Transparency and Openness: Has Covid-19 Changed Data Practices?

Diana Kornbrot

This note is an idiosyncratic review of goals and relevant information sources, followed by comments as to change. Much more may be gleaned from medstats and radstats lists.

Goals for tackling Covid-19

Decisions with the following main goals, need to be informed by data. They must be assessed against base lines, of populations size and age – at very least.

- Reduction in deaths Deaths/million
  - Via lowering transmission R0 <1
  - Via ICU, ventilator, professional resources Capacity re-serve
- Economic activity GDP, GDP change
- Individual wellbeing Universal credit etc.

Upcoming Decisions Lockdown

Decisions are an amalgam of demography and activity
Essential services are unlocked at all times.
Food is always available, but conditions on supermarkets apply.

Activities/Locations

- Transmission
  - Social distance (home, 2 metre, none), masks, surfaces
  - Location group size, transport, industry
- Industry
  - Education: schools, universities
  - Non-essential health, e.g. elective surgery
  - Home-working not possible: retail, manufacturing, construction, tourism

Demography

- Vulnerability Effects resources (vulnerable more ICU), transmission unknown
- Age Effects resources (older more ICU), transmission unknown
- Sex, ethnicity, SES Not amenable to human control, data needed for base rate

- Location Not amenable to human control, data needed for base rate.

Lockdown decision relevant data

International data on transmission is too muddy to be useful. E.g. on social distancing and border controls. However, many countries are adding Covid19 material to routine national health and social care data. The UK has data at individual level via the Health Research Innovation Gateway, but access requires payment and via the Office of National Statistics.

New Research Initiatives

The crisis has led to global sharing of published Mss. and physiological data, e.g. the virus genome; to fast track funding of C-19 projects. Pre-print numbers have massively increased, but most await peer review, and data is often not available.

The COVID19 SYMPTOM TRACKER has close to 3 million UK respondents, with information as to sex, age, postcode, ethnicity, diabetes, heart, respiration, cancer and immunity. Participants state daily: tested, or not; health is 'normal' or 'not quite right', who are then asked about symptoms, blood pressure medication, and isolating behaviour. The C-19 Symptom Tracker Dataset (linked to NHS gateway) is not open. So, project is currently not transparent open science. Variables are not described, and summaries do not reference Mss. Zoe has commercial interests.

The next sweep of the UK household longitudinal study will add to age standardised mortality and social deprivation Covid19 questions on: symptoms and test results, long-term health, caring responsibilities, loneliness, employment and financial situation, home schooling, food and alcohol consumption, exercise and smoking, mental wellbeing.

Currently, NHS, Google, Amazon, etc are developing and delivering Apps so that contacts of C19 positive may be traced and tracked. Information about availability of data, is conspicuous by its absence.

Summary

New data sets in combination with routine data should inform how government lockdown and individual choice might affect the goals of death reduction, economic viability and individual hardship. However

- Most lockdown questions remain open.

- Statements that decision are 'based on the science' must remain questionable wherever data and publications are not available.

It seems to me that while the volume of data, data sources and pre-prints has increased enormously, practices, particularly as to transparency, have, if anything, got worse.