**Table 2: Application of the PG-PHASE Framework against three case studies**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Evaluation design and data collection methods** | | | |
| **Case study 1: *Listeria monocytogenes*** | **Case study 2: *Mycobacterium tuberculosis*** | **Case study 3: SARS-CoV-2** |
| **Phase 1** | * Routinely operational laboratory data * Interviews with laboratory personnel * Interviews with end users | | | |
| **Phase 2** | * Interviews with bioinformaticians * Interviews with genomic epidemiologists * Interviews with end users * Developed reports | | | |
| **Phase 3 (Part 1)** | * Interviews with end users * Case studies * Public health data | | | |
| **Phase 3 (Part 2)** | **Pre/post design**   * Characterisation of clusters prior to and following introduction of WGS   + Number of notifiable cases   + Mortality due to notifiable illnesses   + Size of clusters   + Spread of clusters   + Percent of cases linked to clusters * Financial loss due to food recalls prior to and following introduction of WGS | **Retrospective analysis**   * Application of WGS to historical isolates in combination with epidemiological data to identify preventable transmissions   + Identified transmission chains that could have been interrupted   + Estimated cases that could have been averted | **Mathematical modelling**   * Development of mathematical modelling using data from similar contexts where WGS is not being used, in combination with epidemiological data, to estimate impact of public health interventions informed by genomic data   + Estimated cases averted   + Estimated differences in mortality |
| * Publicly available health care costs due to pathogen * Routinely collected data on costs of epidemiological investigation | * Estimated changes to health care costs due to cases averted * Estimated changes to costs of epidemiological investigation due to cases averted and use of WGS to identify transmission links | * Estimated changes to health care costs due to cases averted * Estimated changes to costs of epidemiological investigation due to cases averted and use of WGS to identify transmission links * Estimated changes to costs of infection control interventions due to cases averted and use of WGS to identify transmission links |

**Supplementary Table 1. The evaluation framework: Key activities in utilising pathogen WGS in public health and associated outputs, outcomes and indicators**

| **Stage of the evaluation** | **Key activities (what is done at this stage)** | **Outputs (what are the results of these activities)** | **Outcomes (what is achieved)** | **Indicators (how do we know)** |
| --- | --- | --- | --- | --- |
| Pre-analysis and analysis | * Sample selection, collection and preparation * Laboratory processes to conduct whole genome sequencing * Production of sequence and typing data * Quality control of sequence and typing data * Analysis of sequence data | * Streamlined laboratory workflow processes * Isolate sequence data obtained * Basic information on sequence (type, composition and quality) | * Increased efficiency of workflow processes * Reduction in turnaround times * Reduction in analysis costs * Increased resolution of genotypic data | * Number of samples processed per week * Number of samples analysed per week * Diversity of samples processed per week * Sample processing times * Sample analysis times * Staff, equipment and reagent costs * Levels of satisfaction with workflow processes * Quality control outcomes * Discriminatory power * Typeability |
| Reporting and communication | * Reporting mechanisms between end users and laboratory established * Establishment of reporting requirements for phylogeny and typing data * Preparation of routine reports in a timely manner as per established requirements * Preparation of reports on an ad hoc basis in a timely manner as per established requirements * Mechanisms established to facilitate data sharing across jurisdictions * Mechanisms established to facilitate data sharing across sectors * Establishment of appropriate data governance to support data sharing * Standards are in place to ensure interoperability of genomic data management systems * Standards are in place to ensure potability of genomic data and data systems | * Efficient and effective communication between bioinformaticians, genomic epidemiologists and end users * Routine reports for phylogeny and typing data provided as agreed * Ad hoc reports for phylogeny and typing data provided as agreed * Genomic data routinely shared across jurisdictions to support surveillance, investigation and public health interventions * Genomic data routinely shared across sectors to support surveillance, investigation and public health interventions * Effective use of and contribution to genomic databases | * Bioinformaticians have a good understanding of the needs of end users * Genomic epidemiologists have a good understanding of the needs of end users * End users have a good understanding of the possible uses and limitations of sequence data * Information provided to end users is relevant and responsive to needs * Information provided to end users in understandable and useable form * Improved capacity to share genomic data across jurisdictions and sectors * Improved capacity to identify and respond to geographically dispersed clusters * Improved capacity to respond to clusters involving humans, animals and environmental strains | * Agreements in place regarding reporting processes * Agreements in place regarding data sharing and rights to access * Satisfaction with agreements governing data sharing and rights to access * Mechanisms in place to facilitate data archiving, tracking, tracing, and sharing * Satisfaction with mechanisms in place to facilitate data archiving, tracking, tracing, and sharing * Routine reports requested, issued and received * Ad hoc reports requested, issued and received * Retention of key information by end users * Perceptions of end users regarding their own understanding of the uses and limitations of microbial genomics * Perceptions of end users regarding bioinformaticians’ understanding of their needs * Perceptions of end users regarding genomic epidemiologists’ understanding of their needs * Perceptions of bioinformaticians regarding their own understanding of the needs of end users * Perceptions of genomic epidemiologists regarding their own understanding of the needs of end users * Perceptions of bioinformaticians regarding the use of diverse genomic data systems * End users’ perception of the appropriateness of information received (i.e. quality, quantity, utility) * Geographically dispersed clusters identified * Clusters identified across human, animal and environmental samples |
| Implementation in public health practice | * Genomic data is integrated with epidemiological investigations * Routine surveillance conducted across jurisdictions * Routine surveillance conducted across sectors * Coordinated investigations and actions undertaken to respond to outbreaks within and across jurisdictions and international borders * Coordinated investigations and actions undertaken to respond to outbreaks across sectors | * Genomic data is used to support public health decision-making * Epidemiological links are confirmed or excluded based on genomic data * Identification of transmission networks * Cross-jurisdictional and internationally dispersed clusters routinely identified * Clusters routinely identified across human, animal and environmental samples * Source contamination accurately identified in a timely manner * Food recalls are precise and timely | * Increased confidence in public-health decision-making * Public health policies and guidelines are informed by microbial genomics * Public health interventions are appropriately tailored * Improved understanding of transmission networks * More precise allocation of investigative resources * Reduction in time to identify outbreaks * Reduction in time to respond to an identified outbreak * Reduction in average size of clusters * Improvements in identifying source contamination * Improvements in linking cases to source contamination * Reduced financial loss due to food recalls * Reduction in health care costs * Reduction in illness and mortality | * End users’ perceptions of utilisation of genomic data in public health decision-making * Presence of public health policies and guidelines informed by microbial genomics * Perceptions of affected communities of public health actions and interventions * Indirect consequences (e.g. nosocomial infections leading to ward lockdowns) * Time lapse between identification of cluster and public health action * Size of identified clusters * Proportion of cases linked to identified clusters * Capture of cases identified without epidemiological data * Magnitude of food recalls * Precision of food recalls * Number of notifiable illnesses * Health care costs due to notifiable illnesses * Mortality due to notifiable illnesses |

**Supplementary Table 2. Indicators and data collection methods**

|  |  |
| --- | --- |
| **Data collection methods** | **Indicators** |
| * Routine operational laboratory data | * Number of samples processed per week * Number of samples analysed per week * Sample processing times * Sample analysis times * Staff, equipment and reagent costs * Quality control outcomes * Discriminatory power * Typeability * Diversity of samples processed per week |
| * Interviews with public health laboratory personnel | * Levels of satisfaction with workflow processes * Agreements in place regarding data sharing and rights to access * Satisfaction with agreements governing data sharing and rights to access * Mechanisms in place to facilitate data archiving, tracking, tracing, and sharing * Satisfaction with mechanisms in place to facilitate data archiving, tracking, tracing, and sharing * Perceptions of genomic epidemiologists regarding their own understanding of the needs of end users * Perceptions of bioinformaticians regarding their own understanding of the needs of end users * Perceptions of bioinformaticians regarding the use of diverse genomic data systems * Quality of information provided * Routine reports issued * Ad hoc reports issued |
| * Interviews with end users | * Agreements in place regarding reporting processes * Routine reports requested, issued and received * Ad hoc reports requested, issued and received * End users’ perceptions of utilisation of genomic data in public health decision-making * Retention of key information by end users * Perceptions of end users regarding their own understanding of the uses and limitations of microbial genomics * Perceptions of end users regarding genomic epidemiologists’ understanding of their needs * Perceptions of end users regarding bioinformaticians’ understanding of their needs * End users’ perception of the appropriateness of information received (i.e. quality, quantity, utility) * Perceptions of affected communities of public health actions and interventions * Presence of public health policies and guidelines informed by microbial genomics * Indirect consequences (e.g. nosocomial infections leading to ward lockdowns) |
| * Public health data (i.e. notifiable illnesses, public health surveillance data) | * Time lapse between identification of cluster and public health action * Geographically dispersed clusters identified * Clusters identified across human, animal and environmental samples * Size of identified clusters * Capture of cases identified without epidemiological data * Proportion of cases linked to identified clusters * Number of notifiable illnesses * Health care costs due to notifiable illnesses * Mortality due to notifiable illnesses |
| * Publicly available food recall data | * Magnitude of food recalls * Precision of food recalls |

**Supplementary Table 3: Considerations in evaluation design of Phase 3, Part 2 case studies and suggested evaluation measures**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Evaluation design** | | |
| **Case study 1: Listeria monocytogenes: Pre/post design** | **Case study 2: Mycobacterium tuberculosis: Retrospective analysis** | **Case study 3: SARS-CoV-2: Mathematical modelling** |
| **Availability of data** | * Sufficient data from period prior to the introduction of WGS * Sufficient data from period subsequent to the introduction of WGS * Period of time following introduction of WGS and its incorporation into public health practice is sufficient for effects on health outcomes to be apparent | * Sufficient data from period prior to the introduction of WGS * Sufficient data from period subsequent to the introduction of WGS * Period of time following introduction of WGS and its incorporation into public health practice is sufficient for effects on health outcomes to be apparent * Capacity to retrospectively sequence isolates from the period prior to introduction of WGS | * Sufficient data from period prior to the introduction of WGS and the relevant pathogen * Sufficient data from period subsequent to the introduction of WGS and the relevant pathogen * Period of time following introduction of WGS and its incorporation into public health practice is sufficient for effects on health outcomes to be apparent * Available data from a similar context where WGS was not integrated into the public health response |
| **Suggested measures** | * Number of identified clusters * Number of isolates within clusters * Geographical spread of clusters * Number of cases linked to international data * Percent of isolates linked to a cluster * Number of isolates/clusters traced to a common source * Number of resolved outbreaks * Number of isolates contributed to international databases * Costs of epidemiological investigation * Frequency, magnitude and costs of food recalls | * Estimated number of cases averted * Estimated avoided costs to the health care system * Estimated avoided costs of epidemiological investigations | * Estimated differences in mortality * Estimated impact of public health interventions based on use of genomic data * Estimated differences in costs of epidemiological investigations * Estimated avoided costs to the health care system * Number of isolates contributed to international databases |