Predicting NHS patient inflows during an infectious disease outbreak or pandemic

SCC460 Group Project

February 2022

ABSTRACT

Aim: To develop a prediction model that forecasts what the expected number of patient admissions will/might be - per day, N weeks ahead, and per NHS Trust - during an infectious disease pandemic.

Objectives: (a) Understanding, determining, the range of predictors that influence SARS-CoV-2 hospital admissions per NHS Trust, (b) Forecasting N weeks ahead

INTRODUCTION

METHODOLOGIES

Research Strategy

- Ideas: A Model of the Research Process (Oates 2006)

Data Collection

- A diagrammatic illustration of all the data sources, data transformations, feature engineering, data integration, etc., will be most useful.

SARS-CoV-2 & Supplementary Data

Thus far, the focus is the coronavirus.data.gov.uk SARS-CoV-2 data; data extraction via the API. The table below outlines the variables in focus thus far. A few points

- Data variables, i.e., SARS-CoV-2 data measures, are not available at equivalent granularity. For example, only a handful of un-aggregated measures are available at NHS Trust level; the aggregated measures depend on the un-aggregated measures.

- Age demarcations are inconsistent: At the lower tier local authority (LTLA) level, the age groups used for *newCasesBySpecimenDateAgeDemographics*, differ from the age groups used for *vaccinationsAgeDemographics*. Although the API does not include NHS Trust level age group decompositions,
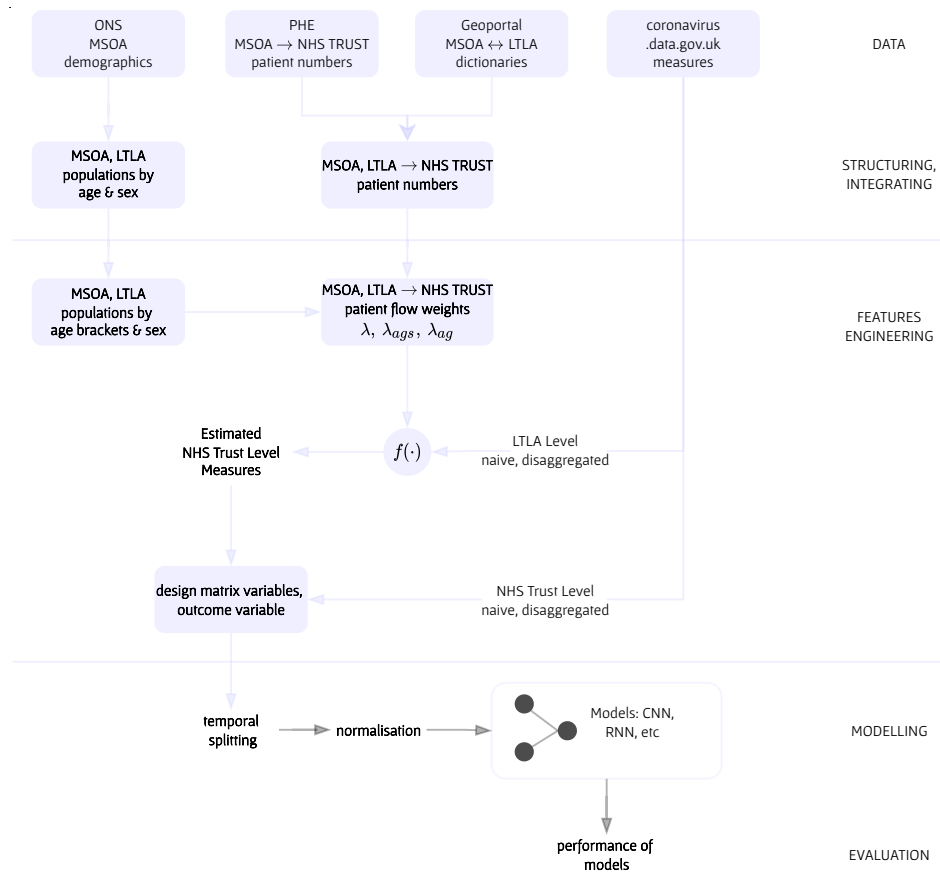
Figure 1: Steps

such decompositions are available via COVID-19 Hospital Activity, which – amazingly – uses a third different set of age group demarcations.

- A few API variables have misleading names, these are renamed; the NHS Trust level variables have been renamed to reflect their original COVID-19 Hospital Activity labels.

Addressing Disparate Data

The prospective modelling measures exists at either NHS Trust or Lower Tier Local Authority (LTLA) level, and because our interest is in NHS Trust level admissions, we need the NHS Trust level forms of the LTLA measures.

Approach: In brief, a section of NHS England records the number of patients that each Middle Super Output Area (MSOA) contributes to each NHS Trust each year; each MSOA contributes patients to one or more NHS Trusts. However, each MSOA region is a member of one LTLA region, hence the number of patients that each LTLA contributes to a NHS Trust, per year, can be calculated via appropriate aggregates. Via these LTLA level aggregates the questions - *per LTLA that contributes to a NHS Trust*

- what fraction of the LTLA's patients did the trust receive?
- what fraction of the LTLA's patients, within age group 5 - 9, did the trust receive?

are calculable, and the resulting fractions guide the apportioning of LTLA level COVID measures to NHS Trusts.

Assumptions: The preceding approach makes a few assumptions, e.g.,

- The 2019 patient flow fractions from a LTLA to a NHS Trust, for all admissions types, are reasonable estimates of flow fractions for any admission type. Not ideal. There are several arguments against this assumption. For example, and in relation to SARS-CoV-2, are all trusts equally equipped to deal with a respiratory pandemic?

- The population characteristics of each LTLA change minimally within a short period. Not ideal either.

Flow Weights

The total # of patients recieved by a trust The population of a LTLA: α The population of a LTLA by age group sex: α_{ags} The population of a LTLA by age group: α_{ag} The patient population of a LTLA: ρ The # of LTLA Patients $\rightarrow$ NHS Trust: β The fraction of a LTLA's patients received by a trust: $\lambda = \beta / \rho$ [weight] The estimated age group sex fraction received by a trust: $\lambda_{ags} = \lambda \alpha_{ags}/\alpha$ [weight] The estimated age group fraction received by a trust: $\lambda_{ag} = \lambda \alpha_{ag}/\alpha$ [weight]

Modelling

- Pre-modelling procedures. [This point refers to the pre-ML/pre-forecasting pre-processing. It is separate from the initial pre-processing & features engineering that leads to the raw design matrix and dependent variable.]
- Data mining techniques?
- Statistical techniques?

RESULTS

Project guide:

- Explain your results and what your analysis revealed
- What implications would your analysis' results have? How do your findings relate to the original question?
- Were there potential biases in your work?
- Validity

CONCLUSIONS

Project Guide:

- Reflection on the approach taken. (Appropriate?)
- How would you have improved the approach in future? (Alternative methodologies, models, etc)

REFERENCES

Oates, Briony J. 2006. *Researching Information Systems and Computing.* SAGE.