08/07/2021

GIRFT

GETTING IT RIGHT FIRST TIME

DiRAC

A Data-Driven Analysis of Covid-19 Nosocomial Infections

William Gray
Katie Tucker
Maria Marcha
Johannes Heyl
Adrian Hopper
Flavien Hardy
Jeremy Yates

Identifying Nosocomial Infections Using HES

Goals and Motivations

- Dedicated study of Covid-19 NI using HES
- Difficult to identify, but may be important:
    - Such patients may be at risk of **poorer outcomes**
    - Relevant for **other infectious diseases**
    - Health care workers navigate between the community and healthcare settings: may play a **role in the dynamics of the pandemic**

- **Exploring HES**: how can we identify NI from HES?
- **Modelling**: Can we use machine learning to estimate the NI that are likely to have been missed?

Sampling NI Using HES

Method 1:
Use of **code Y95**, relative to the codes U071-U072

Method 2:
Infection more than **15 days** after start of a spell

Method 4:
Z208 code on prior admission ***and*** Emergency readmission within 8 days ***and*** Time between admissions > 8 days

Method 3:
Infection **8-14 days** after start of a spell

ICD-10 **Y95**: nosocomial condition
ICD-10 **Z208**: contact with and exposure to other communicable diseases

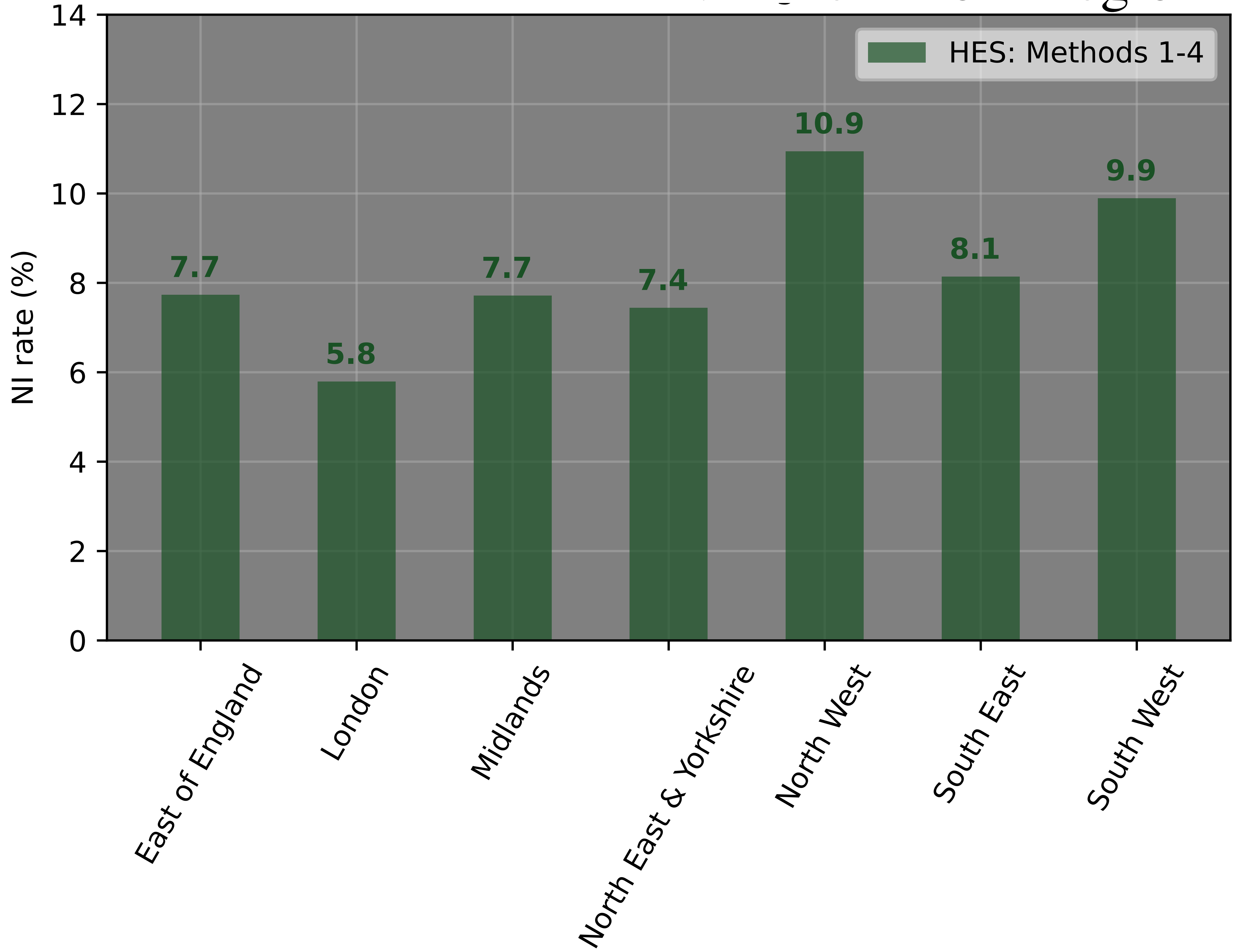
- Methods 1 and 2: definite NI
- Methods 3 and 4: potential NI

Can we use this information to obtain an informed estimate of the NI rate?

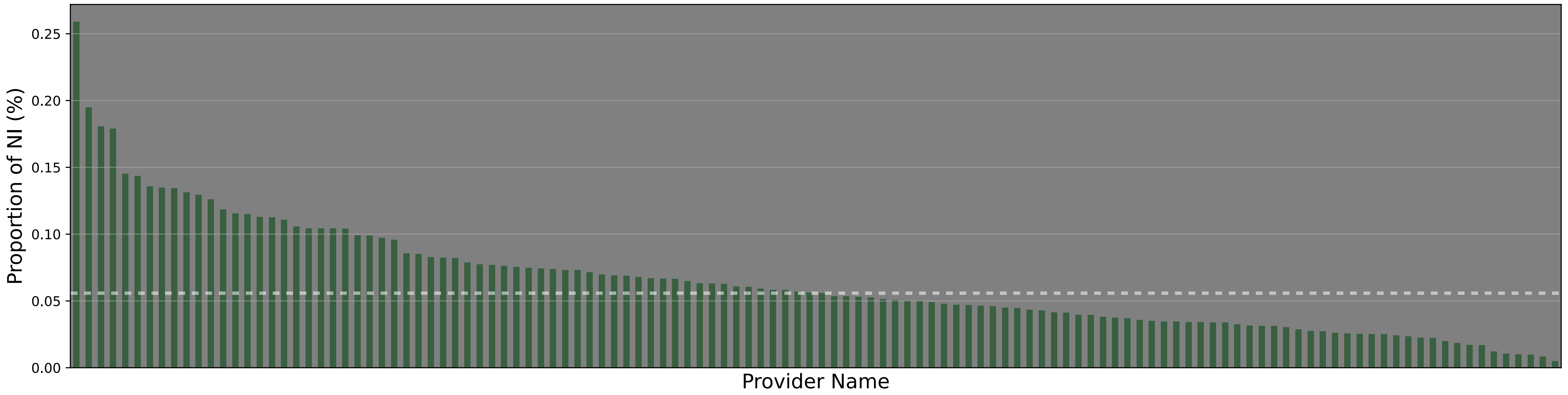
Implementation of Methods 1-4

- March 1st 2020 -March 31st 2021
- Total number of discharges: 374 244
- Number of NI identified: 29 896
- **NI rate** $\sim 8.0\,\%$

NI Rate Across Regions



Proportion of NI across providers



Variation of NI rate across trust could also be a result of **local recording and clinical coding practice**

- We may have missed NI using methods 1-4.
- NI is likely to be an underestimate

Approach:
Use the identified NI to train a model capable of learning the features that are likely to be associated to these infections.

Machine Learning Approach to NI

Data driven approach:

- **Choose a model** capable of learning characteristics of NI
- **Optimise and train** this model using the NI identified
- **Apply this model** to an unseen set of Covid-19 infections

Difficult task for several reasons:

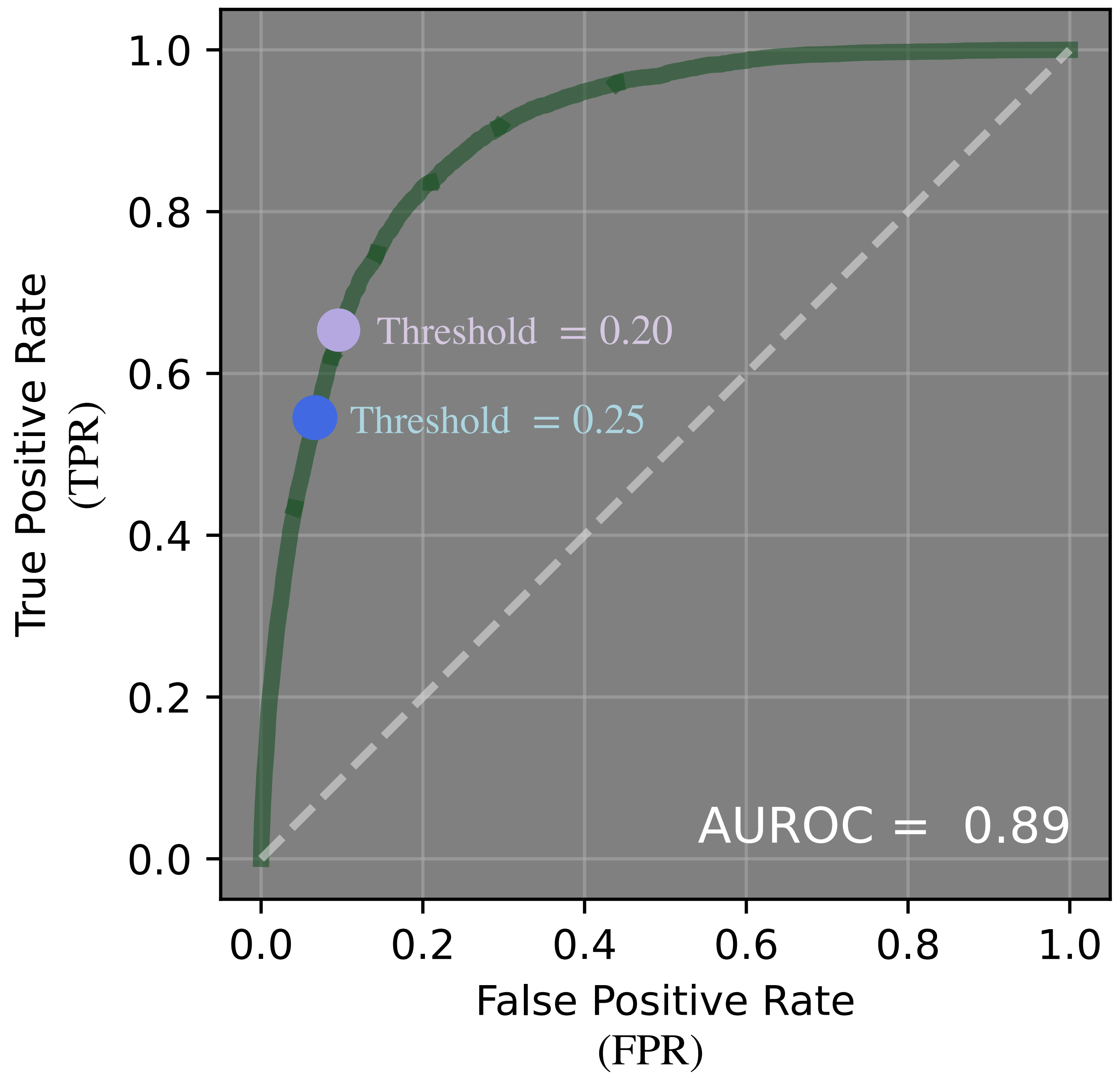
- Imbalanced dataset (of the order of ~ 90:10)
- Unsure labels: we may have FP and FN

Assumptions:

1) We have identified enough NI for the model to pick up on clear patterns

2) The model is robust enough to deal with unsure labels

Modelling Using Random Forests



The trained model offers a trade-off between TPR and FPR

This trade-off is fixed by choosing a **threshold**.

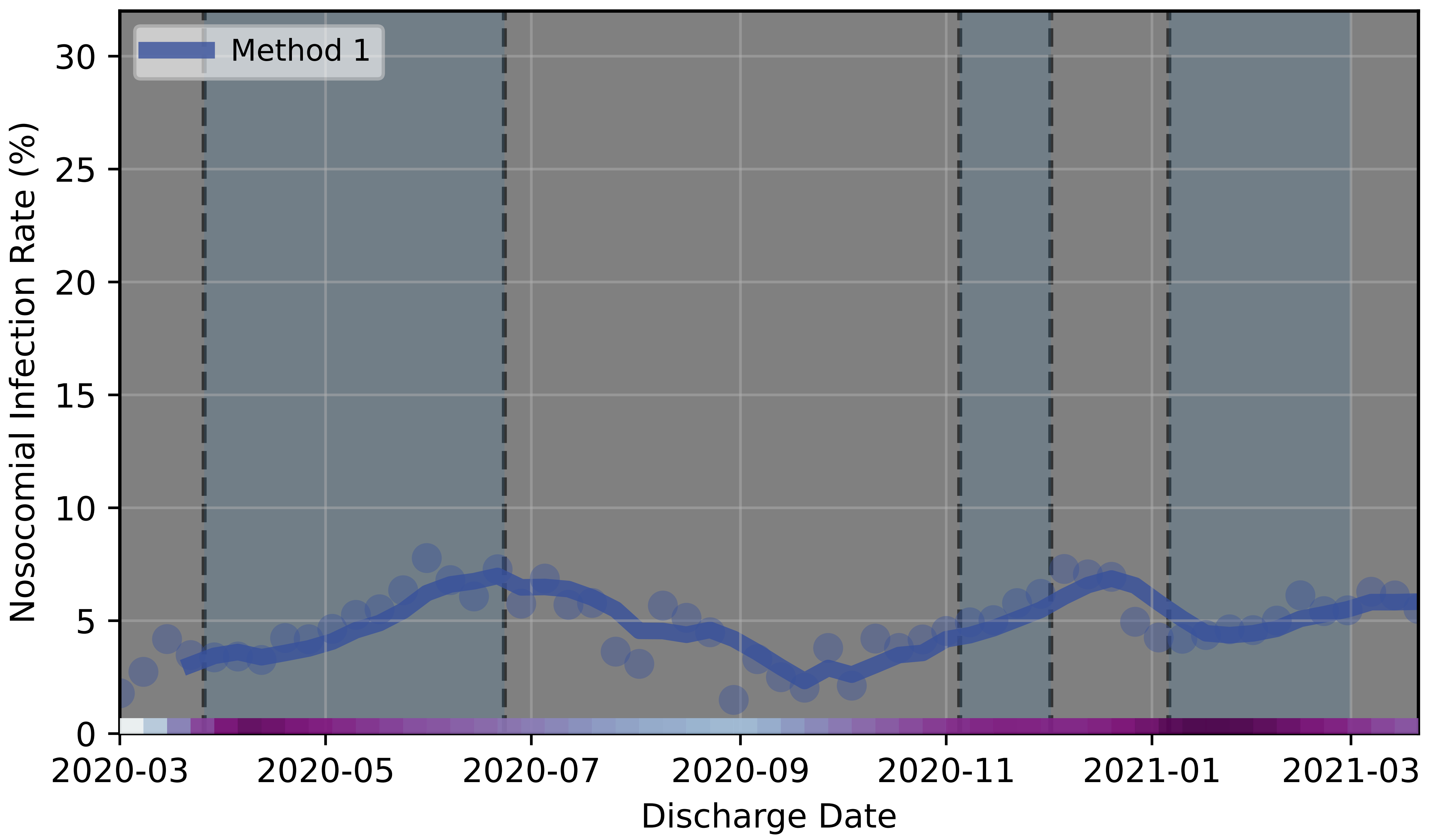
Blue point: threshold of 0.25

How do we constrain this threshold?

1) Relative accuracy of Methods 1-4

2) Design a lower and upper limit for the number of NIs

Predicted NI from the Trained Model:



Method 1: Y95 code

A few remarks:

· The dynamics of the pandemic is illustrated by:

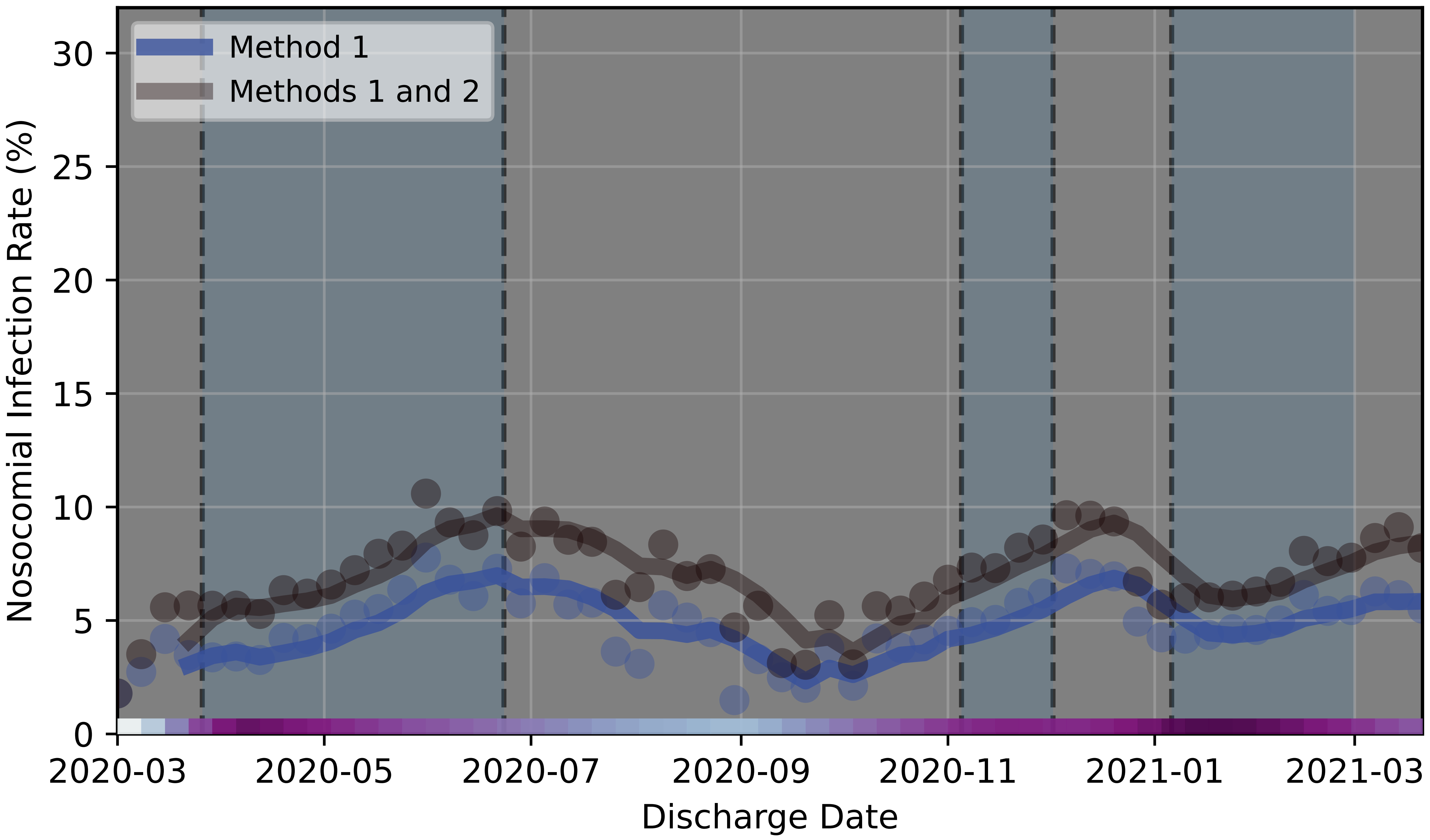
- The blue shaded areas, corresponding to lockdown periods in the UK

- The colour-bar along the x-axis, which shows the number of recorded infections on a log-scale.

· The evolution of the NI rate is expected to lag behind community transmission

· We are plotting against the *discharge date*, which introduces an additional lag—patients with NIs tend show a long length of stay.
We are considering replicating the analysis with the *admission date*.

Predicted NI from the Trained Model:



Method 1: Y95 code

Method 1 + Method 2 (15+ days admission to infection)

A few remarks:

· The dynamics of the pandemic is illustrated by:

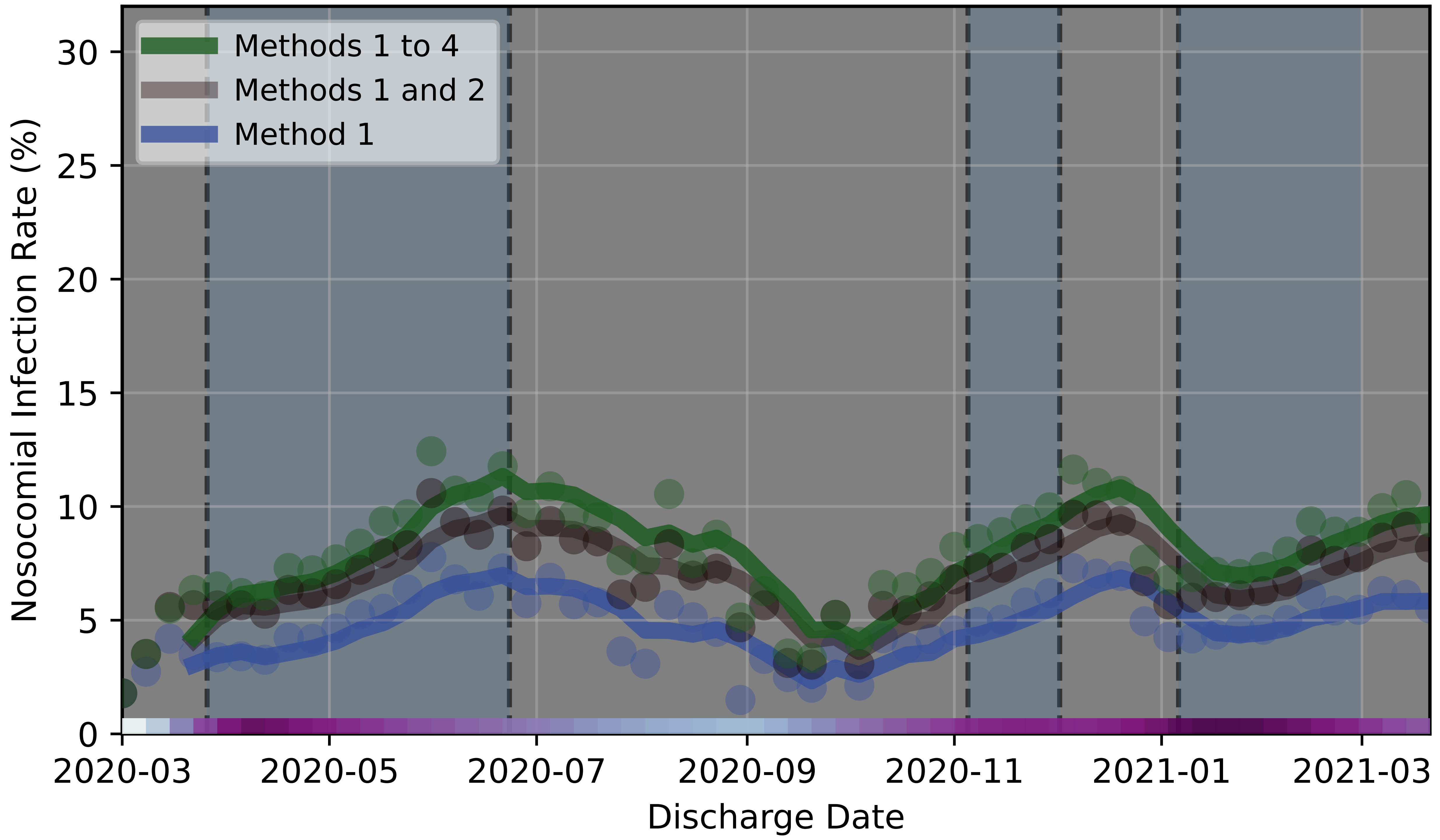
- The blue shaded areas, corresponding to lockdown periods in the UK

- The colour-bar along the x-axis, which shows the number of recorded infections on a log-scale.

· The evolution of the NI rate is expected to lag behind community transmission

· We are plotting against the *discharge date*, which introduces an additional lag—patients with NIs tend show a long length of stay.
We are considering replicating the analysis with the *admission date*.

Predicted NI from the Trained Model:



Method 1: Y95 code

Method 1 + Method 2 (15+ days admission to infection)

Methods 1 to 4

A few remarks:

· The dynamics of the pandemic is illustrated by:

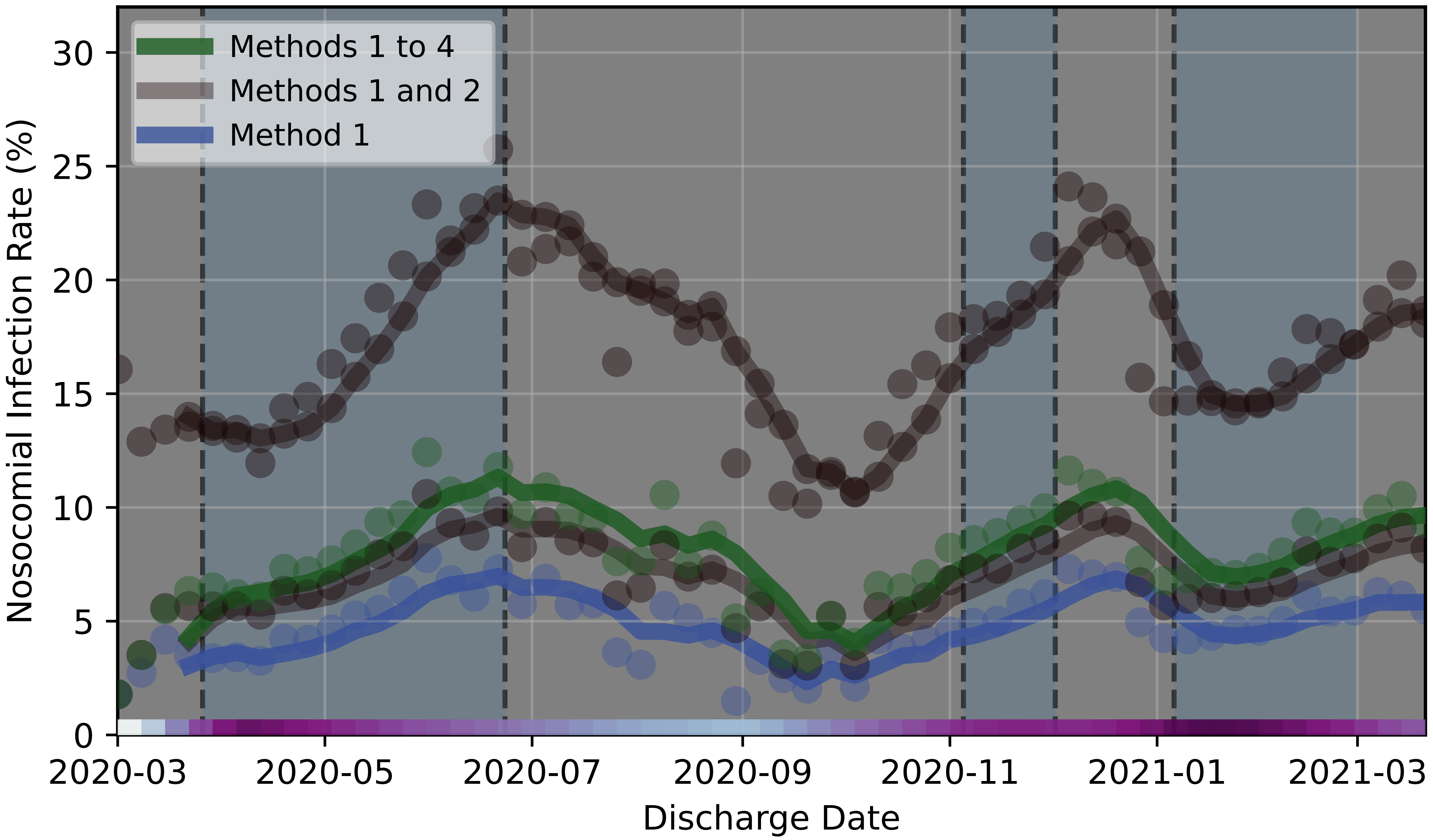
- The blue shaded areas, corresponding to lockdown periods in the UK

- The colour-bar along the x-axis, which shows the number of recorded infections on a log-scale.

· The evolution of the NI rate is expected to lag behind community transmission

· We are plotting against the *discharge date*, which introduces an additional lag—patients with NIs tend show a long length of stay.
We are considering replicating the analysis with the *admission date*.

Predicted NI from the Trained Model:



Method 1: Y95 code

Method 1 + Method 2 (15+ days admission to infection)

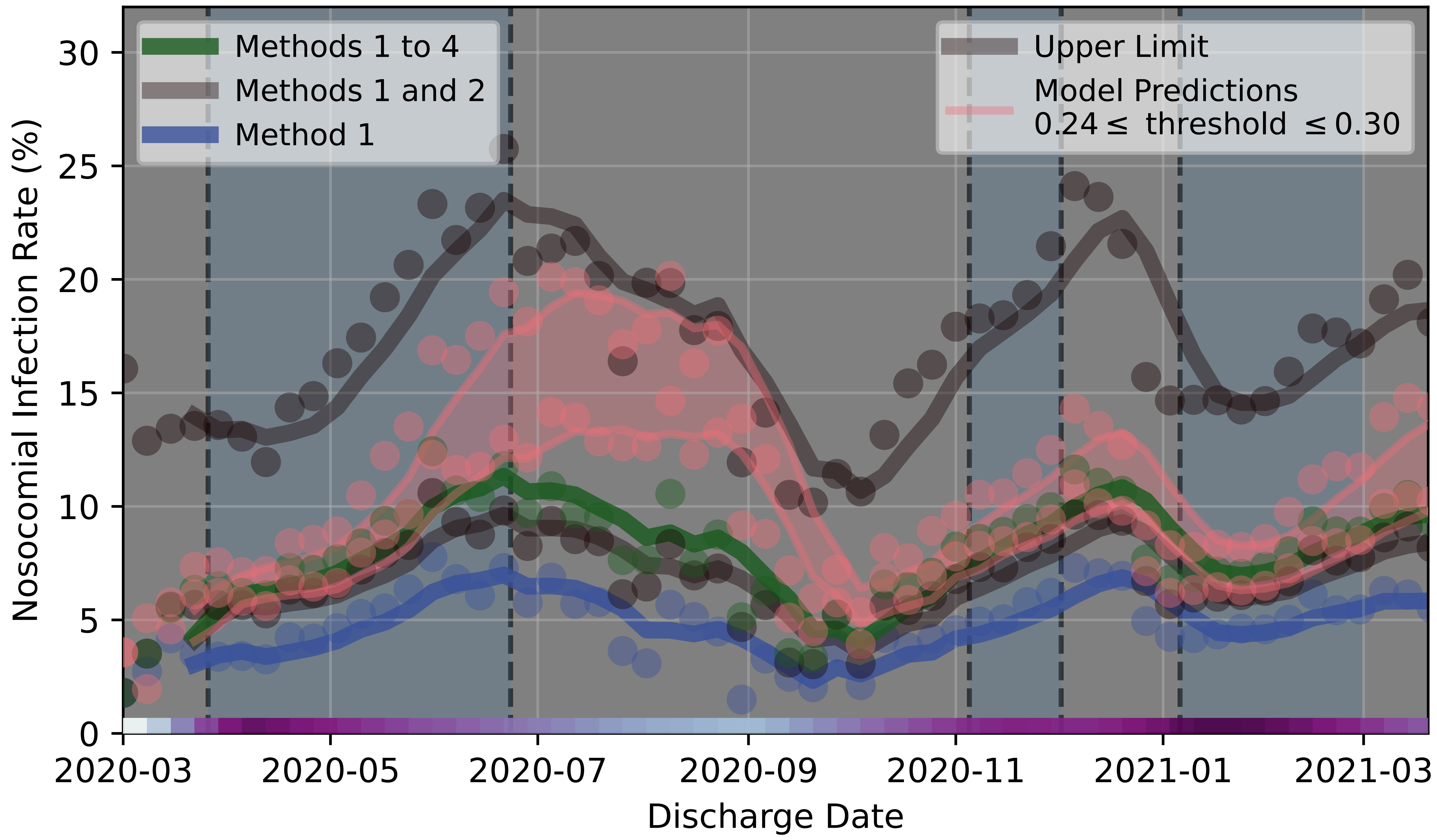
Methods 1 to 4

Upper limit

Upper Limit designed to include *all possible* NIs:

- Method 1: Use of Y95 Code
- Method 4: Use of Z208 prior to emergency admission
- Elective admission with length of stay > 2 days
- Emergency admission with infection after start of the spell + length of stay > 2 days

Predicted NI from the Trained Model:



Method 1: Y95 code

Method 1 + Method 2 (15+ days admission to infection)

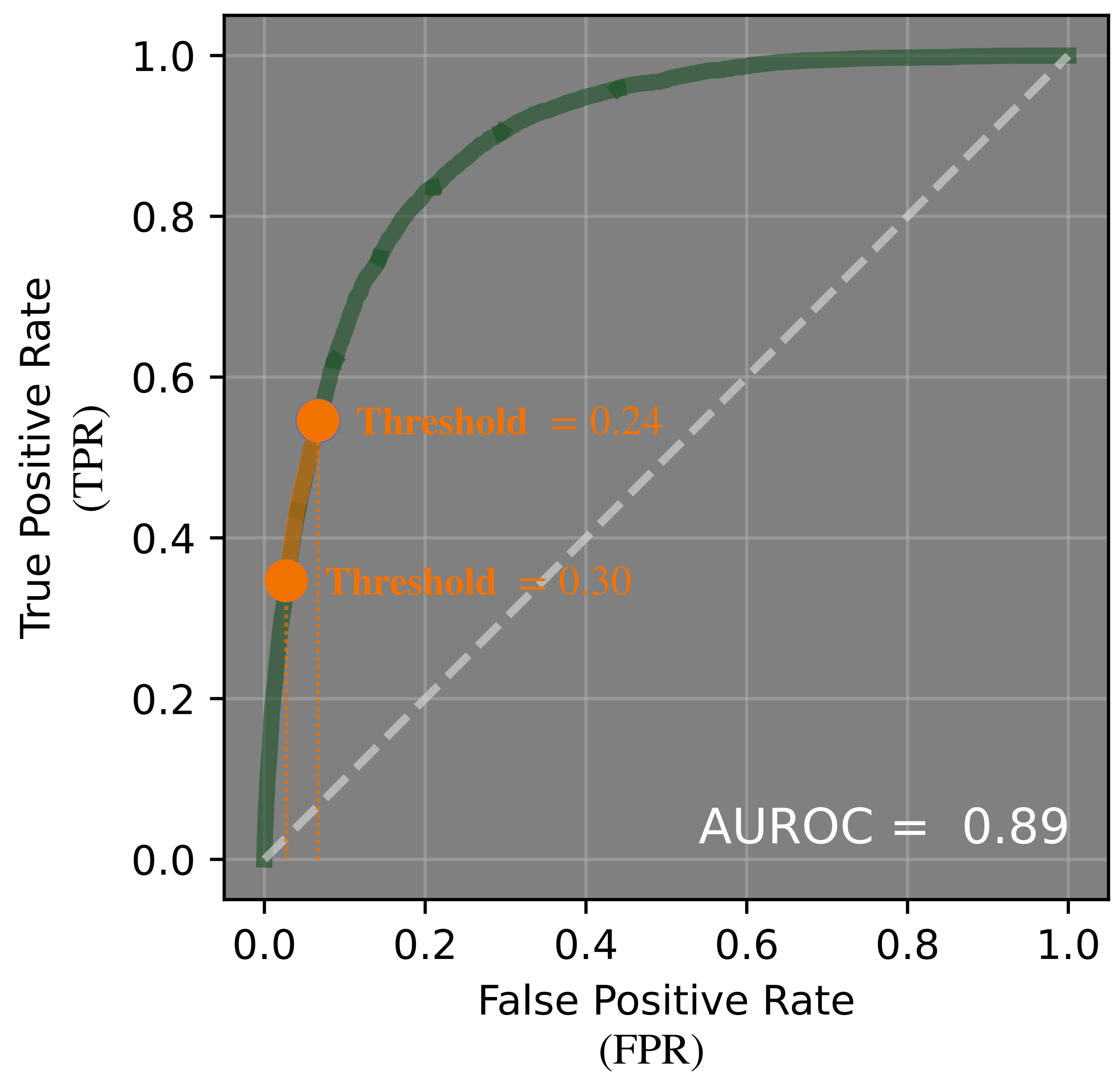
Methods 1 to 4

Upper limit

Predictions from model ($0.24 \leq$ **Threshold** ≤ 0.30)

Upper Limit designed to include *all possible* NIs:

- · Method 1: Use of Y95 Code
- · Method 4: Use of Z208 prior to emergency admission
- · Elective admission with length of stay > 2 days
- · Emergency admission with infection after start of the spell + length of stay > 2 days



Methods 1+2 identify definite NIs: this fixes a **lower limit**.

The two limits constrain the model threshold to $[0.24, 0.30]$

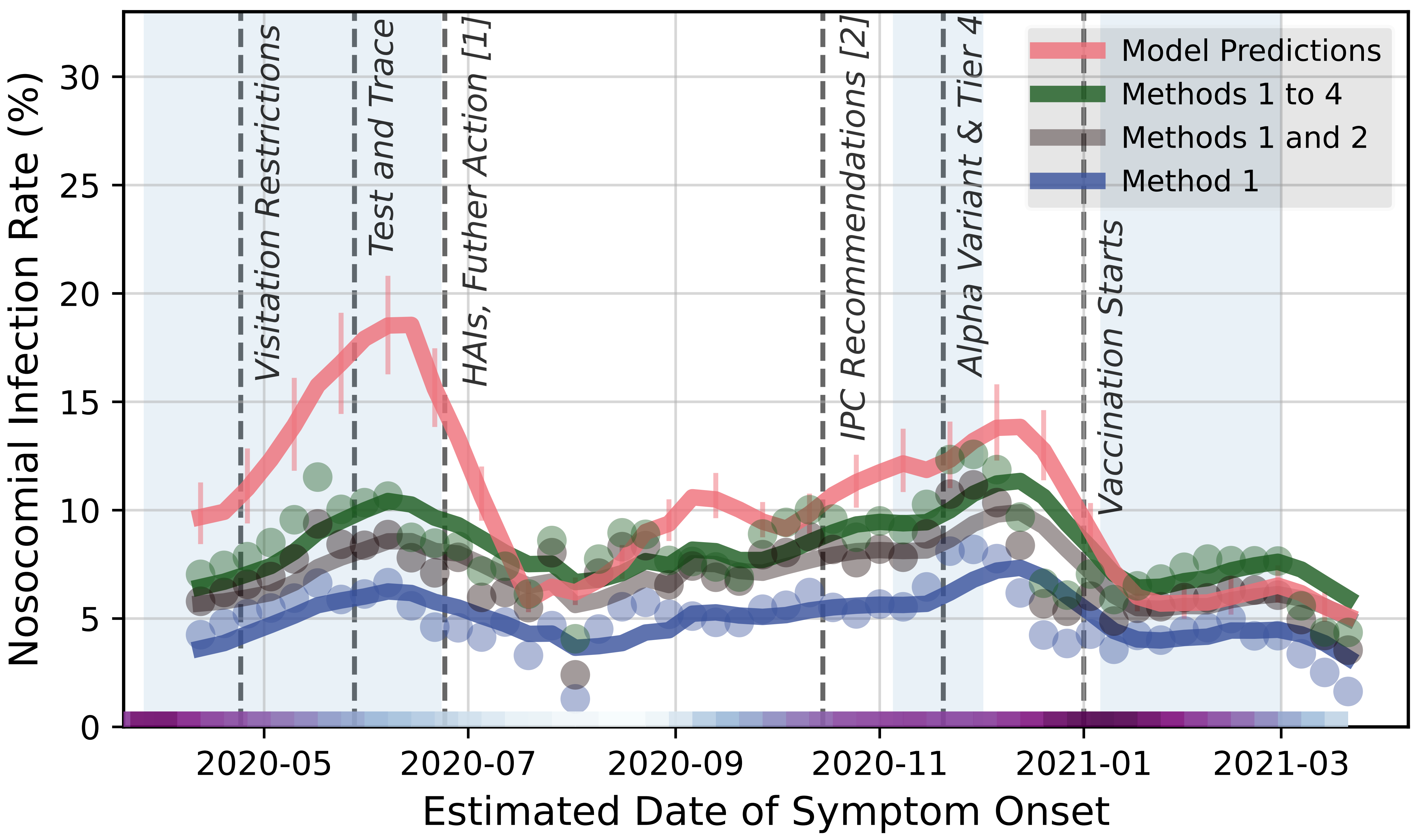
Features Identified as Important by the Model

Out of 130 features:

Features	Importance
Spell_Los	0.032
PropMaxPatientsWave	0.014
HFRS_Band_Severe	0.013
Counts_SameDay	0.013
Charlson_Score	0.011
Total_Hopper_Domain	0.011
age_of_patient	0.010
Mortality	0.009

Also:

ICD-10_**J90X** (*Pleural effusion not elsewhere classified*)	0.004
ICD-10_**N179** (*Acute kidney failure*)	0.004
ICD-10_**J189** (*Pneumonia*)	0.003
ICD-10_**J181** (*Lobar pneumonia*)	0.003



Features engineered to act as proxies for how busy the trust was, relative to its capacity:

Counts_SameDay:
Number of patients admitted on the same day

PropMaxPatientsWave:

$$\frac{\text{Number of patients admitted on the same day}}{\text{Maximum number of patients admitted during the wave}}$$

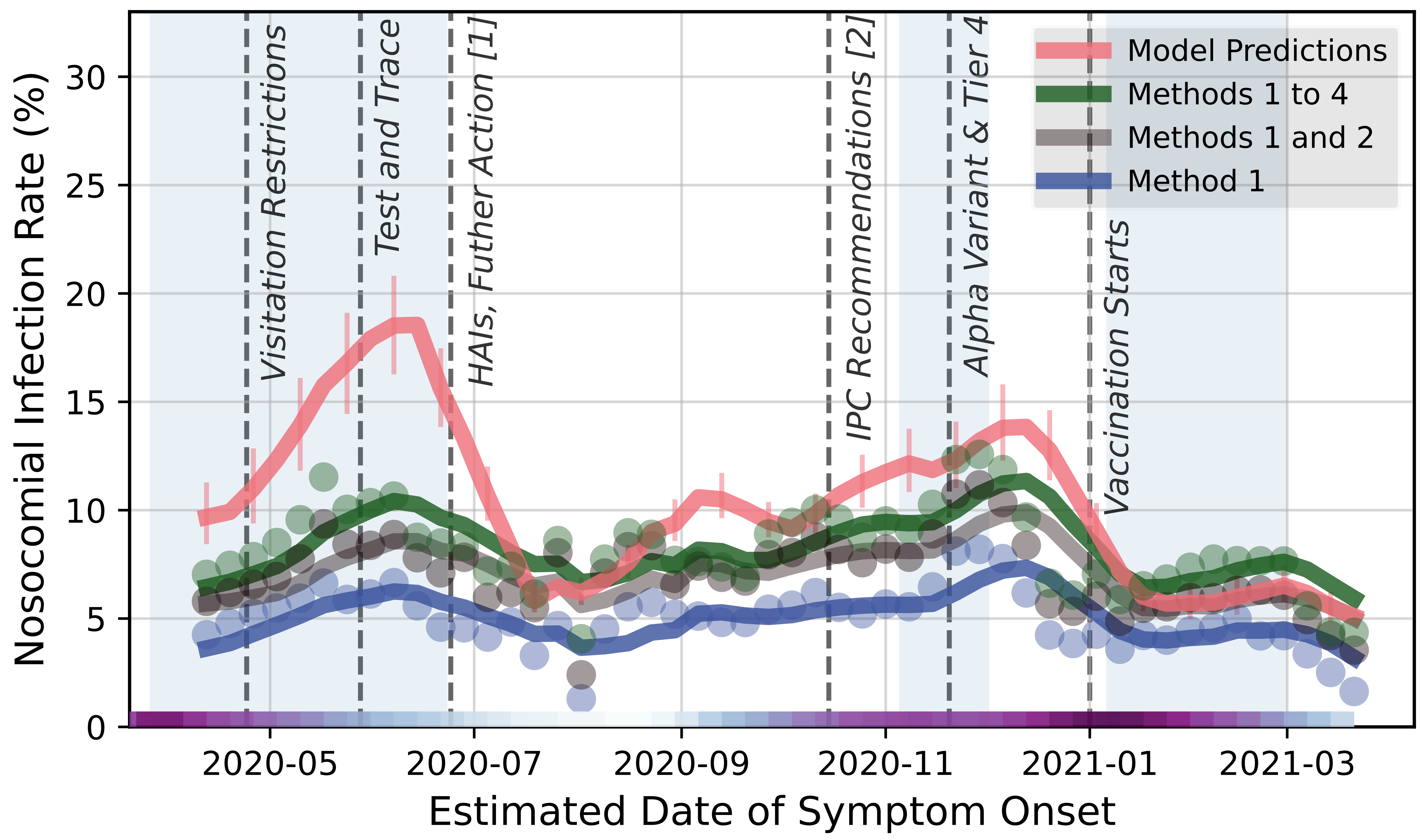
Next Steps:

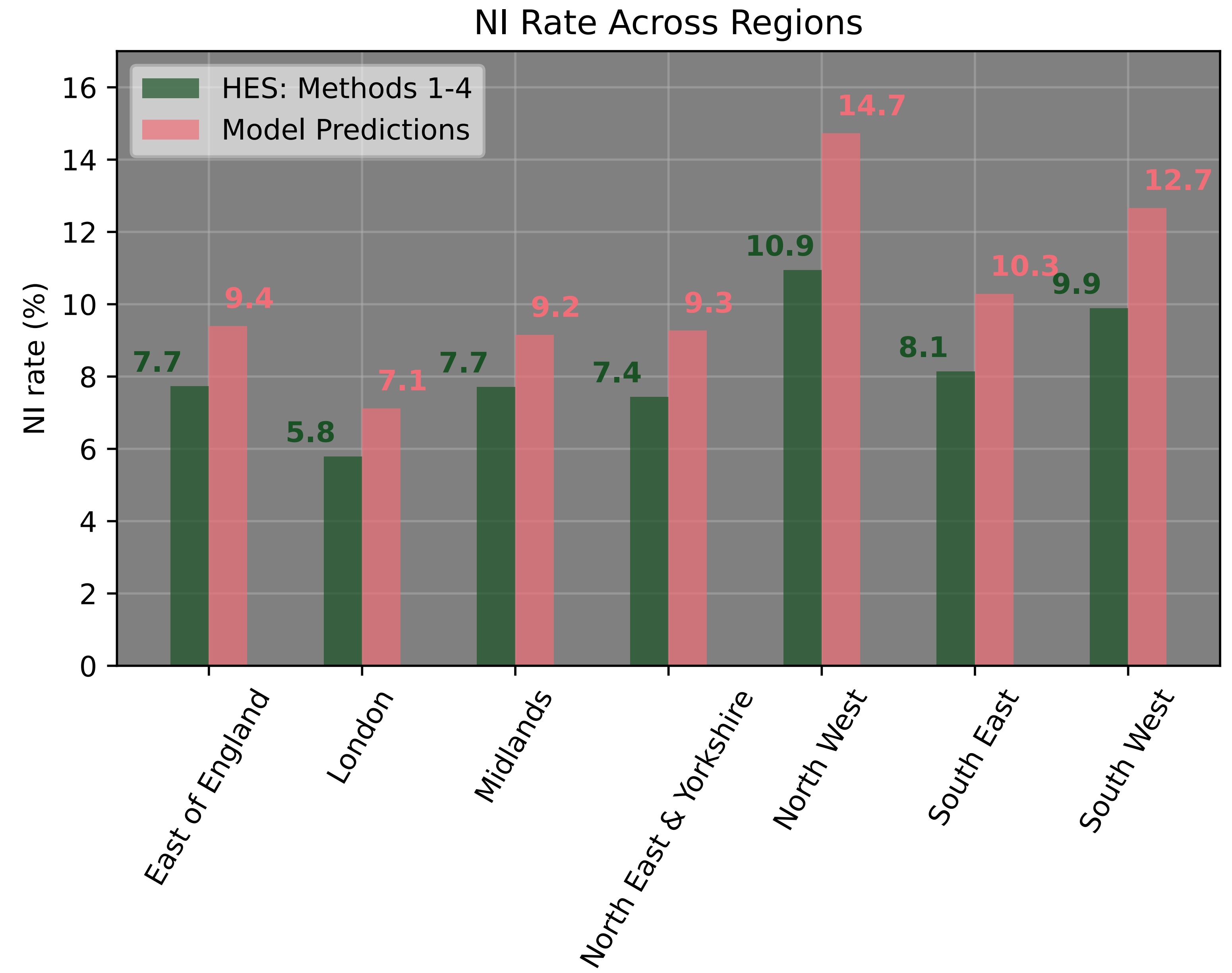
- Refine features related to **severity**:
  - Discussions with Andy and Sue, who will draw a list of ICD-10 codes to potentially include
  - Is there value in analysing the types of patients having acquired NIs?

e.g. find ICD-10 codes that are most prevalent in the identified cohort, and introduce them as features in the model (high prevalence of UTIs, urinary retention)

- Look at the time evolution of the NI rate using the **admission date**, instead of discharge date.

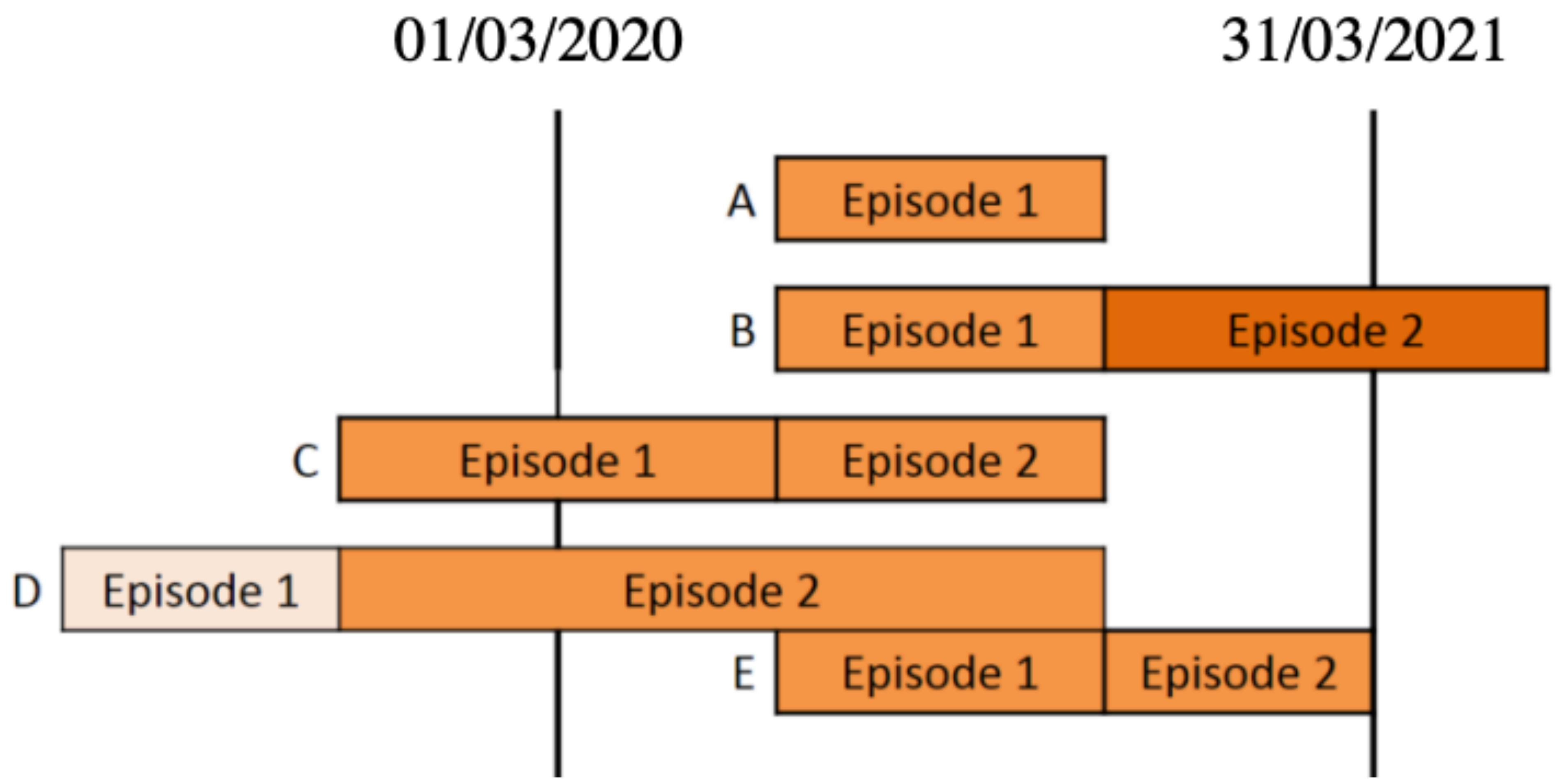
- Split the NI rate by elective and non-elective surgery; is there any evident clustering around **elective admissions**?





Appendix

HES: Spells and Finished Consultant Episodes (FCEs)



Adapted from *Methodology to create provider and CIP spells from HES APC data, NHS England*

· In HES, a hospital admission is referred to as a **spell**: it is an **uninterrupted inpatient stay at one hospital**.

· Spells may include several **Finished Consultant Episodes** (FCEs) if the patient was seen by multiple consultants during the same stay.

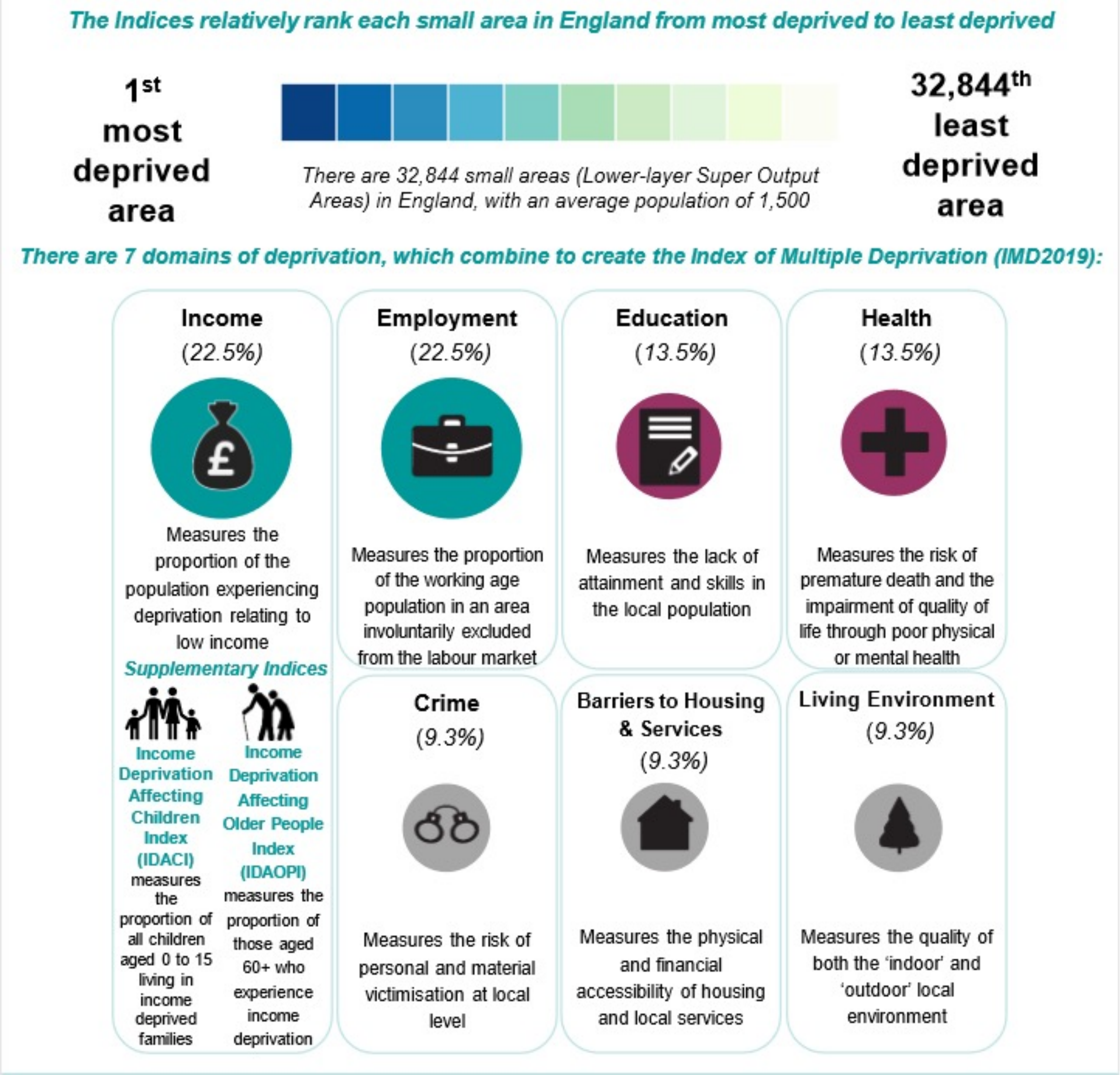
· Here: Spell A includes a single episode
Spell E includes two episodes

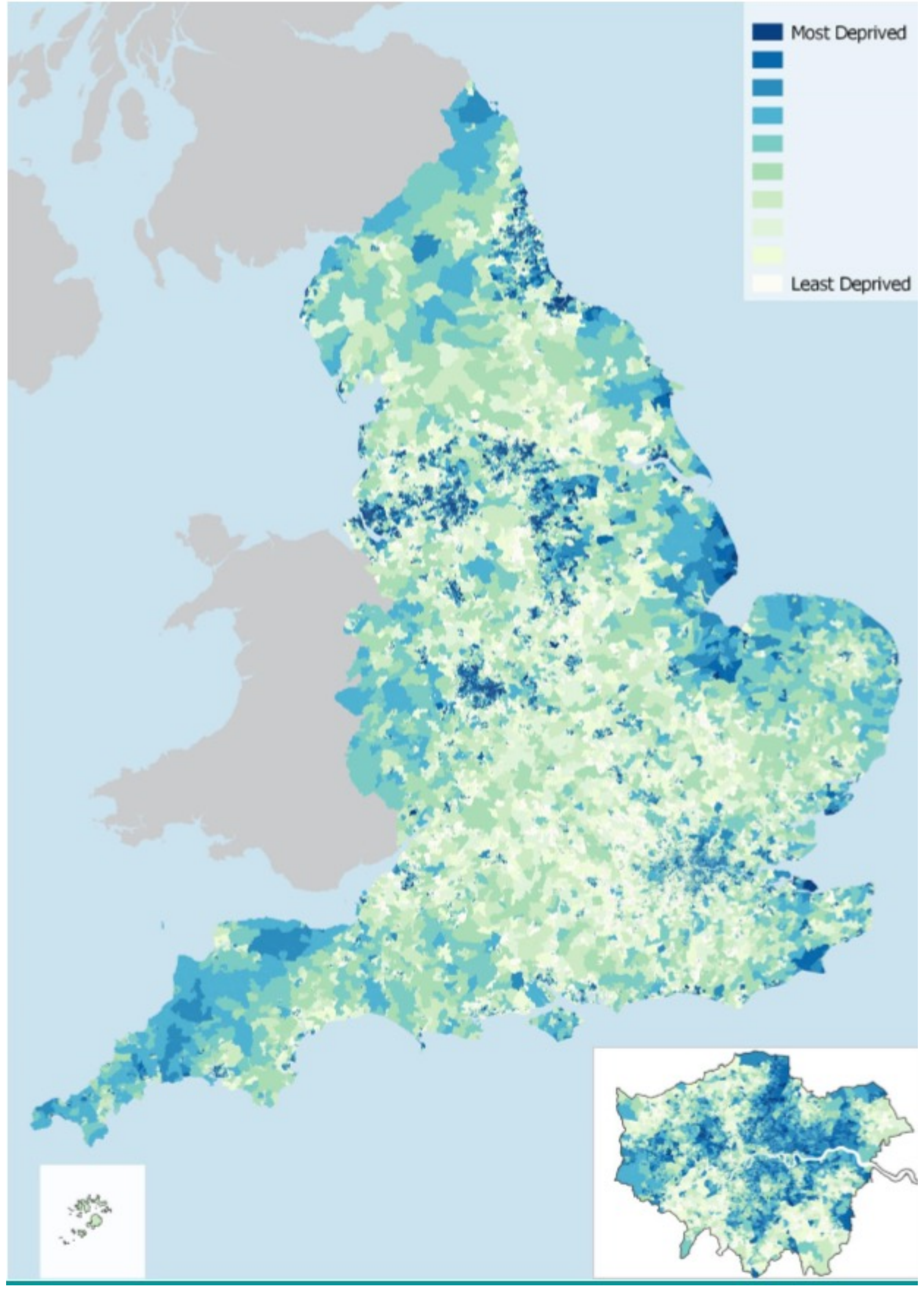
· In slide 2, we introduced **Method 2**: *infections recorded > 15 days after admission*.
This corresponds to the first episode having no mention of U071/U072, and the first occurrence of a Covid-19 diagnosis corresponding to a later episode starting more than 15 days later, within the same spell.

For example: if the U071/U072 codes in spell E first appeared in episode 2, the infection would be flagged as nosocomial if episode 2 started more than 15 days after the start of episode 1.

IMD indices of deprivation

Guy's and St Thomas' NHS Foundation Trust





https://www.gov.uk/government/collections/english-indices-of-deprivation