

# Tracking COVID19-like Symptoms using Online Self-Report at Scale

Dr Eoin Travers, UCL

[Eointravers.com](http://Eointravers.com) |   
[eoin.travers@gmail.com](mailto:eoin.travers@gmail.com)

Like most scientists, I've been sitting at home watching the news, and asking myself if there is anything I could do to help. I'm a cognitive psychologist, so I can't contribute to the biomedical efforts, and I'll leave the analysis of the existing epidemiologists to the experts. Here's my idea.

## The Problem with Diagnostic Tests

The government can't afford coronavirus tests for everyone, and so only particularly at-risk individuals will be tested. Similarly, case histories are only recorded for people who are admitted to hospital, so while we know how many of the people hospitalised had a fever in the last week, we don't know how many people had a fever in the last week but didn't go to hospital. This would be good to know.

The solution is obvious. **We need to ask everyone how they're feeling.** This means not just asking people admitted to hospital, but also people sick at home, people self-isolating just in case, and even people who feel fine aren't worried about the whole thing. By recording people's symptoms, along with some demographic data, we can get a sort of State of the Nation.

The only reasonable way to do this is online, and the only simplest way of doing this, when we take into account ease of access, aesthetics, and scalability, is using Google Forms. There might be a case for using custom html, but let's come back to that.

So, my big coronavirus citizen science idea is this: **let's set up a Google Form asking about coronavirus symptoms and demographics information, and circulate it as widely as possible** (within the UK).

Amazingly, it seems like no one is currently doing this, although [YouGov](https://www.yougov.com/) are asking the important questions, like *"If you were due to meet somebody and they offered you a handshake upon arrival, would you refuse?"*.

# What Needs to Happen

## Epidemiologist Needed

First, this idea cannot and should not be implemented without an actual epidemiologist on board. As a psychologist, I know very well how to get this data, but I'm not qualified to know a) **whether this is a good idea** at all, and b) **what we should be asking** people.

**If you're an epidemiologist and would like to be involved or offer some advice, including advising that this is a bad idea and should be scrapped, please get in touch.**

## Anyone Else Needed

Beyond that, there are plenty of things that I can do, but you might know how to do better:

- Survey design.
- Front- and back-end web development, if Google Docs isn't used.
- Participant recruitment (basically, digital marketing).
- Ethics and Data protection compliance (see below).
- Data engineering, if this becomes big.
- Statistics, particularly epidemiological statistics and polling (e.g. MRP methods).
- Project management, if lots of people do get involved.
- Something else I haven't thought of.

**If you're an expert on any of these things and want to help, get in touch.**

# Other Concerns

## Ethics & Data Protection

First, ethical approval at UCL (where I work), or any other university, takes far too long for this to be possible. This survey would have to be run either under some emergency ethics protocol, or privately, with no connection to work whatsoever. I don't see this as an issue. Normal people can run whatever surveys they like in their own time. Can't people with research training do the same? The data could never be published in an academic journal anyway, so that point is irrelevant.

Second, this whole thing is a data protection nightmare. If we didn't care about data protection, we would collect full contact details so that at risk individuals could be contacted by the NHS. If we were to go full GDPR, we would ask about symptoms and whether the respondent was in the UK, but nothing else. The middle ground is to have data that is useful, but could potentially be de-anonymised, for instance if only one person in a particular postcode has visited Italy in the last two weeks. It's also worth considering that with Google Forms, collecting email addresses allows people to go back and edit their responses, which is obviously useful if symptoms change.

I think the best approach is to clearly state at the top of the form that

- Personally identifiable data (e.g. email, full postcode) will only be accessible to the people running the survey and to, for example, the NHS, Public Health England (or whatever the devolved versions are), and the WHO.
- Anonymised data will be made public. This includes any symptoms, demographics, and area codes (e.g. EC1).
- By completing the form, you consent to these terms.

This meets several of the GDPR [legal grounds for processing](#): legitimate, public, and vital interest, and consent.

## What to Ask

So far, I've thought of the following. Personal data that maybe can't be publicly released is in **bold**.

- Age
- Sex
- **Postcode** (if full postcode is provided, can be reduced to area code for public release)
- **Email** (so individuals can update as symptoms develop)
- Travel history (at a provincial level)
- A socioeconomic status indicator
- Underlying health complications.
- Have you tested positive for coronavirus?

- Symptoms

My best idea at present is to do this in closed form (e.g. "*Fever/Cough/Other (please specify)*") and open-ended (what symptoms developed and when, although this would require coding).

An alternative would be to have some kind of nice interface for indicating symptom time course, but I don't know how feasible this is with Google Forms.

- Do you personally think you have coronavirus?
- How many people do you know who have been...
- ???

Clearly we shouldn't ask too much, as every additional item discourages people. It might be possible to structure the form so that the important stuff is asked first, and the more speculative questions are included on a later screen once the key stuff has been saved.

## What to do with the Data

I'm not an epidemiologist, so I'm not the right person to analyse public health data. As a psychologist, though, I do know how to go about getting it. However, here are a few points

- As noted above, anonymised data should be shared publicly, while full data should be shared with the relevant public bodies (NHS, etc.). Again, this needs to be made clear to respondents at the outset.
- The sampling will certainly be biased. [Multiple Regression with Post-stratification \(MRP\)](#) provides a powerful tool for addressing this.
- In reality, since we don't have a good estimate of covid19 prevalence, and since people stuck at home are more likely to fill on online surveys than those at work, we'll never eliminate this source of bias.