Collaboration across Disciplines for the Health of People, Animals and Ecosystems
L. Stoicu-Tivadar et al. (Eds.)

doi:10.3233/SHTI241077

Clinical Informatics Foundations of 57 Years Sentinel and Genomic Surveillance: Data Quality, Linkage and Access

Simon de LUSIGNAN[a,1] , Mark JOY[a] and Maria ZAMBON[b]
[a]*Nuffield Department of Primary Care Health Sciences, University of Oxford, UK*
[b]*UK Health Security Agency, Colindale, London, UK*
ORCiD ID: Simon de Lusignan https://orcid.org/0000-0002-8553-2641,

Abstract. Sentinel surveillance networks are sophisticated health information systems that warn about outbreaks and spread of infectious diseases with epidemic or pandemic potential, the effectiveness of countermeasures and pressures on health systems. They are underpinned by their ability to turn data into information and knowledge in a timely way. The Royal College of General Practitioners (RCGP) Research and Surveillance Centre (RSC) is one of Europe's oldest. We report its progressive use of technology to improve the scope of sentinel surveillance, with a focus on genomic surveillance. The technologies include terminologies, phenotypes, compute capability, virology including virial genome sequencing, and serology. The RSC's data collection developed from partial, then full extraction of computerised medical record (CMR) data. with increasing sophistication in its creation of phenotypes. The scope of surveillance in 1967 was clinical diagnosis, influenza-like-illness (ILI) was its focus. In the 1992-1993 winter virology sampling started, with progressively more sophisticated sequencing of the viral genome. From 2008 viral sequencing was comprehensive with the Global Initiative on Sharing All Influenza Data (GISAID) the primary repository, supplemented by the COVID-19 Genomics UK (COG-UK) consortium in-pandemic. High quality primary care data captures sociodemographic features, risk group status, and vaccine exposure; linked hospital and death data informs about severe outcomes; virology identified the causative organism and genomic surveillance the variant. Timely data access and analysis will enable identification of new variants resistant to vaccination or other countermeasures and enable new interventions to be developed.

Keywords. Health information systems, sentinel surveillance, general practitioners, data accuracy, medical record systems computerised, phenotype, genomics, vaccines

1. Introduction

Sentinel surveillance is underpinned by sophisticated health informatics systems, that are capable of turning data into information and knowledge in a timely way. Sentinel surveillance warns about potential outbreaks and spread of infectious diseases with epidemic or pandemic potential, its components are described in the World Health Organisation's (WHO) Mosaic Framework [1]. Their scope includes reporting the

[1] Corresponding Author: Simon de Lusignan, Oxford University, UK; E-mail: simon.delusignan@phc.ox.ac.uk.

effectiveness of countermeasures and likely pressures on the health care system. A key component is reference virology laboratory support including viral sequencing.

The Royal College of General Practitioners (RCGP) Research and Surveillance Centre (RSC) is one of Europe's oldest sentinel networks [2]. Its structure and functions have been evaluated against the WHO Mosaic Framework components [3].

We report the RSC's progressive and increasing use of informatics and other technologies to improve its sentinel surveillance, with a focus on genomic surveillance.

2. Method

The RSC has progressively adopted more sophisticated technologies.

Terminologies have changed. Initially a small number of International Classification of Disease (ICD) codes were used to label the conditions monitored by the sentinel network. Primary care clinicians labelled their diagnosis, these were collected in spreadsheets and aggregated centrally. As UK primary care is a registration-based system this allowed weekly reporting of disease incidence; the RSC's Weekly Returns service has monitored 23 conditions, now increased to 34, every week since then. UK primary care computerised in the 1990, initially using the Read terminology but more recently the Systematised Nomenclature of Medicine (SNOMED) clinical terms (CT) [4].

As terminology's sophistication increased so has the RSC's ontological approach and phenotypes. Where possible we exploit SNOMED's expression constraint language (ECL) creating refsets that retain meaning as SNOMED evolves [5].

Compute capability has grown, from using Microsoft Excel and Access to collect data to structured query language (SQL), Figure 1. Practice data extraction has grown from just the monitored condition, reported by age-band, to a comprehensive extract. Pseudonymisation of the unique personal identifier (NHS Number) allows linkage of primary care data to severe outcomes: hospitalisation, intensive care admission and death. Data from these sources is now in a single repository.

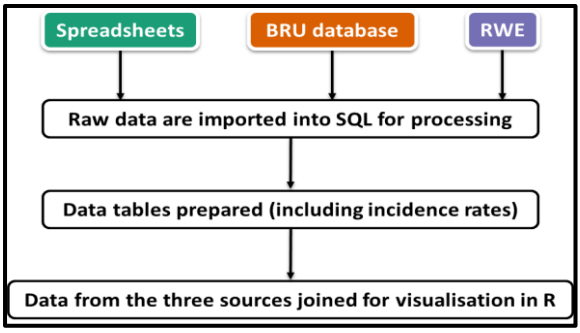

Figure 1. RSC data collection changes, from spreadsheets, to the Birmingham Research Unit (BRU) first database to the Real World Evidence (RWE) SQL server.

Virology sample collection started in the winter of 1992-1993. This started with influenza, with progressive addition of respiratory syncytial virus (RSV), human metapneumovirus (hMPV), then severe-acute-respiratory-syndrome coronavirus-2 (SARS-CoV-2) in 2022. With the start of the pandemic sampling changed from being winter season only with broader virology sampling [3]. Swabs were collected in virus transport media (VTM) and posted to laboratory, with mean time to arrival of 2-3 days

[6]. Analysis was conducted using reverse transcription polymerase chain reactions (RT-PCR) in multiplex formats and updated regularly. Viral genome sequencing has been conducted ever more consistently and in greater depth over this period.

Serology, to observer population levels of immunity and to measure if there is waning vaccine induced immunity was piloted in 2018, and then introduced as an all-year-round activity in 2020.

The ethical approval of surveillance is provided under Regulation 3 of Health Service (Control of Patient Information) Regulations 2002 [7], approved annually by the UK Health Security Agency (UKHSA) Caldicott Guardian.

3. Results

The scope of surveillance in 1967 was clinical diagnosis, of particular interest was influenza-like-illness (ILI). Notwithstanding the lack of virological confirmation this longitudinal data provides considerable insight into the historic incidence of conditions and the impact of vaccination. We show the example of the introduction of vaccination against measles, mumps and rubella (Figure 2).

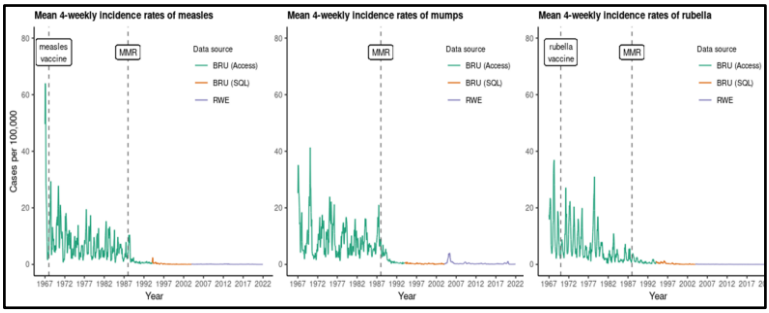

Figure 2. Association of vaccine introduction and incidence of measles mumps and rubella (MMR). The dotted line represents the introduction of a vaccine, the colours the datasets (Figure 1)

High quality primary care data captures sociodemographic features, risk group status, and vaccine exposure, linked hospital and death data informs about severe outcomes [8].

The introduction of virology, 1992-1993 season, for the first time identified the causative organism and genomic sequencing evermore precise details about the variant. From 2008 viral sequencing was comprehensive with the Global Initiative on Sharing All Influenza Data (GISAID) the primary repository; supplemented by the COVID-19 Genomics UK (COG-UK) consortium in-pandemic. The GISAID number is being progressively added to the RSC's linked clinical records so that sociodemographic features, the pseudonymised computerised medical record (CMR) data including vaccine exposure and be linked to health systemwide and death certificate data.

Our surveillance database contains primary care virology samples taken from symptomatic individuals from 1992. 13,619 were positive for influenza, and 3,922 for RSV. We hold 5,073 SARS-CoV-2 results from 2000 onwards. The total numbers sequenced were 2,819 for influenza, 1,251 RSV, and 2,486 for SARS-CoV-2. The proportion where we have linked to the medical records is 97.1%, 96.8% and 98.9% for influenza, RSV and SARS CoV-2 respectively. Methodologies for influenza genomic sequence data evolved from partial sequence analysis of haemagglutinin (HA) and

neuraminidase (NA) genes to whole genome sequence [9], with similar processes for RSV and SARS-CoV-2. Surveillance virology also detected the cross-over of viruses from animal to human, in 2024 a novel swine flu case [10].

Serology sampling exceeded 11,000 samples per year in 2023. These samples, unlike virology are not instantly analysed but instead reserved for analyses around the need for booster vaccinations where immunity might be waning [11].

4. Discussion

To be effective sentinel surveillance needs to be underpinned by sophisticated health informatics infrastructure. Part of this will be predetermined by the nature of the health systems, if there are high quality data collected, whether there are nationally or regionally available datasets and if so whether they can be linked and rapidly accessible for research.

The first components of this are high quality data, particularly primary care data, that can provide near-real-time data about potential cases and is able to calculate an incidence rate ideally though having an accurate population denominator. The data captured needs to include sociodemographic characteristics, if people are in a risk group status, vaccine or other treatment exposure.

The second component is the ability to link at individual patient level to reference laboratory data about circulating pathogens, particularly those that might be resistant to current countermeasures and have epidemic or pandemic potential. Linked hospital and death data informs about sever outcomes. Where virology identifies the causative organism, genomic sequencing should help identify the relevant variant.

The third component is data access. There needs to be timely data access within a high-performance computing environment. This should enable identification of new variants resistant to vaccination or other countermeasures and enable others to be identified. The relative speed with which messenger ribonucleic acid (mRNA) vaccines can be produced makes identifying viral genomic sequence data about resistant strains a finding for which there might be a timely intervention. [12].

There are consortia working across Europe who are able to mobilise data in this way at scale. These include Influenza – Monitoring Vaccine Effectiveness in Europe (I-MOVE), [13] and there are others. However, the informatics is generally the epiphenomenon rather than core. The International Medical Informatics Association (IMA) primary care working group has flagged our disciplines importance [14,15].

5. Conclusions

Sentinel surveillance requires sophisticated health informatics underpinning to ensure data quality, linkage, and prompt access to data in a high performance environment.

Acknowledgements

General practices in the RCGP RSC network who volunteer to take samples for surveillance and share data with us; and their patients and carers who share data and provide samples. The Wellcome Trust funded the creation of a Biomedical Resources

grant https://wellcome.org/grant-funding/people-and-projects/grants-awarded/royal-college-general-practitioners-rcgp-research KHSA are the primary sponsor of the RSC.

References

[1] World Health Organization (WHO). Mosaic Respiratory Surveillance Framework. WHO [Internet] c2023 [cited 28.06.2024]. Available from https://www.who.int/initiatives/mosaic-respiratory-surveillance-framework.

[2] de Lusignan S, Correa A, Smith GE, Yonova I, Pebody R, Ferreira F, Elliot AJ, Fleming D. RCGP Research and Surveillance Centre: 50 years' surveillance of influenza, infections, and respiratory conditions. Br J Gen Pract. 2017 Oct;67(663):440-441. doi: 10.3399/bjgp17X692645.

[3] Gu X, Watson C, Agrawal U, Whitaker H, Elson WH, Anand S, et al. Postpandemic Sentinel Surveillance of Respiratory Diseases in the Context of the World Health Organization Mosaic Framework: Protocol for a Development and Evaluation Study Involving the English Primary Care Network 2023-2024. JMIR Public Health Surveill. 2024 Apr 3;10:e52047. doi: 10.2196/52047.

[4] de Lusignan S. Codes, classifications, terminologies and nomenclatures: definition, development and application in practice. Inform Prim Care. 2005;13(1):65-70. doi:10.14236/jhi.v13i1.580.

[5] Jamie G, Elson W, Kar D, Wimalaratna R, Hoang U, Meza-Torres B, et al. Phenotype execution and modeling architecture to support disease surveillance and real-world evidence studies: English sentinel network evaluation. JAMIA Open. 2024 May 10;7(2):ooae034. doi:10.1093/jamiaopen/ooae034.

[6] Ellis JS, Fleming DM, Zambon MC. Multiplex reverse transcription-PCR for surveillance of influenza A and B viruses in England and Wales in 1995 and 1996. J Clin Microbiol. 1997 Aug;35(8):2076-82. doi:10.1128/jcm.35.8.2076-2082.1997.

[7] Taylor MJ. Legal bases for disclosing confidential patient information for public health. Med Law Rev. 2015;23(3):348-74. doi: 10.1093/medlaw/fwv018.

[8] Leston M, Elson WH, Watson C, Lakhani A, Aspden C, Bankhead CR, et al. Representativeness, Vaccination Uptake, and COVID-19 Clinical Outcomes 2020-2021 in the UK Oxford-Royal College of General Practitioners Research and Surveillance Network: Cohort Profile Summary. JMIR Public Health Surveill. 2022 Dec 19;8(12):e39141. doi: 10.2196/39141.

[9] Galiano M, Agapow PM, Thompson C, Platt S, Underwood A, Ellis J, et al. Evolutionary pathways of the pandemic influenza A (H1N1) 2009 in the UK. PLoS One. 2011;6(8):e23779. doi:10.1371/journal.pone.0023779.

[10] Cogdale J, Kele B, Myers R, Harvey R, Lofts A, Mikaiel T, et al. A case of swine influenza A(H1N2)v in England, November 2023. Euro Surveill. 2024 Jan;29(3):2400002. doi:10.2807/1560-7917.ES.2024.29.3.2400002.

[11] Whitaker HJ, Tsang RSM, Byford R, Andrews NJ, Sherlock J, Sebastian Pillai P, et al. Pfizer-BioNTech and Oxford AstraZeneca COVID-19 vaccine effectiveness and immune response amongst individuals in clinical risk groups. J Infect. 2022 May;84(5):675-683. doi:10.1016/j.jinf.2021.12.044.

[12] Kis Z, Kontoravdi C, Shattock R, Shah N. Resources, Production Scales and Time Required for Producing RNA Vaccines for the Global Pandemic Demand. Vaccines (Basel). 2020 Dec 23;9(1):3. doi:10.3390/vaccines9010003.

[13] Bagaria J, Jansen T, Marques DF, Hooiveld M, McMenamin J, de Lusignan S, et al. Rapidly adapting primary care sentinel surveillance across seven countries in Europe for COVID-19 in the first half of 2020: strengths, challenges, and lessons learned. Euro Surveill. 2022 Jun;27(26):2100864. doi:10.2807/1560-7917.ES.2022.27.26.2100864.

[14] Liaw ST, Kuziemsky C, Schreiber R, Jonnagaddala J, Liyanage H, Chittalia A, Bahniwal R, et al. Primary Care Informatics Response to Covid-19 Pandemic: Adaptation, Progress, and Lessons from Four Countries with High ICT Development. Yearb Med Inform. 2021 Aug;30(1):44-55. doi:10.1055/s-0041-1726489.

[15] Jonnagaddala J, Hoang U, Wensaas KA, Tu K, Ortigoza A, Silva-Valencia J, et al. Integrated Management Systems (IMS) to Support and Sustain Quality One Health Services: International Lessons from the COVID-19 Pandemic by the IMIA Primary Care Working Group. Yearb Med Inform. 2023 Aug;32(1):55-64. doi:10.1055/s-0043-1768725.