



Tackling COVID-19 through electronic health records

By Dan Ryan for

COVID-19 Actuaries Support Group – Learn. Educate. Inform. Influence

As we in England change our mantra from “Stay at Home” to the more nuanced “Stay Alert” and the possibility of increasing numbers of socially distanced contacts, it is vital to understand the range of data sources that the UK government is using to understand the spread and impact of this crippling disease. And indeed, what data is being shared with us now and what will be shared in the future.

Our changing diet of data, thanks to Public Health England

Back at the beginning of April, there were only three metrics from Public Health England (PHE):

- Numbers of tests (all of the PCR swab variety).
- Numbers of confirmed cases of COVID-19.
- Numbers of deaths (but in those days limited to those happening in hospitals).

Since then, the breadth of modalities of data has spread to cover many different activities in hospitals, care homes and the community.

In particular, since 23 April, PHE has been publishing a COVID-19 epidemiology surveillance summary report each Thursday at 2pm. Each report is a veritable *smorgasbord* or treasure trove of information, that highlights the wealth of different data sources that PHE can call upon, using new and existing channels.

The series of reports aim to provide a picture of COVID-19 in the wider community that is intended to help plan the national response to the pandemic and assist regional stakeholders in local planning.

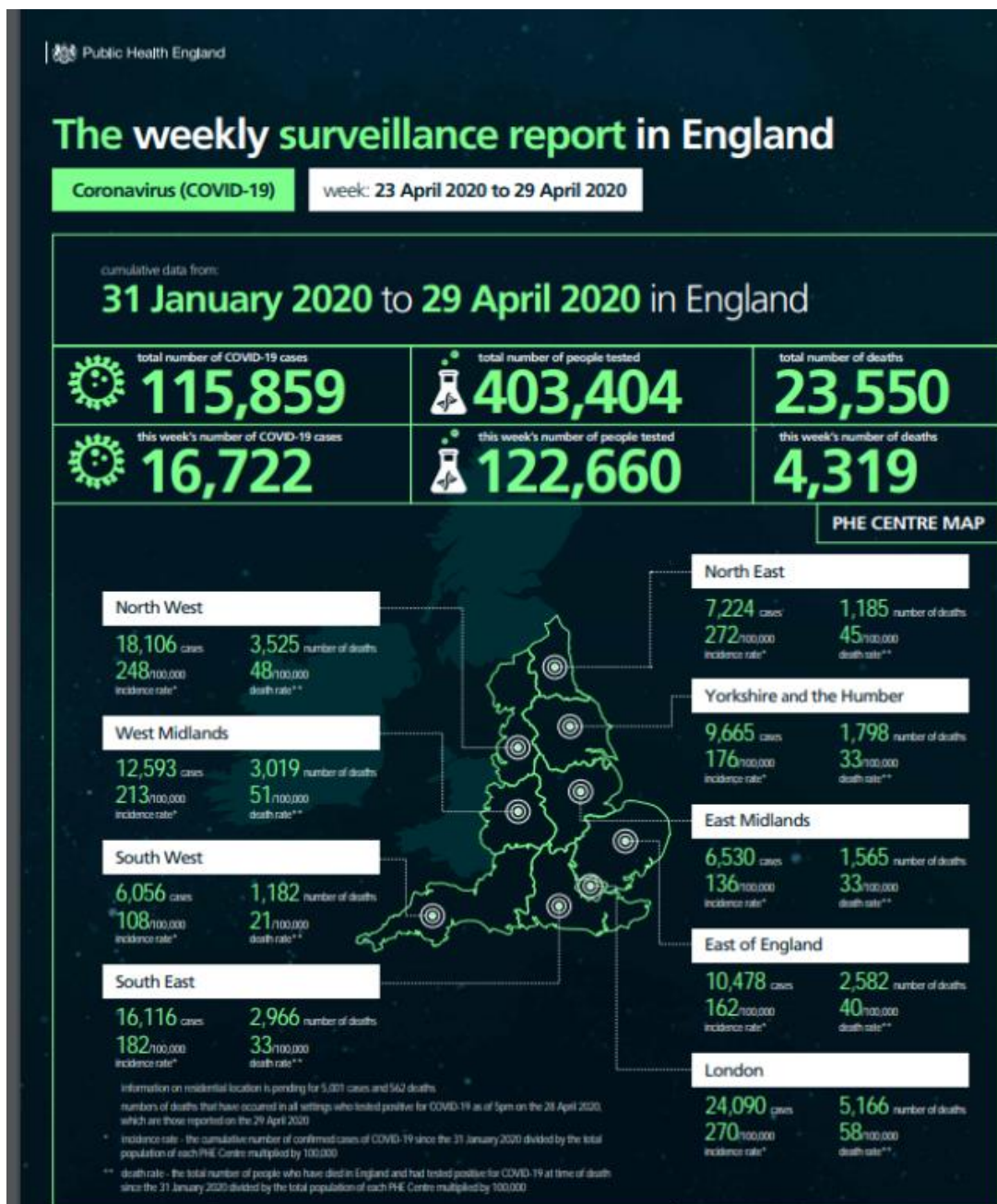
Here is a snapshot of the different key sections in the most recent [20-page report](#) from 7 May, providing data from week 17 (April 20-26):

- 1) *Confirmed cases* – setting out total number of tests and those confirmed with SARS-CoV-2 in England up to 29 April, broken down by age, sex, date of sample, ethnicity and region
- 2) *Community surveillance* – reports on 1,006 new acute respiratory outbreaks from PHE’s Health Protection Teams and groups in the devolved countries and on internet-based surveillance systems covering both Google search queries and a national survey of 4,425 on trends in influenza-like-illness (ILI) using the FluSurvey template.
- 3) *Primary care surveillance* – numbers of GP visits (covering 55% of England), numbers of unscheduled visits and calls during out of hours (covering 70% of England), and the sentinel swabbing scheme at 200 GP practices which shows that 20-30% of those presenting with influenza-like-illness (ILI) are testing positive for SARS-CoV-2.
- 4) *Secondary care surveillance* – attendances at Emergency Departments and admissions to hospital, including data from CHES (Covid-19 Hospitalisation in England Surveillance System) that tests all those admitted with ILI, lower respiratory tract infections (LRTI) or pneumonia for SARS-CoV-2.

- 5) *Virological surveillance* – data collated in Respiratory Datamart from the various PHE laboratories that have been monitoring circulating viruses since the last influenza pandemic in 2009.
- 6) *Mortality surveillance* – cumulative numbers of deaths by date of occurrence split by age, sex and ethnic grouping. As this report only summarises data up to 27 April, it predates the inclusion of deaths in nursing homes from the Care Quality Commission. It remains to be seen how this will be presented in future weeks. And finally, reference to excess mortality analysis provided by the ONS each Tuesday.

All of this is helpfully summarised in [a multi-part infographic](#) together with a [fuller explanation](#) of the underlying sources and motivation:

Figure 1 – Infographic from PHE Surveillance Report for 7 May 2020



Source: Public Health England

An impressive feat of data collation that provides vital insights. And yet. And yet. It is much like the “curate’s egg”: tantalising in parts. It probably meets its core objective which is to provide those inside and outside the health system with reassurance that the situation is being monitored.

It leaves those wanting to develop forward-looking models and scenarios to continue their search. But it gives them the vital sign that such data does exist. All the activity information on GP and out-of-hours calls and visits are but a single slice of the patient electronic health records that collectively are one of the most valuable assets of the nation for research. Closely and correctly guarded, the various custodians of this mosaic of health datasets have supported approved researchers for years from leading research institutions around the world in improving our understanding of health, disease and treatment through tightly governed and ethically approved research projects.

Custodians of our individual health data

One of these custodians, the Clinical Research Practice Datalink (CPRD), has long collected de-identified patient data from GP practices across the UK to provide a longitudinal picture of UK population health^[1]. This dataset alone includes records from 14 million registered patients. Over the last 6 weeks, CPRD has approved 9 different cohort studies from leading research institutions such as University College London, King’s College London and the London School of Hygiene and Tropical Medicine to investigate the many different aspects of our battle with COVID-19. Impressive, but might we not have expected more given the quality and quantity of such data?

Figure 2 – Approved research studies by CPRD focusing on COVID-19

ISAC Approval Date ▼	Title
05/05/2020	Antimicrobial stewardship during the COVID-19 epidemic. Cohort study
05/05/2020	Consultations for non-specific symptoms commonly reported with COVID-19 and other acute respiratory infections in the Clinical Practice Research Datalink: long term trends and patient characteristics
27/04/2020	Underlying conditions in coronavirus (COVID-19) infection: definition, prevalence, mortality and understanding risk
20/04/2020	COVID-19 related risks in cancer survivors: a matched cohort study using linked UK electronic health records data
20/04/2020	Psychological distress associated with the current COVID-19 outbreak: evidence from electronic health records
16/04/2020	COVID-19 and drugs acting on the renin angiotensin system
09/04/2020	Pharmacological risk factors for COVID-19 infection: a matched prospective cohort study of patients in primary care
25/03/2020	Identifying risk factors for infection with the new coronavirus (COVID-19)
23/03/2020	Characterisation, prediction of complications, and safety of emergent therapies for viral outbreaks and pneumonia: a rapid network study to inform management of COVID-19

Source: Clinical Research Practice Datalink

The CPRD and other similar electronic health record datasets contain rich data built up from patient consultations in primary and secondary care, whether directly or through the referral process. Rich data covering diagnoses, laboratory tests, immunisation records, risk factors such as smoking and blood pressure, medications, outpatient and inpatient stays. Linkages with other key datasets dramatically extend the granularity and list of clinical outcomes that can be investigated, such as:

- Small area level data from ONS on deprivation measures
- [Hospital episode statistics](#) from NHS Digital
- Cancer registration, treatment and quality of life data from PHE

Many will be familiar with the attempts in the past to synthesize all these many different healthcare datasets into a single entity, increasing both the power and depth of health insights that could be drawn. Now is not the time to debate or rehash that history. Instead, I would rather highlight how the threat of COVID-19 has led to the rapid development of a new dataset that illustrates how electronic health records can help researchers in identifying new insights and strategies for how we tackle COVID-19 more effectively in the future.

Rapid innovation in unleashing the power of electronic health records

Last week on May 7, we were introduced to the OpenSafely platform that used NHS electronic health records to look for patterns in hospitalised patients who die from COVID-19. This platform has been built in a remarkable 5 weeks, bringing together 24 million patient primary care records into a virtual data centre.

OpenSafely represents a multi-disciplinary collaboration between the [DataLab](#) at the University of Oxford, the [EHR group](#) at London School of Hygiene and Tropical Medicine, electronic health record software companies such as [TPP](#) and wider groups such as [ICNARC](#), working on behalf of NHS England and [NHSX](#), the group in the NHS driving forward the digital transformation of health and social care. The full background can be found at the [OpenSafely website](#). This collaboration released its [first analytical report](#) at the launch, providing multi-variate analysis on which patients are most at risk of death in hospital from COVID-19^[2].

Figure 3 – Multi-variate analysis of mortality risk, including ethnicity and deprivation

Ethnicity	Primary Analysis	Adjusted for completeness of smoking/ ethnicity data	IMD quintile	Primary Analysis	Adjusted for completeness of smoking/ ethnicity data
White	1.00	1.00	1 (least deprived)	1.00	1.00
Mixed	1.64	1.64	2	1.13	1.19
Asian or Asian	1.62	1.62	3	1.23	1.26
British					
Black	1.71	1.71	4	1.49	1.53
Other	1.33	1.33	5 (least deprived)	1.75	1.70

Source: OpenSafely^[2]

OpenSafely is taking a deliberately open approach to both its collaborations with leading research institutions and its analytical approaches. Approved researchers will be able to carry out large scale cohort analyses within the TPP data centre. Further active areas of research would include the following:

- Identify those treatments that increase and decrease both the risk and severity of COVID-19.
- Identify those individuals at highest risk of hospital admission, ventilation or death to inform advice and planning at all levels.
- Use local clinical data to predict local spread and health service need.
- Provide early warnings on disrupted clinical services such as cancer referrals and emergency interventions for heart attacks and stroke to better monitor and measure the likely indirect impact of COVID-19 on population health.

All of which serves to highlight the power of well-maintained electronic health records to answer questions that are simply not possible with snapshots of activity data and limited morbidity and mortality data. Rapid innovation and iteration is enabling research that previously was either impossible or would have taken years to orchestrate.

Our data scientists and researchers have never been more crucial as we chart a course through the maelstrom that is COVID-19, and accessible electronic health records are an invaluable resource in that battle.

Dan Ryan
13 May 2020

REFERENCES

1. Wolf, A., et al., *Data resource profile: Clinical Practice Research Datalink (CPRD) Aurum*. Int J Epidemiol, 2019. **48**(6): p. 1740-1740g.
2. Williamson, E., et al., *OpenSAFELY: factors associated with COVID-19-related hospital death in the linked electronic health records of 17 million adult NHS patients*. medRxiv, 2020: p. 2020.05.06.20092999.