Plan Overview

A Data Management Plan created using DMPonline

Title: Using a back-calculation model to estimate the scale of asymptomatic Covid-19 prevalence by age and determine the critical threshold of available susceptible persons within the community

Creator: Prakashini Banka

Principal Investigator: Prakashini Banka, Catherine Comiskey

Data Manager: Prakashini Banka

Affiliation: Other

Funder: Health Research Board (HRB) Ireland

Template: Health Research Board DMP Template

Project abstract:

Classical mathematical epidemiology has highlighted the need to identify a critical population size for an epidemic to drive across a community. This threshold depends not only on the nature of the epidemic but also on the scale of the available susceptible population. Clearly if few are susceptible, as is the case within an immunised community, there can be no epidemic. However if the immunisation rate drops below a certain level, the number susceptible increases and localised, minor epidemics can break out in schools, nursing homes or other vulnerable community settings. In the early stages of a new epidemic where no vaccine is available all persons are susceptible. As the epidemic progresses and the number of infectious individuals increases the number of susceptible individuals will decrease. However when an epidemic can produce both asymptomatic and symptomatic cases the identification of the numbers infected becomes more challenging. Yet it is the estimates of this very number that is required to enable decisions on when a community has reached its critical threshold point and when policy makers and planners can advise on school openings, safety for nursing homes and protection of the vulnerable communities. Mathematical and statistical models of back-calculation have been used successfully both internationally and in Ireland to produce estimates of the scale of a hidden infected population within HIV/AIDS, heroin use and more recently bio-terrorism, where the comparatively short incubation periods are particularly applicable to Covid-19. Working with observed symptomatic cases and the known incubation period, these models predict through the incubation period distribution the total numbers of infected and asymptomatic cases these observed cases arose from. Using back calculation methods with reporting delays, age structure and a range of models for the observable cases, this study will provide crucial estimates for planners of the scale of the asymptomatic Covid-19 population.

ID: 69695

Last modified: 20-07-2021

Grant number / URL: COV19-2020-010

Copyright information:

Using a back-calculation model to estimate the scale of asymptomatic Covid-19 prevalence by age and determine the critical threshold of available susceptible persons within the community

Data description and collection or re-use of existing data

How will new data be collected or produced and/or how will existing data be re-used?

New data will not be collected. Secondary data analysis will be conducted using data collected by the Health Protection and Surveillance Centre (HPSC).

The sample will be a census sample of all cases notified and we will work with our collaborators, Dr. Howard Johnson, the clinical lead at the Health Service Executives Knowledge Management Directorate and Mr. Ajay Ozau, Surveillance Scientist at HPSC who will provide the anonymised dataset for the purposes of our research and will provide feedback on the findings.

The research team consists of four researchers within Trinity. Professor Catherine Comiskey (PI) TCD. Dr Katy Tobin, Co-applicant (TCD), Lead the numerical solution of the models in R. Assistant Professor in Biostatistics School of Medicine. Background in mathematics and statistics and a PhD in mathematical modelling of the transmission dynamics of the HPV vaccine. Dr Sonam Banka and Karen Galligan (Final year PhD student and researcher).

Software such as R, SPSS and Excel will be used on existing data.

What data (for example the kind, formats, and volumes), will be collected or produced?

This study involves a retrospective study design. The TCD research team will receive a dataset of Covid case notifications from our collaborators, consisting of pre collected COVID19 case notifications and deaths in Ireland from the start of the epidemic to December 2020. The sample will be a census sample of all cases notified. The data will be anonymized to TCD.
No names, ID keys or contact information are included in it. The researcher will only have access to data which will be insufficient to identify any individual. The data being provided in this dataset consists of:
Gender
Age bands
County of case origin
CHO area of case origin
Date of symptom onset
Nature of symptoms
Nature of any additional underlying health challenges
Date of test
Date of test result
Nature of test result
Number of close contacts of confirmed case
Date of test report to GP
Date of test report to patient
Date of test result registration
Date of admittance to hospital
County of hospital or CHO area
Reason for admission
Date of admittance to ICC
Date of release from ICU
Date of release from hospital
Date of death
Date death reported
No additional data will be collected by the PI for the purposes of this research.

All of the listed items are chosen to enable the statistical model to explore the timing, gender and age of disease progression across Ireland. In additional the differences between dates of results and dates of reporting are to explore the impact of reporting delays on case notifications and hence on model predictions.

Excel database (.xls, .csv) is approximately 150 kb (input), will be shared and imported into the statistical package SPSS for descriptive statistics and the R package for modelling solutions.

Code files and outputs is estimated to be approximately 200kb.

The database will be stored on the School of Nursing and Midwifery research projects drive in the Comiskey Covid-19 folder located within Trinity College Dublin, all files are appropriately backed up on institutional storage.

Access to this drive is restricted by the TCD IT services and permission to access has to be granted by Comiskey and operationalized by the IT Service department within Trinity.

Documentation and data quality

What metadata and documentation (for example the methodology of data collection and way of organising data) will accompany data?

The methodology of data management and analysis will be available and relevant codes/equations for analysis will be made available.

What data quality control measures will be used?

The final case notification dataset received from the Health Protection Surveillance Centre will not be altered in any way by the research team. The team will be performing statistical analysis and modelling with the data but will not be altering or updating the data set received.
The HPSC are the national centre for case notifications of all notifiable diseases and are responsible for all infectious disease notifications, they have with internal data accuracy and control procedures as required.

Storage and backup during the research process

How will data and metadata be stored and backed up during the research process?

Dataset will be encrypted and transferred by Fileshare.
The database will be stored on the School of Nursing and Midwifery research projects drive in the Comiskey Covid-19 folder located within Trinity College Dublin, all files are appropriately backed up on institutional storage.
All computers storing data will be password protected. The external hard drive and remotely accessible computer will be encrypted and locked in an office on Trinity's campus. Access is restricted to designated staff only.
Documentations defining the variables will be provided by HPSC.
Code and analyses will be documented.

How will data security and protection of sensitive data be taken care of during the research?

All data received will be anonymised. All computers storing data will be password protected. Access is restricted to designated staff only.

Legal and ethical requirements, codes of conduct

If personal data are processed, how will compliance with legislation on personal data and on security be ensured?

No personal data will be collected and will not be processed.

How will other legal issues, such as intellectual property rights and ownership, be managed? What legislation is applicable?

We do not envision any legal or ethical issues, as the this research involves secondary data analysis and findings will be made available publicly in open access journals.

What ethical issues and codes of conduct are there, and how will they be taken into account?

We do not envisage any ethical issues. The team will not collect new data and will access personal data. Data Protection Impact Assessment has been approved by the Data Protection Officer of Trinity College Dublin and ethical approval has been granted for this study by the Faculty of Health Sciences, Trinity College Dublin.

Data sharing and long-term preservation

How and when will data be shared? Are there possible restrictions to data sharing or embargo reasons?

Data acquired from HPSC will not be shared, but codes and equations of modelling techniques will be made available through open access journals.
Some analysis will be conducted using publicly available data and is therefore already available in the public domain.
Following best practice guidance, the dataset will be retained for 7 years after which it will be destroyed by Professor Comiskey

How will data for preservation be selected, and where data will be preserved long-term (for example a data repository or archive)?

Following best practice guidance, datasets will be retained for 7 years after which it will be destroyed by Professor Comiskey. The database will be stored on the School of Nursing and Midwifery research projects drive in the Comiskey Covid-19 folder located within Trinity College Dublin. Data will not be archived or preserved in a data repository.
At a minimum any datasets which are uploaded as supplemental data to journal articles will be considered for preservation. We are investigating domain specific repositories for sharing this data. If suitable domain specific repositories are not found, then data will be preserved in Institutional (TARA) or third party repositories (e.g., Zenodo)

What methods or software tools are needed to access and use data?

Excel, SPSS and R.

How will the application of a unique and persistent identifier (such as a Digital Object Identifier (DOI)) to each data set be ensured?

Codes, equations and relevant supplemental data will be made available through open access research articles and uploaded to appropriate domain specific, institutional or third party repositories.

Data management responsibilities and resources

Who (for example role, position, and institution) will be responsible for data management (i.e. the data steward)?

The research team, consisting of four researchers within Trinity, will be responsible for the data management. Professor Catherine Comiskey (PI) TCD. Dr Katy Tobin, Co-applicant (TCD), Lead the numerical solution of the models in R. Assistant Professor in Biostatistics School of Medicine. Background in mathematics and statistics and a PhD in mathematical modelling of the transmission dynamics of the HPV vaccine. Dr Sonam Banka and Karen Galligan (Final year PhD student and researcher).

What resources (for example financial and time) will be dedicated to data management and ensuring that data will be FAIR (Findable, Accessible, Interoperable, Re-usable)?

Data management plan will be maintained and updated over the course of the project.
Staff time is allocated for data management, we are evaluating the resources which are going to be needed for this task.
Codes, equations and relevant supplemental data will be made available through open access research articles and uploaded to appropriate domain specific, institutional or third party repositories.