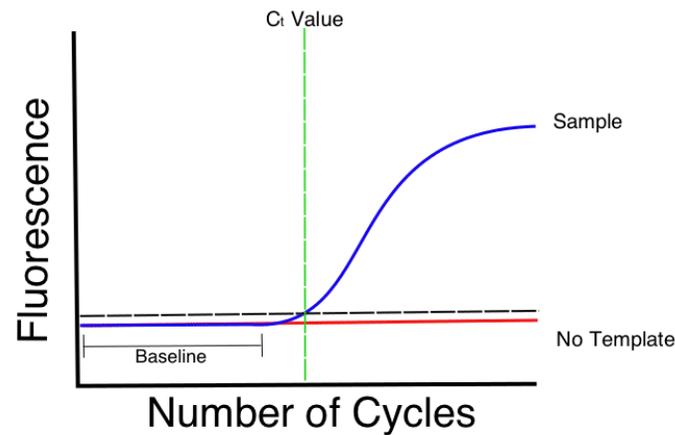
21

How much testing/sequencing
do we need?



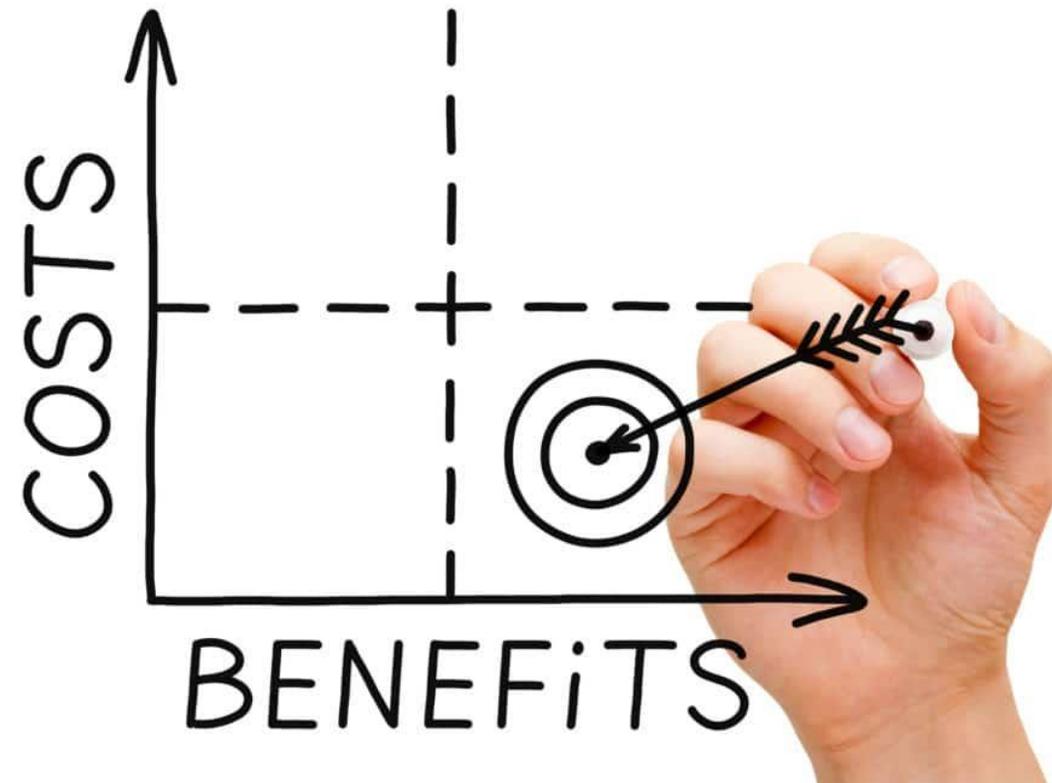


Is it better than other
leading indicators?





Is it cost effective?



Further work



Would it work outside the UK?





Could we apply this method to
other pathogens?



Acknowledgements



MRC Centre for
Global Infectious
Disease Analysis

Imperial College
London

Imperial College London:

Erik Volz



Olivia Boyd



University of Chicago:

Manon Ragonnet-Cronin



Data:

COG-UK, CLIMB, PHE, UK HSA

Funding:

Wellcome Trust