**Summary of discussion: JBC Data Science Advisory Board Meeting 25th June 2021, 14:00-14:50
[held virtually]**

Chair: Daniela De Angelis

1. **Previous minutes were signed off and agreed**
- Andrew Morris, Mark Parsons and Wendy Hall sent their apologies.

2. **JBC update**
- Director of D&DS gave an update on current organisational priorities.

3. **JBC presentation: cluster detection analysis**
- The team introduced the overall context of the project.
    - The aim is to investigate how contacts and cases are linked beyond the traditional method which uses postcode-orientated approach for matching home addresses of citizens.
    - The JBC has developed a tool that links "events" reported by contacts interviewed by the contact tracing service – an "event" record is created for each location or setting that a contact reports to have visited e.g. if a person attended a school in the morning and then a pub in the evening, two events would be created, each with a date, location (in terms of postcode) and setting type ("educational" and "recreational/hospitality" respectively). This allows cluster of potential transmission to be assigned to particular settings.
    - Thinking as a system as opposed to just a postcode is important – where are the clusters *actually* arising? In the buildings, between the people in/near/close to that building or another mode that is a mix of both?
    - Collaborations between JBC, PHE, OS and HSE have been key to the work, particularly with respect to high resolution geographical data.
    - A key observation from this work has been that clusters are arising in settings related to the easing of restrictions.
- The presentation next focused on the quality of the data present in CTAS and the ways and mechanisms by which the DHSC team is working to improve the quality of the data.
    - 13.85m CTAS events are currently available which includes 2.86m backwards tracing events (which record citizens movement **before** they started to self-isolate and is the basis of the DCNA tool). Of this, 1.91m events have a valid postcode and date and make up the input date for the cluster analysis.
    - The age and ethnicity columns are partially populated (10.2% not stated and 24.4% not stated respectively) however this is not used during the generation of clusters, and instead is used in indicative post-cluster-generation analysis.
    - The missing nature / data quality of data must be stated with any analysis for transparency.
- The team are working to improve data quality in the following ways:

- o Inclusion of UPRNs – to provide additional information about the exact building in a postcode.
- o Improving postcode data quality.
- o The team has also run qualitative interviews with front-line contact tracers to provide greater insight into the tracing data.

4. Discussion

The main discussion points were as follows:

- A board member asked what constitutes a cluster.
  - o A member of the JBC presenting team responded that a cluster is defined as two or more events sharing the same time, location and setting.
- A point was raised surrounding time windows and whether these are manually set within the tool.
  - o The JBC presenter responded that there are two versions of the tool, one that runs on default settings (I.e. a six-day window between events, one degree of postcode separation and a setting matching criteria). The other version allows manual adjustment of date windows, postcode degrees of separation and tightness of activity matching criteria.
- A board member asked about how the insights from the tool translate into real actions on the ground.
  - o The JBC presenter responded that the tool is trying to bridge the gap between what humans and machines do. The machine provides a statistical likelihood of where transmission is happening. The current version of the tool is being used by humans to understand how to direct resources.
  - o The chair added that it would be useful for the team to distribute the methodologies to the board.

5. Any other business

- A reminder that deep-dives will be halted for the summer and that the next meeting will take place on 30/07/21.
- Thanked DSAB secretariat Aine Fairbrother for taking technical minutes.
- Daniela was thanked for her first session as chair.

Attendance

Dr Ewan Birney	Deputy Director General of EMBL, Director of EMBL-EBI
Dr Sylvia Richardson	Director, Chair of Biostatistics, University of Cambridge
Professor Chris Holmes	Programme Director for Health and Medical Sciences, The Alan Turing Institute
Dr Zeynep Engin	Urban Dynamics Laboratory, UCL

Professor Daniela De Angelis	Professor of Statistical Science for Health, University of Cambridge, Deputy Director and Programme Leader at the MRC-BSU
Professor Simon Vosper	UK Meteorological Office
Dr Ben Goldacre	Director, DataLab, University of Oxford
Professor Graham Medley	Professor of Infectious Disease Modelling, London School of Hygiene and Tropical Medicine
Johanna Hutchinson	Director of Data & Data Science, Joint Biosecurity Centre
Charlie Dunstan-Rice	Joint Biosecurity Centre (Secretariat)
Aine Fairbrother	Joint Biosecurity Centre (Secretariat)
Edward Wynne-Evans	Joint Biosecurity Centre (Presenting)
Mindaugus Uzubalis	Joint Biosecurity Centre (Presenting)
Bella Grant	Joint Biosecurity Centre (Presenting)
Leo Tagg	Joint Biosecurity Centre (Presenting)