



Real-Time Standard Analysis of Disease Investigation (SADI)—A Toolbox Approach to Inform Disease Outbreak Response

Paul Bingham¹, Masako Wada^{2*}, Mary van Andel¹, Andrew McFadden¹, Robert Sanson³ and Mark Stevenson⁴

¹ Diagnostic and Surveillance Services Directorate, Operations Branch, Ministry for Primary Industries, Wallaceville, New Zealand, ² EpiCentre, School of Veterinary Science, Massey University, Palmerston North, New Zealand, ³ AsureQuality, Palmerston North, New Zealand, ⁴ Faculty of Veterinary and Agricultural Sciences, Melbourne Veterinary School, University of Melbourne, Parkville, VIC, Australia

OPEN ACCESS

Edited by:

Eyal Klement,
The Hebrew University of
Jerusalem, Israel

Reviewed by:

Francois Frederick Maree,
Agricultural Research Council of South
Africa (ARC-SA), South Africa
Anette Ella Boklund,
University of Copenhagen, Denmark

*Correspondence:

Masako Wada
m.wada@massey.ac.nz

Specialty section:

This article was submitted to
Veterinary Epidemiology and
Economics,
a section of the journal
Frontiers in Veterinary Science

Received: 18 May 2020

Accepted: 01 September 2020

Published: 29 September 2020

Citation:

Bingham P, Wada M, van Andel M,
McFadden A, Sanson R and
Stevenson M (2020) Real-Time
Standard Analysis of Disease
Investigation (SADI)—A Toolbox
Approach to Inform Disease Outbreak
Response. *Front. Vet. Sci.* 7:563140.
doi: 10.3389/fvets.2020.563140

An incursion of an important exotic transboundary animal disease requires a prompt and intensive response. The routine analysis of up-to-date data, as near to real time as possible, is essential for the objective assessment of the patterns of disease spread or effectiveness of control measures and the formulation of alternative control strategies. In this paper, we describe the Standard Analysis of Disease Investigation (SADI), a toolbox for informing disease outbreak response, which was developed as part of New Zealand's biosecurity preparedness. SADI was generically designed on a web-based software platform, Integrated Real-time Information System (IRIS). We demonstrated the use of SADI for a hypothetical foot-and-mouth disease (FMD) outbreak scenario in New Zealand. The data standards were set within SADI, accommodating a single relational database that integrated the national livestock population data, outbreak data, and tracing data. We collected a well-researched, standardised set of 16 epidemiologically relevant analyses for informing the FMD outbreak response, including farm response timelines, interactive outbreak/network maps, stratified epidemic curves, estimated dissemination rates, estimated reproduction numbers, and areal attack rates. The analyses were programmed within SADI to automate the process to generate the reports at a regular interval (daily) using the most up-to-date data. Having SADI prepared in advance and the process streamlined for data collection, analysis and reporting would free a wider group of epidemiologists during an actual disease outbreak from solving data inconsistency among response teams, daily "number crunching," or providing largely retrospective analyses. Instead, the focus could be directed into enhancing data collection strategies, improving data quality, understanding the limitations of the data available, interpreting the set of analyses, and communicating their meaning with response teams, decision makers and public in the context of the epidemic.

Keywords: decision support tool, outbreak investigations, outbreak response, real-time analyses, centralised data warehouse, big data, foot-and-mouth disease

INTRODUCTION

Concurrent with globalisation and cross-border movements, the opportunity for the emergence of new infectious pathogens in a country has increased substantially (1, 2). Some transboundary animal diseases important for food safety, international trade and livestock production, such as foot-and-mouth disease (FMD), highly pathogenic avian influenza (HPAI), and African swine fever (ASF) can spread rapidly and require a prompt and intensive response if eradication is to be achieved. However, disease eradication responses are usually resource intensive, costly and may not be justified for some diseases. Following discovery of the index case, the competent authority may decide to respond to a disease outbreak by undertaking a control or eradication program. Strategies for disease control or eradication, as well as important factors to consider before embarking on such programs, have been documented (3).

When trying to follow the course of an epidemic and judge the effects of control measures, the routine analysis of up-to-date data, as near to real time as possible, is essential to allow objective assessment of the patterns of disease spread, assessment of the effectiveness of control measures, and the formulation of alternative control strategies (4). This (often iterative) process is rarely formally documented in the published literature, but examples can be found in the United Kingdom's response to bovine spongiform encephalopathy from 1986 to 2012 (5) and the outbreak of FMD that occurred in 2001 (6, 7).

Responses to animal health epidemics increasingly deal with "big data". Some of the challenges of dealing with big data are encompassed in the often described four Vs: volume (relatively large datasets), velocity (the speed that new data is accumulated), variety (integration of multiple sources of data), and veracity (data typically needs accuracy checking, or cleaning) (8–10). Additional to the four Vs described above, animal health responses typically have an additional V: very short time frames.

Data sources for responses typically include laboratory results from traditional and molecular diagnostic methods, animal movement records sourced from national animal movement databases or farmer records, questionnaire interview data, targeted risk based sampling, and opportunistic sampling data. All these data must be underpinned by national farm or animal level demographic datasets.

The key to achieving real-time assessment of ongoing control measures is the presence of a decision support tool, i.e., a data warehouse capable of integrating all data sources and with functions of automated analyses and reporting tailored to the outbreak, and available as early as possible. The components of decision support tools that can be used in animal health have been previously described (2, 11, 12). These should be designed and set up, wherever possible, during non-response (peace time) periods to address the challenges described, particularly ensuring internal validation of the tool, and understanding the limitations and biases in required datasets. Such tools should ideally be centralised, contain relational databases all-inclusively, and ensure that any updates in the system be reflected instantly. Animal disease response activities necessitate that the tools used for management or analysis of data be developed within the

regulatory authority. This is due to real issues around data consistency between response teams, sharing, and confidentiality (13). The software platform plus analytics needs to have both utility and usability in that the analyses can be run frequently and in real time, the interface allows new users to quickly learn and use the tool, and this in turn frees up limited numbers of epidemiology personnel to interpret the analyses and improve data quality. The data and analytics should be accessible to epidemiologists, for exploration and augmentation as required.

As part of New Zealand's biosecurity preparedness, a tool named real-time Standard Analysis of Disease Investigation (SADI) has been developed for performing standardised analyses during disease outbreaks. Our focus was development of a data warehouse together with a standardised set of analyses for use by epidemiologists seconded into a large FMD response, should one occur. Their usual role outside of the response may not include infectious disease epidemiology or the use of programming languages. Therefore, SADI has a simple, user-friendly interface so that the focus can be on improving data quality, understanding the limitations of the data available and interpreting the set of analyses and their meaning in the context of the epidemic. Our goal was to standardise and automate the analyses and increase the time available to interpret and communicate outbreak metrics and patterns. SADI has been used for domestic and international training of epidemiologists in biosecurity outbreaks, and in the ongoing *Mycoplasma bovis* eradication programme since 2017. The aim of this paper is to describe SADI using a hypothetical FMD outbreak as an example. FMD was chosen because it is the major threat to New Zealand livestock industries due to its high contagiousness and significant economic impact. SADI could be modified and applied to other diseases, for example, HPAI or ASF.

MATERIALS AND METHODS

General Description

SADI was developed as a customised project within an application, Integrated Real-time Information System (IRIS)¹ (EpiSoft Ltd, New Zealand). The following sections describe the structure of SADI in terms of the platform, data, and analysis.

Platform

Integration, data management, and analysis were conducted within IRIS.

IRIS is a secure, web based, data management application, based on a dynamic data storage system. All data administration and processing are achieved via a web portal. Multiple portals can be added and customised according to organisational needs. Data can be accessed from any remote location with an internet connection, using any device with a browser. Data storage uses the adaptive object model (14), and access is restricted to authorised personnel using role-based access control.

The application can import and store virtually any type of data including, but not limited to text, images, vector,

¹<http://www.episoft.co.nz/iris>

and raster spatial data. Data are imported into the system using industry standard formats. The framework allows third party applications to communicate with it via web services. Data can then be filtered, sorted and grouped to create customised views. Project managers have the flexibility to

change and modify data schemas as their requirements change over time.

Figure 1 shows how existing databases and response specific field data came together for analyses within SADI using IRIS as the data warehouse and analytical toolbox.

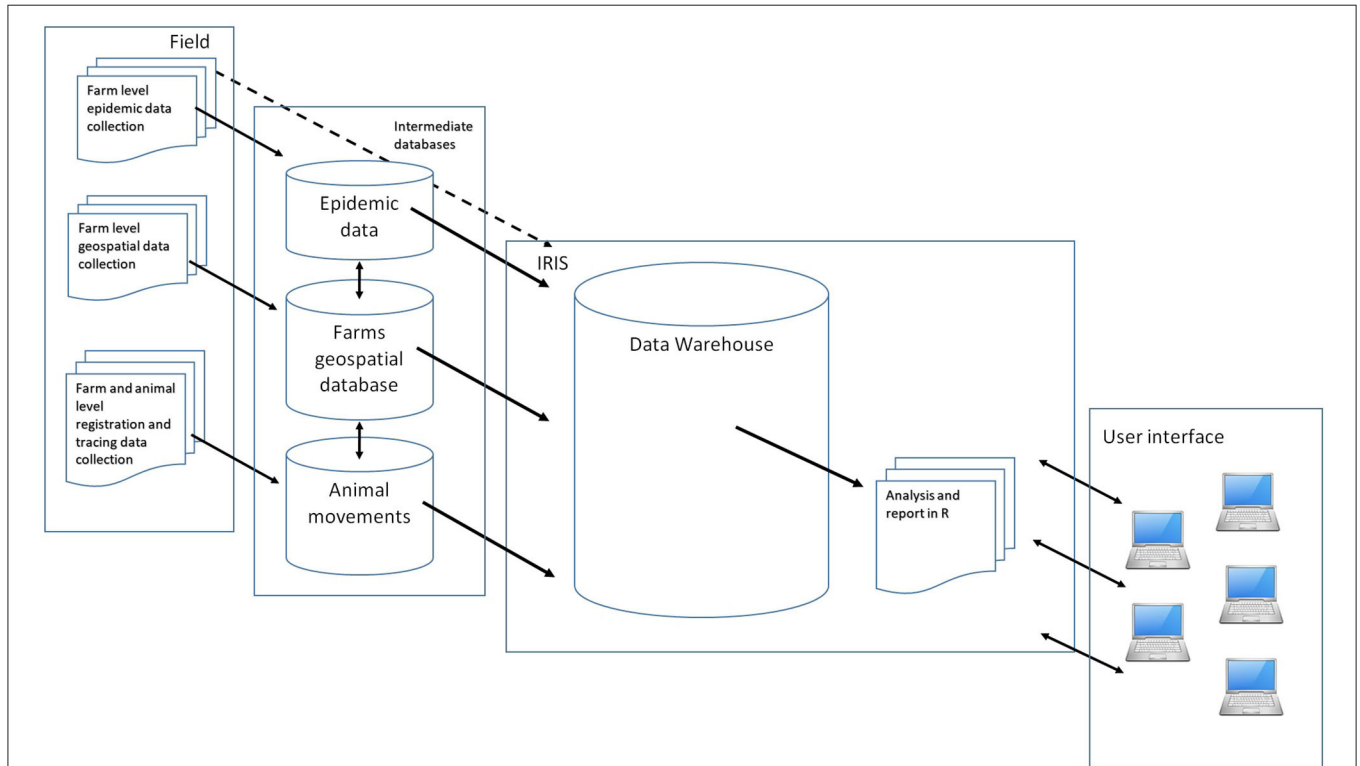


FIGURE 1 | Schematic of the Standard Analysis of Disease Investigation (SADI), showing how relevant field data are collected, integrated, analysed, and reported to the response intelligence using Integrated Real-time Information System (IRIS).

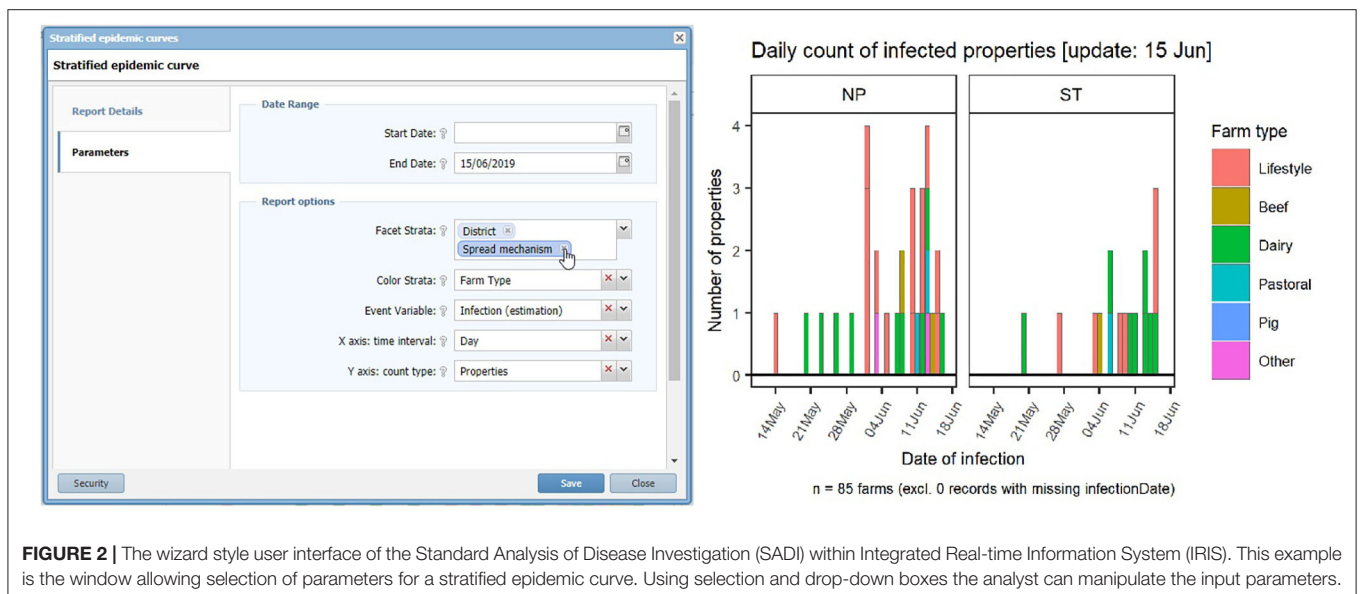


FIGURE 2 | The wizard style user interface of the Standard Analysis of Disease Investigation (SADI) within Integrated Real-time Information System (IRIS). This example is the window allowing selection of parameters for a stratified epidemic curve. Using selection and drop-down boxes the analyst can manipulate the input parameters.

The reporting engine is powered by the R statistical software² (R Core Team, Vienna, Austria). However, IRIS has a wizard style user interface making the running of any R report a relatively simple exercise for non-proficient R users. A typical example of the wizard style user interface, which sits between the user and the R code, is shown in **Figure 2**. Key parameters can be changed easily and the analysis re-run to quickly explore patterns in the data.

Reports can be scheduled to be automatically updated as frequently as required, for example every 24 h, to ensure that the interpretation and the assessments are made based on the most up-to-date data.

Datasets

For our FMD epidemic scenario, three datasets are required to perform the standardised analyses: an outbreak dataset, a tracing dataset, and a population dataset. They are linked by a common field, a unique farm identifier. For other types of diseases, additional datasets (e.g., laboratory data or slaughterhouse data) could be included, if required. The data frame can flexibly be modified, such as accommodating additional fields or using animals as epidemiological units. The database was designed so data fields were comprehensive without redundancy, to avoid data inconsistency within the system.

The outbreak dataset would be supplied from field investigations performed on confirmed infected farms (affected farms with infected animals present). This data for individual farms can be entered into the platform directly in the field, using for example, a handheld device. The data for multiple farms can also be imported into the platform after transcribing field questionnaire data into a comma separated value (CSV) file. Alternatively, the data can be imported indirectly from an external response database. The design of the outbreak data is described in **Table 1**.

As no actual FMD outbreak data was available in New Zealand during the development of this tool, the authors simulated hypothetical outbreak data from the New Zealand Standard Model for FMD (NZSM) (15–17).

The tracing dataset required is described in **Table 2**. During an outbreak, this dataset would be sourced as part of the epidemiological interview and from national livestock traceability systems. As traceability systems are usually focused on live animal movements, both of these methods would be used and possibly others to collect a comprehensive list of possible disease conveyors. In New Zealand, the three main tracing data sources would be epidemiological interview; the National Animal Identification and Tracing (NAIT) system³ (OSPRI, Wellington, New Zealand) (which at the time of publication covers cattle and deer); and the Animal Status Declaration (ASD) system which is a hard copy traceability system covering all FMD susceptible species⁴ (Ministry for Primary Industries; MPI, Wellington, New Zealand).

TABLE 1 | The design of the outbreak data for the Standard Analysis of Disease Investigation (SADI).

Field name	Field description	Data format
FarmID	Unique farm identifier [†]	String
InfectedNo	Infected place number	Integer
NoticeNo	Legal notice number	Integer
PublicReport	If the infection is publicly reported	Boolean
SurveillanceType	Surveillance type which identified the farm	Categorical (public report/surveillance zone/tracing)
VisitDate	First visit date	Date
InfectionDate	Estimated infection date	Date
LesionAge	Age of oldest lesion seen	Numeric
ClinicalDate	Date of onset of clinical signs [‡]	Date
DiagnosisDate	Date of diagnosis or laboratory confirmation	Date
SlaughterDate	Date when slaughter is completed	Date
DisposalDate	Date when disposal is completed	Date
CleanDate	Date when cleaning and disinfection is completed	Date
SpreadMechanism	Likely spread mechanism	Categorical (movement/contiguous property/associated property/local spread [§] / other)
SourceFarm	Likely source farm (if known)	String
DairyInfected	Number of clinically infected dairy cattle	Integer
BeefInfected	Number of clinically infected beef cattle	Integer
SheepInfected	Number of clinically infected sheep	Integer
PigInfected	Number of clinically infected pigs	Integer
GoatInfected	Number of clinically infected goats	Integer
DeerInfected	Number of clinically infected deer	Integer
UpdateDate	Date when the data were last updated	Date

[†]Common farm identifiers to link with the other datasets, i.e., tracing data and population data.

[‡]This field can automatically be filled in based on the visit date and oldest lesion age.

[§]Local spread includes direct and indirect spread mechanisms such as unknown movements, general local farmer activities, aerosol or wind spread, fence contacts, and potentially wildlife spread.

Again, as no actual infected farms were available during development of SADI, tracing data were simulated by the NZSM.

The population dataset needs to be collected prior to outbreak responses and updated regularly, as part of disease preparedness. For New Zealand livestock population, the data were sourced from AgriBase[®] which is a commercially available, comprehensive, spatially explicit, farm level, demographic

²<https://www.R-project.org/>

³<http://www.nait.co.nz/>

⁴<https://www.foodsafety.govt.nz/elibrary/industry/animal-status-declaration/>

TABLE 2 | The design of the tracing data for the Standard Analysis of Disease Investigation (SADI).

Field name	Field description	Data format
TraceID	Trace event unique identifier	String
TraceFarmID	Unique identifier [†] of a farm being traced	String
LinkedPlaceID	Unique identifier [†] of a farm identified as being connected with the traced farm	String
MovementDate	Movement date	Date
MovementDirection	Movement direction	Categorical (forward/backward)
Conveyor	Conveyor type	Categorical (live animal/carcass/genetic material/dairy tanker/other animal product/non-animal product/fomites/other)
Risk	Risk level based on the assessment of the conveyor type	Categorical (low/medium/high/unknown)
CountAnimals	Count of animals moved (if conveyor type is animals)	Integer
DataSource	Data source	Categorical (interview/NAIT [‡] /ASD [§])
UpdateDate	Date when the data were last updated	Date

[†]Common farm identifiers to link with the other datasets, i.e., outbreak data and population data.

[‡]National Animal Identification and Tracing system.

[§]Animal Status Declaration system.

TABLE 3 | The design of the population data (sourced from AgriBase™) for the Standard Analysis of Disease Investigation (SADI).

Field name	Field description	Data format
FarmID	Unique farm identifier	String
Owner [†]	Farm owner	String
Manager [†]	Farm manager	String
Phone [†]	Contact phone number	String
Email [†]	Email address	String
Address [†]	Farm address	String
District	Farm district	Categorical (67 territorial local authorities)
PostCode	Farm postal code	Integer
X coordinates	The longitude of the farm centroid	Numeric
Y coordinates	The latitude of the farm centroid	Numeric
FarmType	Types of farms by animal species, production or management, which are of epidemiological importance	Categorical (dairy/beef and sheep/lifestyle/pigs)
Dairy	Number of dairy cattle	Integer
Beef	Number of beef cattle	Integer
Sheep	Number of sheep	Integer
Goat	Number of goats	Integer
Pig	Number of pigs	Integer
Deer	Number of deer	Integer
UpdateDate	Date when the data were last updated	Date

[†]Data fields that are required for operational response purposes and with restrictive access.

database, describing commercial and non-commercial properties holding production animals in New Zealand (18, 19). The design of the population data is shown in **Table 3**. Details of farms such as the names of farm owner/manager and contact details are not required to perform standardised analyses but are required for other operational response purposes. The access to these data fields can be restricted to authorised persons only.

Analytics

A set of standard analyses was collected by reviewing literature or gathering opinions from MPI staff. The use of these analyses is for summarising and visualising data for response or tracing teams; describing the current situation for informing intelligence and public awareness; building hypotheses about risk factors; or measuring efficiency and effectiveness of the ongoing response efforts.

For each analysis, a report template composed of a variable table, data queries and an R code was created within SADI. A variable table listed a set of input parameters that were necessary for conducting this particular analysis and would accommodate parameter values entered by an analyst, as shown in **Figure 2**. Data queries specified the data fields necessary for conducting this particular analysis. Based on these queries, the most up-to-date datasets were drawn from the internal database in SADI each

time the analysis was carried out. An R script was developed, which would process the datasets using the input parameter values, analyse, and output a report in image (png, jpg, svg, etc.), web page (hypertext markup language; HTML), or map (keyhole markup language; KML) format. Data manipulation and visualisation was commonly conducted using R packages reshape2 (20), plyr (21), and ggplot2 (22).

This set of analyses were programmed to run at an optimal interval (e.g., 24 h) so that the updated analyses would reflect new data values that were entered after the last analyses.

RESULTS

The list of analytic reports that were collected for the use of FMD outbreak response and incorporated in SADI is shown in **Table 4**. There were 16 reports, of which 12 could be used for assessment of response effectiveness and efficiency, seven for informing intelligence and public awareness, five for hypothesis building and four for assisting tracing (some reports were counted multiple times).

With a hypothetical FMD outbreak scenario, infection and detection in 51 farms in New Plymouth and South Taranaki were simulated over 5 weeks. An animated figure showing the spread

TABLE 4 | Summary of analytic reports incorporated in the Standard Analysis of Disease Investigation (SADI) for a foot-and-mouth disease (FMD) outbreak response, with function, input parameters, dependent data* (O, outbreak data; T, tracing data; P, population data), and primary use.

Name	Function	Input parameters	Data*	Primary use
Individual farm timeline	Timeline of a particular infected farm with backward and forward traced movements to/from this farm with the dates and movement details, mapped on a line which represents the course of infection of the farm.	Farm identifier, incubation period range, maximum possible sub-clinically infectious period	O, T	Tracing
Response timeline	Temporal trend of the duration taken to complete each response activity (detection/depopulation/disposal/cleaning) for each infected farm on the scale of calendar date.	Date range, farm sorting criteria	O	Efficiency assessment
Response efficiency boxplot	Temporal trend of the duration taken to complete each response activity.	Strata, response activity (detection/depopulation/disposal/cleaning), time interval	O	Efficiency assessment
Network map	A map showing the point location of a particular infected farm and the farms associated with this infected farm identified from backward or forward tracing. The map was generated as a Keyhole Markup Language (KML) format, allowing it to be zoomed in and out or display the details of farms or movements by clicking the features.	Farm identifier, incubation period range, maximum possible sub-clinically infectious period	O, T, P	Tracing, intelligence/awareness
Outbreak map	A map showing the point locations of all identified infected farms and at-risk farms within a buffer (e.g., 3 km protection zone) or traced farms. The map was generated as a Keyhole Markup Language (KML) format, allowing it to be zoomed in and out or display the details of farms or movements by clicking the features.	Buffer width, maximum possible sub-clinically infectious period	O, P	Tracing, intelligence/awareness, efficiency assessment
Stratified cumulative epidemic curve	Temporal change in the cumulative counts of infected farms (or animals) by estimated infection date or date of diagnosis, with an option to stratify by potential risk factors (e.g., district, farm type, spread mechanism).	Strata, infection stage (infected/diagnosed), count unit (animals/farms)	O	Intelligence/awareness, hypothesis building, efficiency assessment
Stratified epidemic curve	Counts of newly infected farms (or animals) per unit time by estimated infection date or date of diagnosis, with an option to stratify by potential risk factors (e.g., district, farm type, spread mechanism).	Strata, infection stage (infected/diagnosed), count unit (animals/farms), time interval	O	Intelligence/awareness, hypothesis building, efficiency assessment
Number of infectious farms	The total counts of infectious farms each day over time by farm states (subclinical infectious, clinical and undiagnosed, diagnosed and waiting for slaughter).	Maximum possible sub-clinically infectious period	O	Intelligence/awareness, efficiency assessment
Estimated dissemination rate (EDR)	Temporal change in the EDR for n days, calculated as the number of new cases in one time period (day i to day [i-n+1], inclusive) divided by the number of new cases in the previous time period (day [i-n] to day [i-2n+1], inclusive), based on the estimated infection date or diagnosis date. Confidence intervals around each daily EDR were derived from simulated EDRs of 99 iterations using random numbers drawn from Poisson distributions with the calculated numerator and denominator values as the mean. The analysis used an R package epiR (23).	Strata, infection stage (infected/diagnosed), base number of days (n), inclusion of loess smooth line, span for loess smooth	O	Efficiency assessment
Estimated reproduction number (R_0)	Temporal change in R_0 with a week window, with the given serial interval distribution. The posterior mean and 95% credible intervals for R_0 is obtained within a Bayesian framework. The serial interval distribution was based on the parametric method with the given mean and standard deviation of the serial interval. This analysis used an R package EpiEstim (24).	The average number of days from the date of infection to when they become infectious and standard deviation	O	Efficiency assessment
Kernel smoothed density map	A map of kernel smoothed density of infected farms (expected number of farms per square km). The smoothing bandwidth was determined by the standard deviation of an isotropic Gaussian kernel. The bandwidth could be adjusted by a specified factor. This analysis used an R package spatstat (25).	Adjustment factor for the smoothing parameter	O, P	Intelligence/awareness
Nearest neighbour distance	Temporal trend of the distance from a newly infected farm to the nearest infectious farm (potential infection source), with an option to stratify by risk factors (e.g., farm type, district). Distance was computed using a package sp (26).	Strata, plot type (boxplot/histogram)	O	Hypothesis building, efficiency assessment

(Continued)

TABLE 4 | Continued

Name	Function	Input parameters	Data*	Primary use
Areal attack rates	Temporal change in the estimated areal attack rates for each day over time, calculated as proportions of farms that became infected among all the susceptible farms located within a specified buffer from infected farms during their infectious period (27). Distance was computed using a package <i>sp</i> (26). Confidence intervals around the estimated areal attack rates were calculated by the exact method for probabilities using <i>epiR</i> package (23, 28).	Buffer width, maximum possible sub-clinically infectious period, inclusion of loess smooth line, span for loess smooth	O, P	Hypothesis building, efficiency assessment
Area under control	Temporal change in the total area size of a buffer (e.g., 3 km protection zone) around identified infected farms and the total number of susceptible premises within the buffer. The area size was computed using <i>rgeos</i> package (29).	Buffer width	O, P	Intelligence/awareness, hypothesis building, efficiency assessment
First day incidence	Temporal trend of the proportions of animals showing clinical signs in an infected farm on the day of first visit divided by farm type or animal species, which indicates the infectiousness of the farm in terms of forward risk potential during the period from infection to diagnosis (30). Confidence intervals for the incidence risk estimates were calculated using Wilson's approximation (31). This analysis was conducted using <i>incursion</i> package (32).	Strata	O	Hypothesis building, efficiency assessment
Incubation period	The distributions of the number of days since the estimated infection date until the onset of clinical signs based on the data from all identified infected farms.	Strata	O	Tracing, hypothesis building

of this hypothetical FMD epidemic between farms is available as a **Supplementary Material**. A subset of reports produced 4 weeks (25 June 2019) after detection of the index case (22 May 2019) is shown in **Figure 3**. The stratified epidemic curves provided an indication of the temporal pattern of incidence, importance of local spread as the common spread mechanism and predominance of infection in lifestyle blocks (hobby farms) and dairy farms (**Figures 3A,B**). Note using estimated infection dates instead of diagnosis dates removed some of the influence of surveillance intensity after the recognition of disease (**Figure 3B**). The area under control showed the presence of over 1,100 susceptible farms locating within the 3 km radius protection zones, dominated by lifestyle farms (**Figure 3C**). The majority of infectious farms were undiagnosed farms, warranting enhanced surveillance for early detection of these farms, as well as increased capacity for depopulation (**Figure 3D**).

The areal attack rates showed a higher rate of secondary infection within 3 km of infected farms on the 4th weeks of the outbreak, indicating that disease mainly propagated locally (**Figure 3E**). Both districts had an Estimated Dissemination Rate (EDR) decreasing over time and approaching 1 at the time of the analysis (**Figure 3F**). If this trend continued, it would indicate that control measures were bringing dissemination of infection under control. The effective reproduction number (R_{eff}) was approaching 1, which had a similar indication as the EDR (**Figure 3G**).

Figure 4 shows the timeline of a particular farm (ST0029-) that was recently diagnosed (5 July 2019). This timeline demonstrated identifying seven farms as having contacts with this farm in the potential introduction period and potential infectious period. From backward tracing, two farms were identified as the potential source of infection, whereas five

farms were identified as potentially infected from this farm by forward tracing.

Figure 5 is a snapshot of a network map on a particular date (15 June 2019), showing the point locations of the farms in various state (infected, suspect, traced, unknown, at risk). The map also showed the details of a selected farm (NP003xx) as well as two traced movements from or to this farm. This would allow field investigators to prioritise surveillance of linked properties. Additionally, in efforts to identify risk factors associated with disease spread, network analyses could be used to select controls for case-control studies matched on time.

Response timelines depicted the timeliness of response activities for all infected properties and indicated the operational capacity of the response organisation (**Figure 6**). For example, long delays (8–12 days) from the onset of clinical signs to diagnoses were highlighted for three farms (e.g., ST0017-, ST0092-, NP0022-), indicating extra resources may be required to improve communication between farmers and veterinarians and increase public awareness.

DISCUSSION

Here we described SADI, which functions as a centralised data warehouse and performs real-time analyses during a response to an animal health epidemic. This paper demonstrates how the standardised analyses prepared in advance and largely automated, allow description of disease spread as near to real time as possible, assessment of effectiveness of response control measures and input into the formulation of new strategies. By automating the analysis steps and using a user-friendly interface, a wider group of epidemiologists can focus their time away from

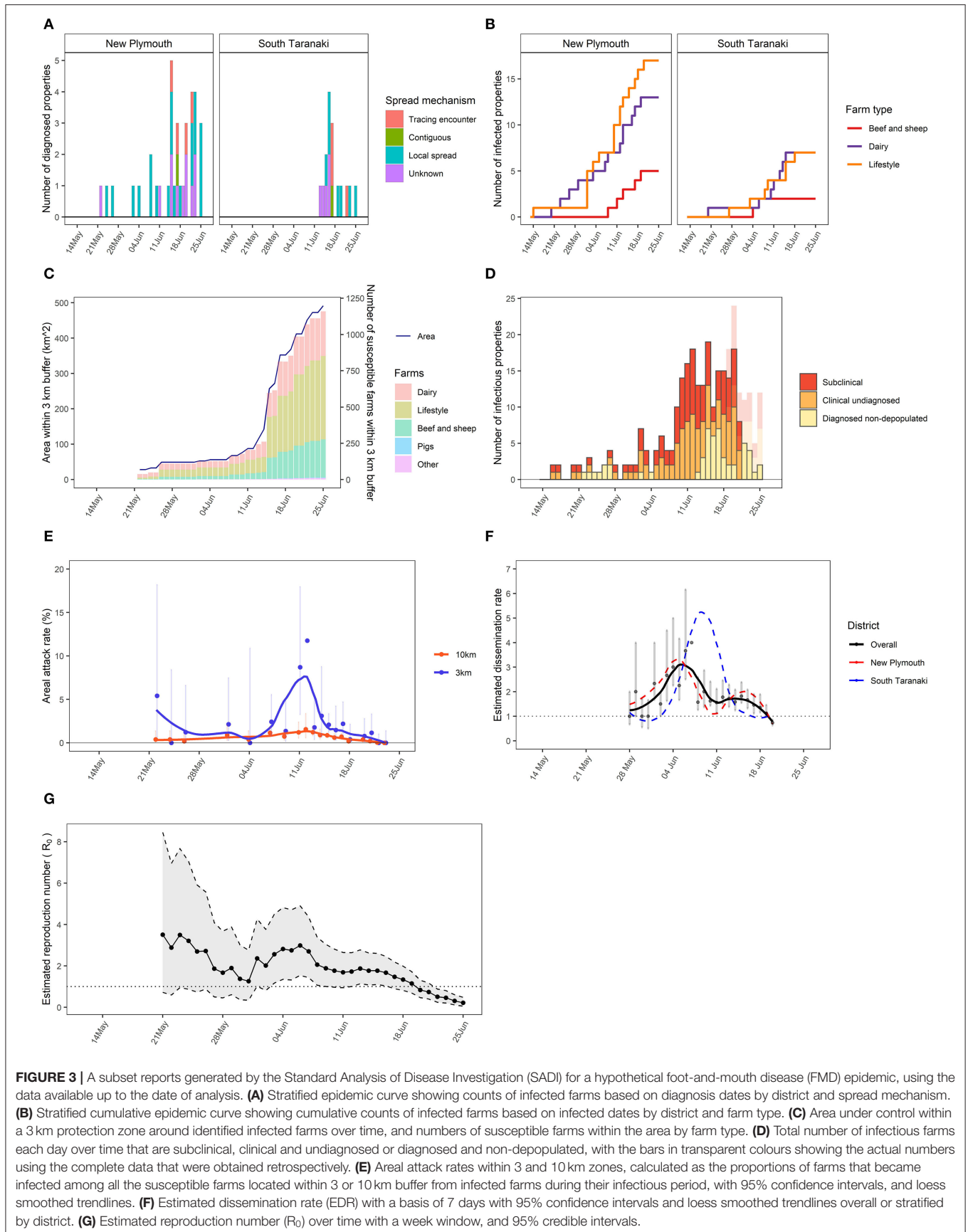
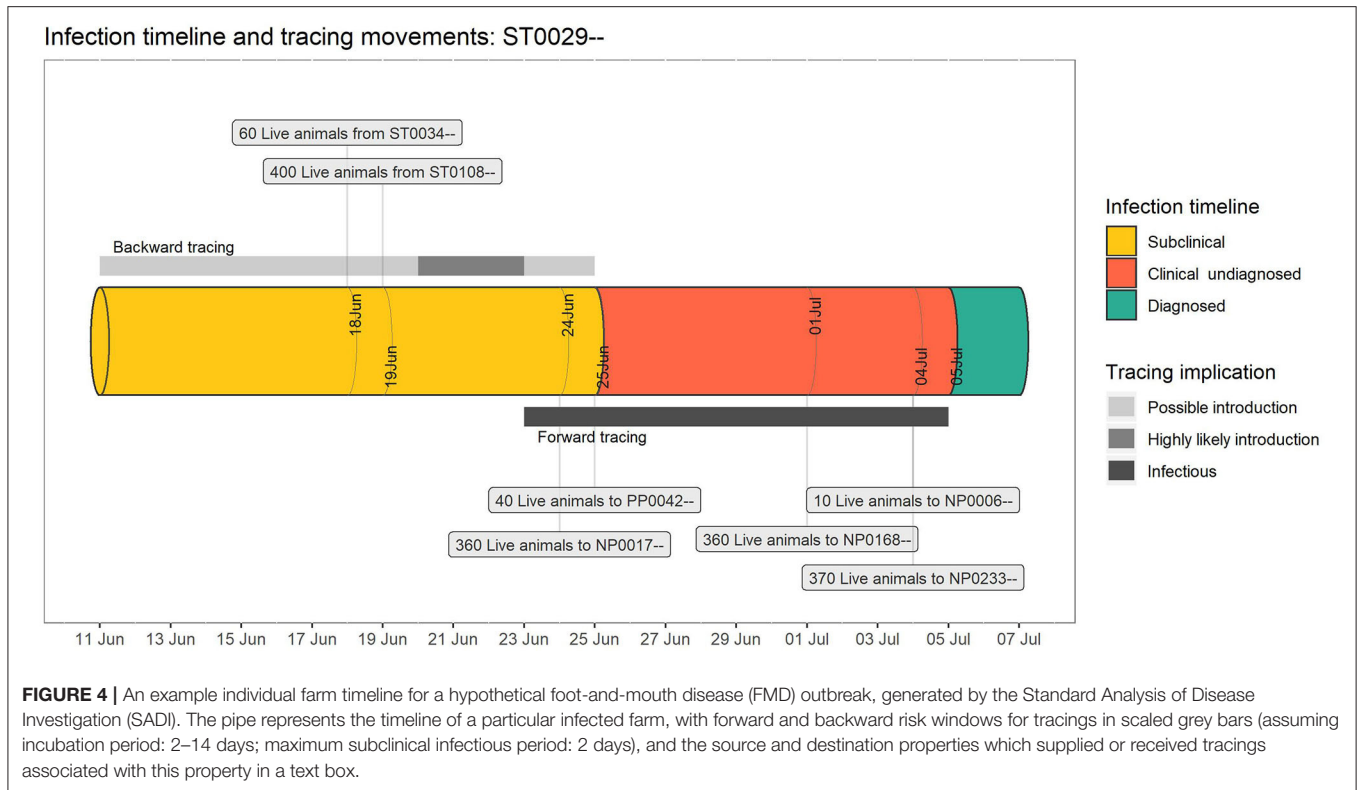


FIGURE 3 | A subset reports generated by the Standard Analysis of Disease Investigation (SADI) for a hypothetical foot-and-mouth disease (FMD) epidemic, using the data available up to the date of analysis. **(A)** Stratified epidemic curve showing counts of infected farms based on diagnosis dates by district and spread mechanism. **(B)** Stratified cumulative epidemic curve showing cumulative counts of infected farms based on infected dates by district and farm type. **(C)** Area under control within a 3 km protection zone around identified infected farms over time, and numbers of susceptible farms within the area by farm type. **(D)** Total number of infectious farms each day over time that are subclinical, clinical and undiagnosed or diagnosed and non-depopulated, with the bars in transparent colours showing the actual numbers using the complete data that were obtained retrospectively. **(E)** Areal attack rates within 3 and 10 km zones, calculated as the proportions of farms that became infected among all the susceptible farms located within 3 or 10 km buffer from infected farms during their infectious period, with 95% confidence intervals, and loess smoothed trendlines. **(F)** Estimated dissemination rate (EDR) with a basis of 7 days with 95% confidence intervals and loess smoothed trendlines overall or stratified by district. **(G)** Estimated reproduction number (R_0) over time with a week window, and 95% credible intervals.



daily “number crunching,” or providing largely retrospective analyses. Instead, the focus can be directed toward optimisation of data collection, exploration of data quality, and quantity prior to any analysis occurring, which then (importantly) enable them to understand the limitations of the data, interpret the analyses produced and provide more immediate advice to other response teams and decision makers. Highly specialised epidemiologists and in particular those with experience with data science, R coding and disease outbreak investigation can be used to refine the analyses in place. To the best of our knowledge, this is the first time such a tool was developed for the livestock population in New Zealand.

Although the system has not been used for real FMD data, it was tested with various simulated FMD incursion scenarios through a series of internal and external workshops involving epidemiologists and programmers. These workshops have helped improving the system, detecting any misfunctions to be fixed and discussing limitations of the system. SADI has also been used for the real outbreaks of *M. bovis* in New Zealand (2017). For the *M. bovis* outbreak, additional analytical reports were developed to meet the specific needs of *M. bovis* epidemiology and response activities. The outcomes of the tool have been communicated widely among the epidemiologists, response teams, tracing teams and decision makers, and demonstrated its value in providing timely information. Particularly, SADI has shown its advantage in timeliness as well as consistency in automatically providing up-to-date reports over 2 years with minimum resource use, in comparison with other systems or the traditional manual approaches.

The ready availability of near-real time graphs, maps and models present some challenges. During a large disease outbreak, staff who are unfamiliar or undertrained, or imported foreign veterinarians may not understand the implicit biases and caveats, misrepresenting the progress of the disease control operation. It is therefore important that these reports are intuitive and clear. There is also a need for cartography standards for outbreak situation reports.

Also, the outbreak data would typically become available with a lag equivalent to the incubation period plus detection delay. Due to this lag, there is a varying extent of difference between real-time analyses using the incomplete data available on the date of analysis and the retrospective analyses using the complete data. Typically, this results in the underestimation of disease risks shortly before the date of analysis (e.g., **Figure 3D**). The analytic reports should be interpreted with caution or the data might be right censored prior to the date of reporting. The tool is therefore best in the hands of epidemiologists who should be involved in communicating at all levels of the programme.

For FMD, a standard set of useful analyses has been described (4). Even though most of these analyses, as well as additions, have been developed in SADI, the method described is equally applicable to most if not all epidemics and probably to all biosecurity domains (domestic animal health; plant health; marine health). As inferred above, a well thought through set of analytics specific to the disease being considered is better prepared in peace-time. Refinements can then be undertaken during an outbreak.

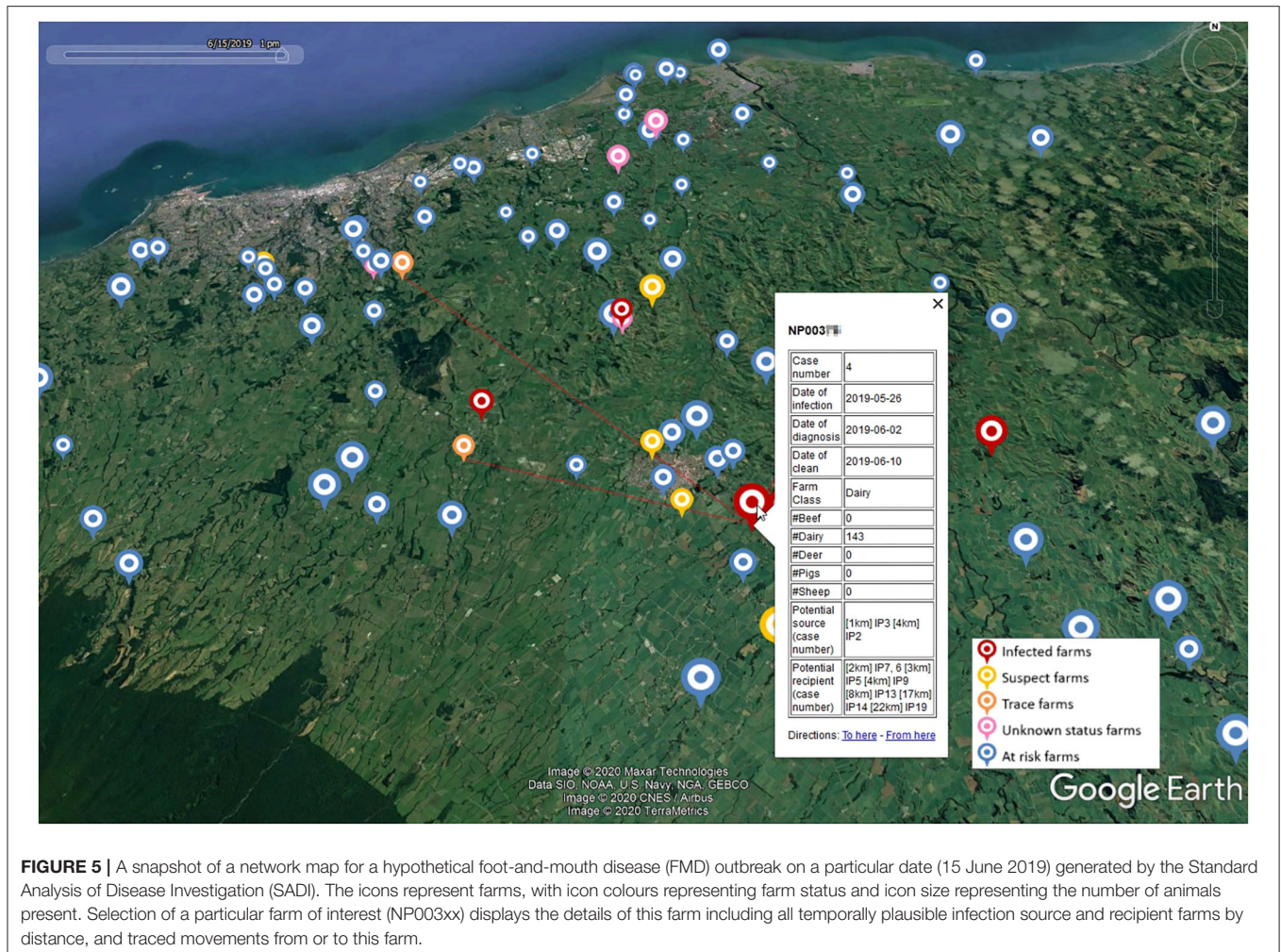


FIGURE 5 | A snapshot of a network map for a hypothetical foot-and-mouth disease (FMD) outbreak on a particular date (15 June 2019) generated by the Standard Analysis of Disease Investigation (SADI). The icons represent farms, with icon colours representing farm status and icon size representing the number of animals present. Selection of a particular farm of interest (NP003xx) displays the details of this farm including all temporally plausible infection source and recipient farms by distance, and traced movements from or to this farm.

Large biosecurity events can occur unpredictably and can put significant, competing demands on the resources of the regulatory authority well beyond usual levels. For high impact diseases such as FMD, many countries have contingency plans in place to allow a pre-programmed set of rapid actions, and set in place a structure for decision making early in the response. This is important because the economic impacts resulting from FMD outbreaks can be enormous (33–36).

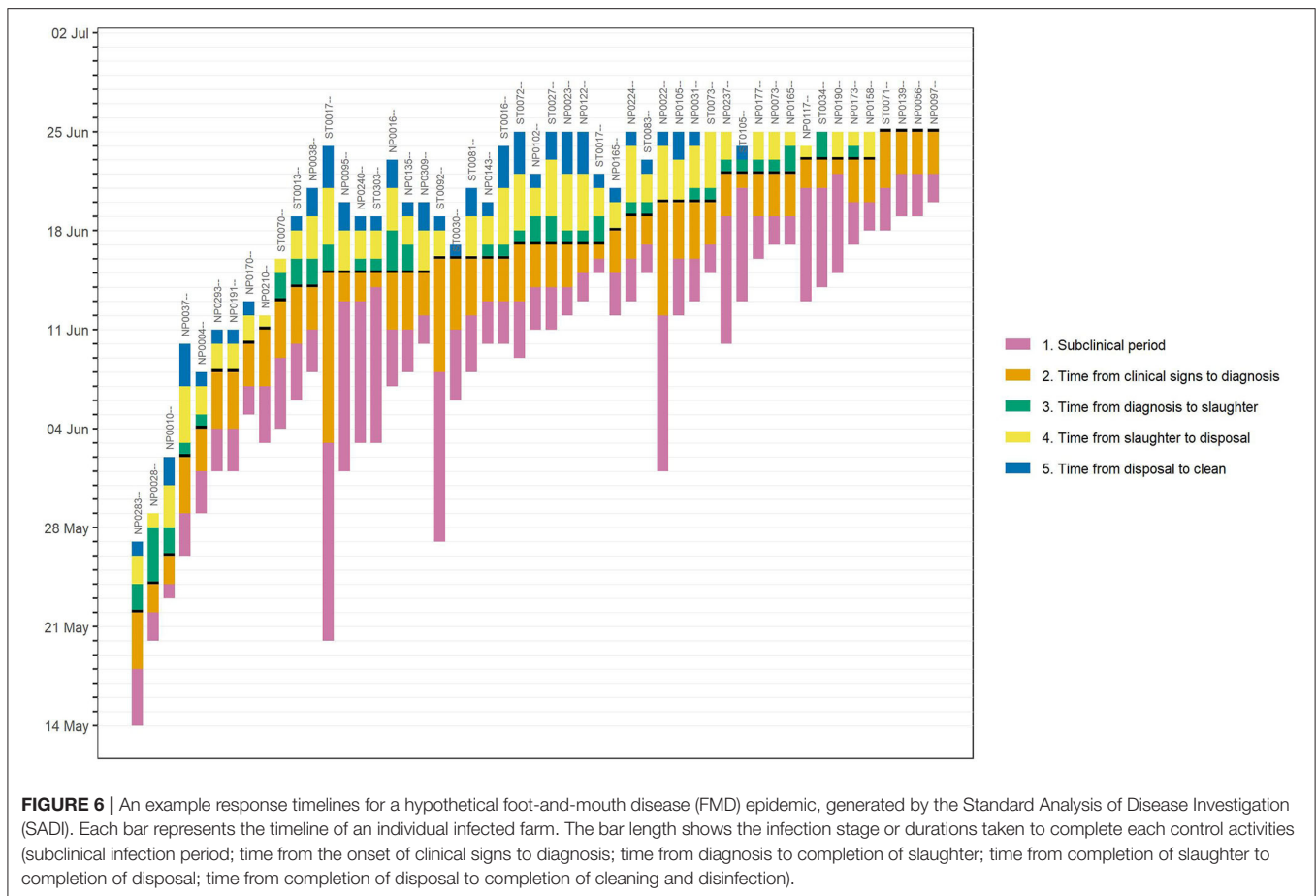
However, even with the presence of response plans, mounting an effective response to a large animal health outbreak can be challenging. Animal health professionals and in particular epidemiologists are well-suited for many roles in disease response and are usually in short supply. To compound this, new and existing staff may have no experience of the disease being controlled, may be unfamiliar with required data sources, data collection and collation methods, or the specific analyses required. Defining the data requirement, setting up data collection strategies and defining and then performing analyses all during the response is not ideal, and is an approach likely to fail.

SADI can form an integral part of the suite of intelligence tools used by epidemiologists during a response. As noted earlier,

many of the data sources used in a response are common to syndromic scanning surveillance. An example would be a national farm demographic dataset which can additionally be used by an epidemic outbreak model. Multiple uses avoid development of tools for siloed applications (13).

As the volume and complexity of infectious disease data increases, professionals must synthesise highly disparate data to facilitate communication with the public and inform decision makers (13). The need for integration of data from a range of sources, into a single data warehouse for analysis is a strong argument in favour of setting up such platforms as a part of readiness between outbreaks. In this paper we have described integration of national farm demographic data, field outbreak data, and individual animal tracing datasets. There are many other possible sources of useful data including laboratory data, industry data such as milk recording at the farm level or meat processing data and vehicle tracing data. If the unit of interest for an outbreak changed from the usual farm level to the individual animal level, other existing data sources will become more common as precision agriculture progresses.

The exploration of data integration including alternative data sources is potentially valuable in augmenting the operation



of the tools and improving the response efficiency. The way that data are generated has changed radically over the last 30 years, mainly as a result of the emergence of electronic methods of measuring, recording, storing, and distributing data (1). Syndromic surveillance systems are becoming increasingly important tools to monitor disease outbreaks by making use of available data (37). Integrating such systems with SADI may help early detection of disease and a prompt start of response activities. While many of these data sources may be protected by legislation from use during “peacetime” surveillance, they could become available during biosecurity responses. The custodians of these datasets may be willing to help ensure integration of data as contingency planning to protect their industries.

While the amount of data potentially available for integration and analysis continues to increase, the development of suitable analytical tools for converting this raw data into useful knowledge has been much slower (9, 38).

Key themes in the development of effective visualisation and analytical tools for infectious disease epidemiology have been described (13). These include: the importance of knowledge regarding user needs and preferences, the importance of user training and the integration of the tool into routine work practices, understanding the complications associated with use of visualisation, the role of user trust and organisational support

in the ultimate usability and uptake of these tools. The paper also noted that individual tools and datasets are rarely sufficient, even for local decision making. Therefore, it is important that the systems under development are tested well in advance by a group of potential users during training exercises, and improvements made through feedback from them. Also, interoperability of tools, data sharing and integration, and sustainability of the tools are important goals that should factor into the design of tools.

Additional to these themes, analytics which are targeted to the objectives of a response and best approaches used in an animal control or eradication program are essential. For example to control an epidemic of FMD, it is essential to understand the mechanisms by which FMD virus is being spread (39). A substantial amount of research has been conducted and described on methods for analysing animal health epidemic data (1, 4, 11, 40–45). However, many of the analyses described were conducted retrospectively and therefore were not available to decision makers in real time during the outbreak. Disease spread patterns are complex, affected by the underlying susceptible population, climate and geography, and the priorities of the stakeholders change over time and vary by country. Therefore, there is no established best strategy that works for every epidemic. The key to successful decision making is based on a good understanding

of the disease in question, based on the timely analysis of the field data.

SADI with simulated outbreak datasets can be used as training materials. The authors have produced simulated FMD outbreaks using InterSpread Plus (46) with deliberately introduced sub-optimal response parameters. The subsequent simulated datasets have then been analysed during training exercises exploring response effectiveness and efficiency. By not dedicating large amounts of time to performing the analyses but rather to interpreting them, rapid understanding of the epidemic and response effort is achieved as well as an appreciation that different analytics are useful at different phases of the epidemic. Conversely, during a true disease outbreak, the standard set of analyses includes response specific parameters for the simulation model. The model would then be tuned to the particular strain of FMD. A set of economic analytics and resource calculators would be a next logical step.

DATA AVAILABILITY STATEMENT

The datasets presented in this article may be available on request. Restrictions apply to the availability of these data, which were used under license for this study.

REFERENCES

- Pfeiffer DU, Stevens KB. Spatial and temporal epidemiological analysis in the Big Data era. *Prev Vet Med.* (2015) 122:213–20. doi: 10.1016/j.prevetmed.2015.05.012
- Stevenson MA, Sanson RL, Miranda AO, Lawrence KA, Morris RS. Decision support systems for monitoring and maintaining health in food animal populations. *N Z Vet J.* (2007) 55:264–72. doi: 10.1080/00480169.2007.36780
- Thrusfield M. *Veterinary Epidemiology*. Oxford: Blackwell (2007). p. 337–48.
- Mansley LM. The challenge of FMD control in the 2001 UK FMD epidemic. In: *Animal Production in Europe: The Way Forward in a Changing World “in-Between” congress of the International Society for Animal Hygiene*. Saint-Malo: Citeseer (2004).
- Ortiz-Pelaez A, Stevenson MA, Wilesmith JW, Ryan JB, Cook AJ. Case-control study of cases of bovine spongiform encephalopathy born after July 31, 1996 (BARB cases) in Great Britain. *Vet Rec.* (2012) 170:389. doi: 10.1136/vr.100097
- Ferguson NM, Donnelly CA, Anderson RM. Transmission intensity and impact of control policies on the foot and mouth epidemic in Great Britain. *Nature.* (2001) 413:542. doi: 10.1038/35097116
- Ferguson NM, Donnelly CA, Anderson RM. The foot-and-mouth epidemic in Great Britain: pattern of spread and impact of interventions. *Science.* (2001) 292:1155–60. doi: 10.1126/science.1061020
- Baro E, Degoul S, Beuscart R, Chazard E. Toward a literature-driven definition of big data in healthcare. *BioMed Res Int.* (2015) 2015:639021. doi: 10.1155/2015/639021
- Gandomi A, Haider M. Beyond the hype: big data concepts, methods, and analytics. *Int J Inf Manage.* (2015) 35:137–44. doi: 10.1016/j.ijinfomgt.2014.10.007
- Hudson C, Kaler J, Down P. Using big data in cattle practice. *In Pract.* (2018) 40:396–410. doi: 10.1136/inp.k4328
- Stevenson MA, Wilesmith JW, Ryan JB, Morris RS, Lawson AB, Pfeiffer DU, et al. Descriptive spatial analysis of the epidemic of bovine spongiform encephalopathy in Great Britain to June 1997. *Vet Rec.* (2000) 147:379–84. doi: 10.1136/vr.147.14.379

AUTHOR CONTRIBUTIONS

All authors have contributed to the work and approved of publication.

FUNDING

This project is funded by Ministry for Primary Industries, New Zealand.

ACKNOWLEDGEMENTS

The authors thank Graham Mackereth (Department of Primary Industries and Regional Development) for his early contribution for the development of SADI. We also thank Bryan O’Leary (EpiSoft Ltd) for his support with IRIS.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fvets.2020.563140/full#supplementary-material>

- Morris RS, Sanson RL, Stern MW, Stevenson M, Wilesmith JW. Decision-support tools for foot and mouth disease control. *Revue scientifique et technique-Office international des épizooties.* (2002) 21:557–64. doi: 10.20506/rst.21.3.1363
- Carroll LN, Au AP, Detwiler LT, Fu TC, Painter IS, Abernethy NF. Visualization and analytics tools for infectious disease epidemiology: a systematic review. *J Biomed Inform.* (2014) 51:287–98. doi: 10.1016/j.jbi.2014.04.006
- Yoder JW, Balaguer F, Johnson R. Architecture and design of adaptive object models. *ACM SIGPLAN Not.* (2001) 36:50–60. doi: 10.1145/583960.583966
- Sanson RL, Stevenson MA, Moles-Benfell N. Quantifying local spread probabilities for foot-and-mouth disease. In: *Proceedings of the 11th International Symposium on Veterinary Epidemiology and Economics*. Cairns: Cairns Convention Centre (2006).
- Sanson RL, Stevenson MA, Mackereth GF, Moles-Benfell N. The development of an interspread plus parameter set to simulate the spread of FMD in New Zealand. In: *International Symposia on Veterinary Epidemiology and Economics (ISVEE) Proceedings*. Cairns (2006).
- Owen K, Stevenson MA, Sanson RL. A sensitivity analysis of the New Zealand standard model of foot and mouth disease. *Revue Scientifique et Technique-OIE.* (2011) 30:513. doi: 10.20506/rst.30.2.2052
- Sanson RL. Agribase-introduction and history. In: *Proceedings of the New Zealand Veterinary Association Epidemiology and Animal Health Management Branch Seminar*. Upper Hutt (2000).
- Sanson RL. The AgriBase™ farm location database. In: *Proceedings-New Zealand Society of Animal Production*. Christchurch: New Zealand Society of Animal Production (2005).
- Wickham H. Reshaping data with the reshape package. *J Stat Softw.* (2007) 21:1–20. doi: 10.18637/jss.v021.i12
- Wickham H. The split-apply-combine strategy for data analysis. *J Stat Softw.* (2011) 40:1–29. doi: 10.18637/jss.v040.i01
- Wickham H. *ggplot2: Elegant Graphics for Data Analysis*. New York, NY: Springer-Verlag (2016). doi: 10.1007/978-3-319-24277-4_9
- Stevenson M, Nunes T, Heuer C, Marshall J, Sanchez J, Thornton R, et al. *epiR: Tools for the Analysis of Epidemiological Data* (2020).

24. Thompson RN, Stockwin JE, van Gaalen RD, Polonsky JA, Kamvar ZN, Demarsh PA, et al. Improved inference of time-varying reproduction numbers during infectious disease outbreaks. *Epidemics*. (2019) 29:100356. doi: 10.1016/j.epidem.2019.100356
25. Baddeley A, Rubak E, Turner R. *Spatial Point Patterns: Methodology and Applications with R*. London: Chapman and Hall/CRC Press (2015). doi: 10.1201/b19708
26. Bivand RS, Pebesma E, Gomez-Rubio V. *Applied Spatial Data Analysis With R*. New York, NY: Springer-Verlag (2013). doi: 10.1007/978-1-4614-7618-4
27. Thrusfield M, Mansley L, Dunlop P, Taylor J, Pawson A, Stringer L. The foot-and-mouth disease epidemic in Dumfries and Galloway, 2001. 1: Characteristics and control. *Vet Rec*. (2005) 156:229–51. doi: 10.1136/vr.156.8.229
28. Collett D. *Modelling Binary Data*. Boca Raton, FL: Chapman & Hall/CRC (1999).
29. Bivand R, Rundel C. *rgeos: Interface to Geometry Engine - Open Source ('GEOS')* (2019).
30. Hutber AM, Kitching RP. The use of vector transition in modelling of intra-herd foot-and mouth disease. *Environ Ecol Stat*. (1996) 3:245–55. doi: 10.1007/BF00453013
31. Rothman KJ. *Epidemiology An Introduction*. London: Oxford University Press (2002). p. 130–43.
32. Stevenson M. *Functions for the Analysis of Infectious Disease Outbreaks in Animal Populations* (2012).
33. James AD, Rushton J. The economics of foot and mouth disease. *Revue scientifique et technique-office international des epizooties*. (2002) 21:637–41. doi: 10.20506/rst.21.3.1356
34. Kim M, Tejada H. Implicit cost of the 2010 foot-and-mouth disease in Korea. *Stud Agric Econ*. (2018) 120:166–73. doi: 10.7896/j.1804
35. Thompson D, Muriel P, Russell D, Osborne P, Bromley A, Rowland M, et al. Economic costs of the foot and mouth disease outbreak in the United Kingdom in 2001. *Revue scientifique et technique-Office international des epizooties*. (2002) 21:675–85. doi: 10.20506/rst.21.3.1353
36. Knight-Jones T, Rushton J. The economic impacts of foot and mouth disease—What are they, how big are they and where do they occur? *Prev Vet Med*. (2013) 112:161–73. doi: 10.1016/j.prevetmed.2013.07.013
37. Dorea FC, Vial F. Animal health syndromic surveillance: a systematic literature review of the progress in the last 5 years (2011–2016). *Vet Med*. (2016) 2016:157–70. doi: 10.2147/VMRR.S90182
38. Kambatla K, Kollias G, Kumar V, Grama A. Trends in big data analytics. *J Parallel Distrib Comput*. (2014) 74:2561–73. doi: 10.1016/j.jpdc.2014.01.003
39. Kitching RP, Thrusfield MV, Taylor NM. Use and abuse of mathematical models: an illustration from the 2001 foot and mouth disease epidemic in the United Kingdom. *Revue Scientifique et Technique-Office International des Epizooties*. (2006) 25:293. doi: 10.20506/rst.25.1.1665
40. Cowled B, Ward MP, Hamilton S, Garner G. The equine influenza epidemic in Australia: spatial and temporal descriptive analyses of a large propagating epidemic. *Prev Vet Med*. (2009) 92:60–70. doi: 10.1016/j.prevetmed.2009.08.006
41. Ward MP, Carpenter TE. Techniques for analysis of disease clustering in space and in time in veterinary epidemiology. *Prev Vet Med*. (2000) 45:257–84. doi: 10.1016/S0167-5877(00)00133-1
42. Farnsworth ML, Ward MP. Identifying spatio-temporal patterns of transboundary disease spread: examples using avian influenza H5N1 outbreaks. *Vet Res*. (2009) 40:1–14. doi: 10.1051/vetres/2009003
43. Thrusfield M, Mansley L, Dunlop P, Pawson A, Taylor J. The foot-and-mouth disease epidemic in Dumfries and Galloway, 2001. 2: Serosurveillance, and efficiency and effectiveness of control procedures after the national ban on animal movements. *Vet Rec*. (2005) 156:269–78. doi: 10.1136/vr.156.9.269
44. Gibbens JC, Sharpe CE, Wilesmith JW, Mansley LM, Michalopoulou E, Ryan JB, et al. Descriptive epidemiology of the 2001 foot-and-mouth disease epidemic in Great Britain: the first five months. *Vet Rec*. (2001) 149:729–43. doi: 10.1136/vr.149.24.729
45. Gibbens JC, Wilesmith JW. Temporal and geographical distribution of cases of foot-and-mouth disease during the early weeks of the 2001 epidemic in Great Britain. *Vet Rec*. (2002) 151:407–12. doi: 10.1136/vr.151.14.407
46. Stevenson MA, Sanson RL, Stern MW, O'Leary BD, Sujau M, Moles-Benfell N, et al. InterSpread Plus: a spatial and stochastic simulation model of disease in animal populations. *Prev Vet Med*. (2013) 109:10–24. doi: 10.1016/j.prevetmed.2012.08.015

Conflict of Interest: RS was employed by the companyASUREQuality.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Bingham, Wada, van Andel, McFadden, Sanson and Stevenson. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.