**Long Covid: The Risks and Factors using OpenSAFELY**

Josh Bibby

Leon Danon

Department of Computer Science

University of Bristol

Bristol, BS8 1UB, UK

Is20347@bristol.ac.uk

1. **Fairy-tale Abstract**

One page of ‘fairy-tale’ abstract.

***Abstract —* We use the OpenSafely database to explore long covid. Looking at the primary care health records of 40% of the English population to analysis factors around long covid. We explore how long covid has been defined so far and how we will define it here and therefore how we will decide who has or has had long covid. We then explore factors such as comorbidities, medications, COVID-19 symptoms of patients suffering from long covid with a particular look at mental health prescription as well as demographic analysis. Who is more likely to suffer from long covid? Do certain underlying or previous health conditions affect your likelihood of suffering from long covid? Our primary findings are WHAT and WHAT and WHAT.**

***Keywords— Long Covid; OpenSAFELY.***

1. **What I hope to say about my project when it is complete**

The coronavirus pandemic has been very impactful but so has the response to it. We have been able to react strongly to something that has affected so many, so harshly. As we continue to gain control over this pandemic the longer lasting effects have become more prevalent. One of these affects being a condition commonly known as Long Covid, that can hinder patients many weeks after they have recovered from their initial COVID-19 infection.

Long covid is still poorly understand, its causes, symptoms and duration have not yet been strictly defined. There have been several studies conducted demonstrating how early in our understanding about long covid we are at this stage. Some studies have looked at patients’ symptoms 4 weeks after their initial infection, some 12 weeks, and some 60 days. Some have conducted follow-ups on patients who were admitted to hospital with COVID-19 infection, and some have focused on surveying people within patient-lead social media groups based around long covid. The diversity of these studies exemplifies how we are still in the early stages of our understanding of this condition.

These studies have been of varying sizes but mainly revolved around patient surveys and check-ups/follow-ups as well as patient voice. These studies have understandably limited scope without large investment. In this paper we utilise the OpenSAFELY platform to analyse primary care electronic health records (EHR) in the NHS with the potential to utilise pseudonymised health record of over 58 million patients. With this greatly increased scope, we can dig into details not available in previous smaller studies.

We started our analysis reproducing common ideas from previous studies but with the key difference of performing it on a much larger dataset of HER records within OpenSAFELY. We first looked at patients who received an official long covid diagnosis (described in NICE guidelines as, ‘Ongoing symptomatic COVID-19’ (4-12 weeks) or ‘post-COVID-19 syndrome’ (12+ weeks), REFS), initially exploring common characteristic factors such as age, sex and ethnicity and then moving onto investigating comorbidities, medications, and mental health in long covid patients. Our findings are WHAT and WHAT and WHAT .

Once this analysis was completed, we went onto compare acute COVID-19 infection (short covid) with long covid. Exploring questions such as, does hospital admission/severity/number of symptoms of the short covid illness affect how long covid manifests itself? What proportion of people with no positive COVID-19 test or diagnosis have a long covid diagnosis? Our findings here are THIS.

Once this was completed, we noted that diagnosis of short covid is not a prerequisite for getting long covid and from this we explored all people describing symptoms of long covid but without an official diagnosis, being careful to make sure we are aware that these symptoms may come from other conditions such as ME/CFS. We did this because long covid is such a new condition that many GPs may not have thought to give this diagnosis in their initial consultations. Our findings here are THIS.

Once this was completed, we explored COVID-19 vaccination and long covid. Answering questions such as, does the vaccine help protect against long covid and if so to what extent? Our findings here are THIS.

Finally, because we are utilising such a large dataset then we were able to perform some cluster analysis on long covid patients. Long covid patients report symptoms that are diverse and differ from patient to patient. The questions we asked were, can we significantly cluster long covid patients based on their symptoms and if so, what we say about this? Are there different types/variations of long covid and if so, do patients of each cluster have representative symptoms for that cluster? Our findings here are THIS.

1. **Introduction/Overview**

At least two pages.

The COVID-19 virus has had a devastating effect on the world population. As of June 2021, the world has seen over 175 million reported cases and almost 4 million deaths. The recent propagation of vaccines for acute COVID-19 has shone a light at the end of the tunnel during this pandemic.

However, the effects of this pandemic will linger long after this virus is brought under control; physically and mentally for some, as well as socially and economically for larger society.

1. **Literature Review**

At least three pages.

1. **Bibliography/References**
2. **Appendix – Time-plan**

One page.

1. **Appendix – Risk Assessment**

One page.

**Project not being accepted -** Our first risk would be our project not being accepted by the OpenSAFELY team and therefore us not having access to the EHR database. The OpenSAFELY team determines whether each project is suitable and whether it poses any risks to patient confidentiality etc. This is unlikely to happen at this stage as the project has already been discussed with a member of the OpenSAFELY team and this project is like others already completed with OpenSAFELY. However, were problems to occur we would first clarify/adjust anything that the OpenSAFELY team is concerned about to help get the project accept. But if that ended up not working out, we would have to complete the project using different dataset such as Openpresribing.net.

**Disclosive information being released -** The next risk would be potentially releasing highly disclosive information about patients within the database. This is an ever-present risk when analysing primary care EHR records and something the team at OpenSAFELY is well aware of. We believe this is a substantial risk and that is why there are mechanisms within the OpenSAFELY process that aim to limit the chance of this risk occurring. Additionally, the OpenSAFELY team does well to set out guidance and best practices for what they believe will limit the chances of patients being identified. We believe following these best practices will ensure this risk is kept to a minimum. If this risk were to occur then we would again follow the best practices set out by the OpenSAFELY team by doing things such as contacting their team as soon as possible, making any public GitHub repos private, tracking down any copies of these repos and making sure they are private etc.

**Data being released/leaked –**

**Insufficient data to produce results -**

**Analysis/code being stolen -**