Predicting NHS patient inflows during an infectious disease outbreak or pandemic

SCC460 Group Project

February 2022

INTRODUCTION

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 (Knight et al. 2020)

- 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.

METHODS

Research Strategy

In progress ... includes research strategy (Oates 2006)

Data Engineering

- The data is from ... (the data sources)
- Hence, LTLA level aggregates .... (structuring, integrating)
- Hence, create features for modelling ... (features engineering)

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 notes

- 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]

Figure ...: A high level illustration of the project's processing, analysis, modelling, and evaluation steps. The acronyms/abbreviations are API: application programming interface, MSOA: middle layer super output area (a United Kingdom census geography), LTLA: lower tier local authority (a United Kingdom admnitrative geography), ONS: office for national statistics, NHS: national health service, PHE: Public Health England. The ONS, NHS, and PHE, are United Kingdom entities.

The data sources are $\rightarrow$ England's SARS-CoV-2 infections related measures: coronavirus.data.gov.uk API, demographics data: ONS, MSOA $\leftrightarrow$ LTLA geographic codes mappings: Open Geography Portal (geoportal), annual intake of patients from one or

ONS
MSOA
demographics

Open Geoportal
MSOA ↔ LTLA
dictionaries

PHE
MSOA → NHS TRUST
patient flow numbers

coronavirus
.data.gov.uk
measures

DATA

LTLA
populations by
age & sex

LTLA → NHS TRUST
patient flow numbers

STRUCTURING,
INTEGRATING

LTLA
populations by
age brackets & sex

LTLA → NHS TRUST
patient flow weights

transformer $f(\cdot)$

LTLA Level
coronavirus
measures

FEATURES
ENGINEERING

Estimated NHS Trust Level
coronavirus
measures

Original NHS Trust Level
coronavirus
measures

Design & Outcome
Variables

temporal
splitting

reconstructing,
differencing,
normalising

training
data

designing & training
of models

MODELLING

developed
models $f(\cdot)$

training, validating,
testing data

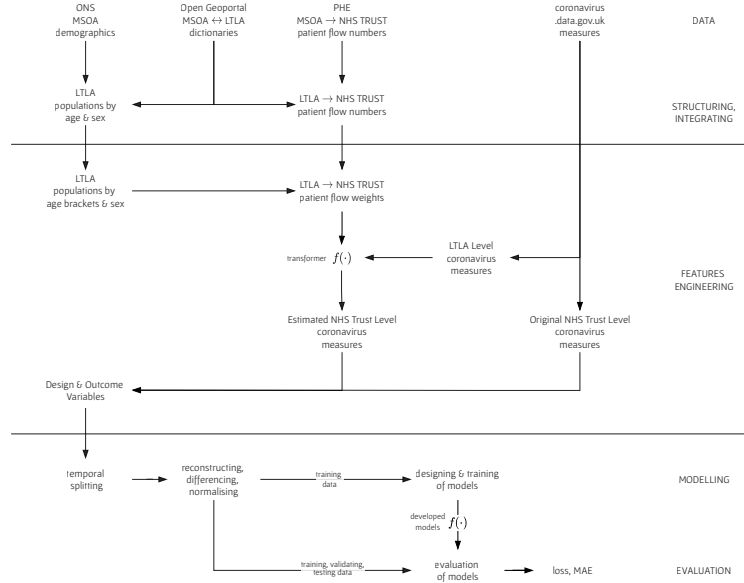evaluation
of models

loss, MAE

EVALUATION

Figure 1: Steps

more MSOA areas to an NHS Trust: **(a)** NHS Trust Catchments Estimation, **(b)** visualisations of

Please refer to the methodologies section for a description of (a) the patient flow weights, and (b) the estimation of NHS trust level measures via flow weights and LTLA level measures.

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?

H.A., J.R., C.O. were discussing:

- Data preparation ... temporal splitting, [rec, diff, ...]
- model design & training ...
    - CNN, LSTM, GRU ... why? briefly discuss type of problem
    - Window ... why? ... forecasting 15 days into the future based on unknown "effective days of history" ... windows allow us to investigate via varying history window sizes

Evaluation

A brief description of the evaluation metrics. [Results discusses the evaluations in detail]

RESULTS

Model Evaluation

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?

Biases & Validity

- Were there potential biases in your work?

- Validity (remember to discuss what would have been done differently to address identified limitations)

CONCLUSIONS

Project Guide:

- Reflection on the approach taken. (Appropriate?)

- How would you have improved the approach in future? (Alternative methodologies, models, etc)

REFERENCES

Knight, Stephen R, Antonia Ho, Riinu Pius, Iain Buchan, Gail Carson, Thomas M Drake, Jake Dunning, et al. 2020. "Risk Stratification of Patients Admitted to Hospital with Covid-19 Using the ISARIC WHO Clinical Characterisation Protocol: Development and Validation of the 4c Mortality Score." Edited by J Kenneth Baillie, Malcolm G Semple, Peter JM Openshaw, Gail Carson, Beatrice Alex, Benjamin Bach, Wendy S Barclay, et al. *BMJ* 370. https://doi.org/10.1136/bmj.m3339.

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