PERSPECTIVE
published: 17 July 2017
doi: 10.3389/fvets.2017.00110

Translating Big Data into Smart Data for Veterinary Epidemiology

Kimberly VanderWaal[1], Robert B. Morrison[1], Claudia Neuhauser[2], Carles Vilalta[1] and Andres M. Perez[1]*

[1] Department of Veterinary Population Medicine, College of Veterinary Medicine, University of Minnesota, St. Paul, MN, United States, [2] Informatics Institute, University of Minnesota, Minneapolis, MN, United States

The increasing availability and complexity of data has led to new opportunities and challenges in veterinary epidemiology around how to translate abundant, diverse, and rapidly growing "big" data into meaningful insights for animal health. Big data analytics are used to understand health risks and minimize the impact of adverse animal health issues through identifying high-risk populations, combining data or processes acting at multiple scales through epidemiological modeling approaches, and harnessing high velocity data to monitor animal health trends and detect emerging health threats. The advent of big data requires the incorporation of new skills into veterinary epidemiology training, including, for example, machine learning and coding, to prepare a new generation of scientists and practitioners to engage with big data. Establishing pipelines to analyze big data in near real-time is the next step for progressing from simply having "big data" to create "smart data," with the objective of improving understanding of health risks, effectiveness of management and policy decisions, and ultimately preventing or at least minimizing the impact of adverse animal health issues.

Keywords: animal movement, big data, machine learning, modeling and simulation, surveillance

OPEN ACCESS

Edited by:
Alejandra Victoria Capozzo, Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina

Reviewed by:
Fernanda Dorea, National Veterinary Institute, Sweden
Brant Andrew Schumaker, University of Wyoming, United States

****Correspondence:***
Kimberly VanderWaal
kvw@umn.edu

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

Received: *24 November 2016*
Accepted: *22 June 2017*
Published: *17 July 2017*

Citation:
VanderWaal K, Morrison RB, Neuhauser C, Vilalta C and Perez AM (2017) Translating Big Data into Smart Data for Veterinary Epidemiology. Front. Vet. Sci. 4:110. doi: 10.3389/fvets.2017.00110

INTRODUCTION

As our capacity to collect and store data continues to expand rapidly, challenges in veterinary epidemiology are shifting from data acquisition to translating data into meaningful insights about animal health. While human medicine and public health have harnessed big data to optimize "precision" care and track trends in human diseases (1–8), big data in the field of veterinary medicine have been mostly focused on spatial analyses and bioinformatics (9–13). However, the use of big data for animal disease surveillance is a rapidly growing field (14, 15). The promise of big data, as has been witnessed in areas ranging from human health to business and marketing, is the capability to target specific populations and track or even anticipate trends (16). The development and refinement of such capabilities in veterinary epidemiology could significantly improve our ability to identify and respond to emerging animal health concerns, especially if collection and analysis of data occurs in near real-time rather than retrospectively.

Big data typically have certain characteristics, referred to as the four "V's" (**Figure 1**) (16, 17): *Volume* refers to the size of the dataset, which is typically an order of magnitude or more than what has previously been available within a given field (10); *variety* refers to different forms of data that may have been generated for different purposes or collected at different spatial or temporal scales; *veracity* addresses uncertainties in data; and *velocity* refers to the rate at which data are accrued. High velocity data should not be thought of as a "dataset," but rather a "data stream" (17, 18). Applying

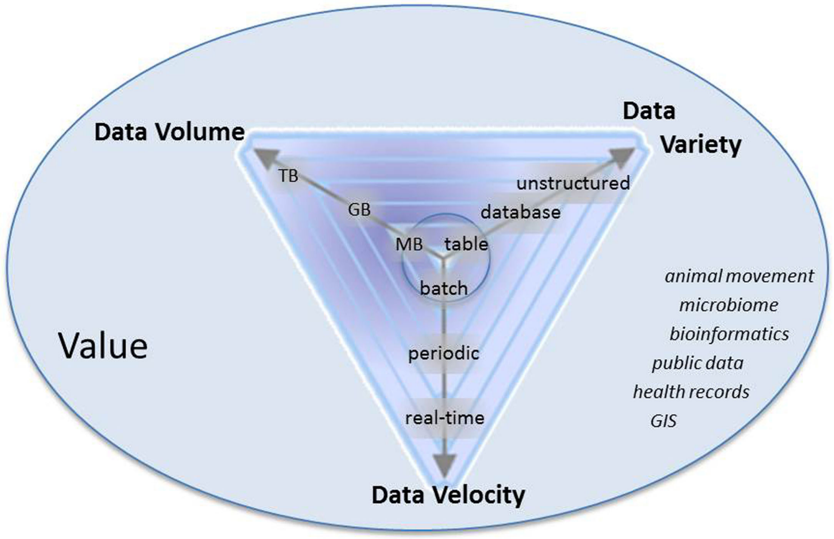

FIGURE 1 | Characteristics of big data: volume, variety, velocity, and value. Arrows represent that data are progressively getting larger (more volume), more variable, and are accruing at faster rates than historically in the field of veterinary epidemiology. Italicized words are examples of types of data in veterinary epidemiology that meet some combination of volume, variety, and velocity.

analytics to volume, variety, veracity, and velocity generates a fifth "V": the *value* of big data to create novel insights and inform decision-making. The analysis of big data, as applied to veterinary epidemiology, is not fundamentally novel compared to traditional or historical practices, but rather differs in complexity, scale, and scope.

Veterinary epidemiological data that are or are becoming "big" include "-omics" data, geospatial data, publically available data repositories such as World Animal Health Information System[1] and EMPRES Global Animal Disease Information System (Empres-i[2]), clinical data or digitized health records from both companion and food animals, data on animal movement from local to international scales, and production data from food animal industries (**Figure 1**) (14, 15, 19). The analysis of such data can be used to understand health risks and minimize the impact of adverse animal health issues by, for example, increasing the effectiveness of control and surveillance by identifying high-risk populations through the analysis of spatial and animal movement data; combining disparate data or processes acting at multiple scales through epidemiological modeling approaches; and harnessing high velocity data to monitor animal health trends and for early detection of emerging health threats.

Generating and storing big data are becoming increasingly easy, but we now face challenges in translating the abundance of available data into meaningful information. This challenge, combined with the capability to analyze epidemiological patterns in near real-time, creates a need to develop effective tools and data pipelines to move from simply having "Big Data" into the creation of "Smart data." Using the four V's as an organizing framework, we review recent examples of big data analytics and highlight insights gained from approaching veterinary epidemiology with a big data perspective.

VOLUME: IDENTIFYING HIGH-RISK POPULATIONS WITH BIG DATA

Collecting and analyzing very large data sets has become increasingly common as technologies for storage and computation advance. For example, research utilizing bioinformatics approaches, detailed data on the demographics and movements of animal populations, and large scale spatial datasets routinely generate terabytes of data, stimulating a new frontier of advanced analytics to handle such data (9–11). Here, we do not provide an exhaustive review of the use of high volume datasets in veterinary epidemiology, but rather select a few diverse examples that highlight the potential use of big data to identify high-risk populations.

Risk of infection is rarely homogenous in a population, and the ability to identify heterogeneities in risk allows for targeted surveillance and control measures. Risk-based strategies are typically more cost-effective than non-targeted strategies, both in terms of early detection and rapid control of a disease (10, 20–22). Because movement of animals between locations is a key risk factor for many infectious diseases, many countries now implement mandatory animal traceability programs (23–26). For example, national or multinational programs, such as the European Union's Trade Control and Expert System and the United Kingdom's Cattle Tracing System, track shipments of production animals across space and time, generating a rich source of information for rapid response to health threats (27–29). In the absence of national regulatory frameworks, large production companies often keep

[1] http://www.oie.int.
[2] http://empres-i.fao.org.

records on the movement of animals between company farms (30). Movement data from a single swine production company in the US contained information on the origin and destination of 9.1 million pigs annually, totaling ~25,000 per day. Such databases can be used to construct contact networks that represent potential transmission pathways in a population, and social network analysis can be used to quantify the connectivity of each node within the network and to assess the population's vulnerability to infectious disease epidemics (26, 31–33). Identifying premises that likely play critical roles in disease spread, such as highly connected farms or farms lying on bottlenecks within the network, can inform control measures that are more effective at limiting disease spread than non-targeted approaches (10, 21, 25). Given the high velocity nature of animal movement data, it is relatively easy to envision how risk estimates could be updated in near real-time, provided that data are efficiently captured in the field, analyzed, and reported to decision-makers.

Substantial spatial heterogeneities exist in the occurrence of infectious diseases, and management and analysis of large spatial datasets represents another facet of voluminous data (34). Numerous spatially explicitly datasets exist for environmental and climatic factors [e.g., Ref. (35)], land cover and use [e.g., Ref. (36, 37)], distributions of at-risk, reservoir, and vector populations [e.g., Ref. (37–39)], and satellite imagery and remote sensing products (40). In addition, the increasing use of GPS tracking devices creates a rich source of data on the movement of people, vehicles, and animals that can be used to dynamically represent exposure and transmission dynamics (41). These data can be combined with geo-referenced disease data [e.g., Ref. (42)] to identify environmental correlates of disease through ecological niche modeling, thus contributing to our ability to understand and map a pathogen's geographic distribution (9, 43–45). By utilizing near real-time updates in environmental data and locations of new cases, risk maps can become evolving rather than static representations of risk (43, 46). Remote sensing, in particular, could be re-framed as a high velocity source of data, as many satellite-based data are updated at regular intervals (43). Ultimately, the ability to predict the occurrence of pathogens through space and time will allow for more effective targeted surveillance and control.

VARIETY: COMBINING DISPARATE DATA

The challenge emerging from the need to assemble datasets from multiple, disparate sources is not new within epidemiology. Analysis of such data is complicated in that data are often aggregated at different spatial and temporal scales, and datasets must be aggregated or disaggregated to harmonize the spatiotemporal scale of the consolidated dataset. Even when combining a single type of data (i.e., diagnostic records) from various institutions, inconsistencies in data structure and vocabulary must be mapped to make the data interoperable.

The use of universally recognized data formats is important for enhancing connectivity of data between different sources (e.g., different laboratories and clinics) (47, 48). Human diagnostic laboratories have long had standardized vocabularies for health records, such as HL7, LOINC, and SNOMED, but standardized vocabularies are typically underused in veterinary medicine (49). To overcome this challenge, the Clinical Wildlife Health Initiative has worked to create a standardized terminology for clinical signs in admitted wildlife. This standardization enhances the ability to pool data from multiple clinics into a common dataset, thus creating a powerful network of clinics in which health trends could be tracked (50). Similar efforts are underway for swine diagnostic data (48). Related to the current lack of data standardization, many sources of animal health data are not readably usable in statistical models, such as "unstructured" text-heavy data (15, 49, 51). While it is possible to use text-mining techniques to extract information from unstructured text fields in clinical or diagnostic records (52, 53), homogenizing and naturalizing data into uniform and standardized formats is critical for maximizing accuracy and ensuring smooth automation (8).

Finally, data relevant to disease dynamics are representative of processes that operate at different spatial and temporal scales. Epidemiological modeling provides a means to connect processes across multiple scales and account for the inherent dynamic elements of disease systems. For stochastic disease models, thousands or even hundreds of thousands of simulations are performed to understand the behavior of the system and optimize parameter values, thus requiring extensive computational resources and generating big data. Machine learning techniques, such as random forests and genetic algorithms, are used to optimize parameter values so that the model simulates epidemiological dynamics that closely resemble real-world data. Computational modeling provides an effective means to link data to processes, and understand mechanistically how disparate data may interact to influence the occurrence of disease.

VELOCITY: HARNESSING HIGH VELOCITY DATA

Of the Vs of big data, velocity represents the largest departure from traditional data processing in veterinary epidemiology, but it also has the most potential for revolutionizing the field, particularly in regard to monitoring and surveillance (2, 3, 7, 15, 19, 54). Sources of high velocity data include digitized records from clinics and diagnostic laboratories, analysis of Google search trends and social media, and mortality and abattoir data (2, 3, 15, 49, 54–56). Analysis of data through time allows for the establishment of baselines to which emerging data can be compared (49). The typical values for a metric that relates directly or indirectly to disease (e.g., incidence or production levels) are summarized for a particular population or spatial location, and deviations outside the normal variation of the metric can be used as an indicator of an outbreak (57, 58). Time-series analysis provides an additional method for mathematically quantifying temporal patterns, incorporating seasonality and long-term trends (58). Short-term predictions can then be made in terms of expected incidence or prevalence over time. Departures from expectations can be considered "anomalies" and may serve as early warnings for emerging threats or altered disease dynamics that warrant further investigation or intervention (54, 56, 58).

For example, a recent initiative involving 700 veterinary hospitals in the US tracked the daily proportion of patients with certain clinical signs, contrasting new data with averages from a retrospective period of time. As a proof-of-concept, the system was able to rapidly detect a simulated outbreak scenario and generate an outbreak alert (19). Similarly, daily condemnation rates of pig carcasses at abattoirs in Canada were evaluated to detect aberrations in the data stream that may signal a disease outbreak. Aberrations detected in a retrospective analysis coincided with several documented disease outbreaks in swine, thus demonstrating the potential timeliness of a syndromic surveillance system based on abattoir data streams (59). Finally, Guernier et al. (54) found that Google search trends could be used to track the occurrence of tick paralysis in companion animals in Australia, and certain search terms could potentially be used as early indicators of high-risk periods. These examples highlight the potential for high velocity and high volume data to enhance surveillance capabilities. Such applications may be particularly relevant for syndromic surveillance, where the causative agent may not be identified and the focus of the analysis is on tracking suites of clinical signs, or a syndrome, that may be associated with an endemic or emerging disease (49).

The capability to acquire and analyze high velocity data streams requires the development of data pipelines, which can automate and streamline the processing of big data for real-time insights and rapid response. For example, the Morrison Swine Health Monitoring Project (MSHMP) is a high velocity data stream that is effective at tracking the spatiotemporal dynamics of several, high-impact infectious diseases in the US swine industry (60, 61). As of November 2016, MSHMP included data from more than 1,000 sow farms managed by 29 production companies; the status of ~46% of breeding sows in the US are tracked through MSHMP on a weekly base. In this case, the data pipeline involves (a) capture of data from disparate data sources, including veterinarians and diagnostic laboratories, (b) storage and (c) processing data to prepare datasets for (d) data visualization and analysis to (e) ultimately interpret and report up-to-date trends in incidence, prevalence, genetic diversity, and spatial occurrence of swine diseases in the US (**Figures 2A–C**). This data pipeline is being used, for example, to establish dynamic baselines for porcine reproductive and respiratory syndrome (PRRS) virus incidence, detect the onset of seasonal PRRS epidemics, and provide value to participating producers.

CREATING VALUE: BIG DATA CHALLENGES AND OPPORTUNITIES IN VETERINARY EPIDEMIOLOGY

In the coming decades, the greatest challenge in big data epidemiology will be to move toward creating value. Putting big data to work requires expanding our definition of the V's of big data to include three A's: accuracy, accessibility, and automation. While some may purport that the sheer quantity of big data means that inaccuracies in the data are washed out, this may not be true if issues with confounding, measurement error, and selection bias scale with sample size (5, 47, 62). For example, data mining based on Google searches or Twitter may misrepresent the population at risk given biases in internet use. In addition, the value of extracting data from clinical records or diagnostic laboratory databases is dependent on the quality of record keeping and data entry. Thus, error checking and quality control should be incorporated into big data processing to ensure reliability (47). Accessibility it also a critical challenge, encapsulating fundamental concerns related to data confidentiality and ownership (8, 14, 18, 47), data engineering issues revolving around data structure

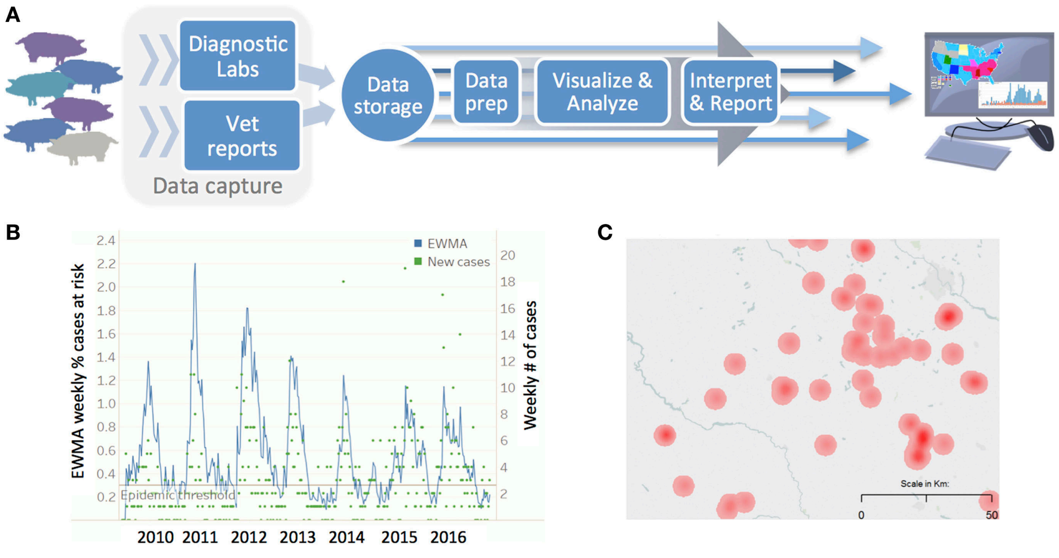

FIGURE 2 | **(A)** Data pipeline utilized by the Morrison Swine Health Monitoring Project for generating near real-time insights about the spatiotemporal incidence of porcine reproductive and respiratory syndrome (PRRS) virus, including weekly reports on the **(B)** incidence of PRRS, with trends reported as an exponential weighted moving average (EWMA), and **(C)** heatmaps of PRRS risk based on the geographic distribution of sow farms shedding PRRS.

and connectivity (8), and limited availability of trained personnel capable of extracting data from databases (14). Finally, high velocity data create a need to automate data pipelines for routine and repeated use. Automation is key for harnessing big data for monitoring and surveillance (8, 49, 56).

A major criticism of big data analytics is that it lacks the rigor of hypothesis-driven, controlled experiments for determining causation (16, 62). However, correlations identified through the analysis of big data are useful for hypothesis generation and prediction (5, 17). In addition, the increasing numbers of measurable explanatory factors available from diverse sources necessitate the use of relatively new (to veterinary epidemiologists) statistical approaches, such as machine learning, that are more appropriate for handing datasets with a large number of covariates (17, 18, 63–65). Due to the large number of potential variables, care must be taken to identify spurious correlations (5, 66), and the use of large datasets does not necessarily increase a study's validity. Sound epidemiological principles for the interpretation of observational data are required (62).

As in all long-term monitoring programs, sustainability of big data surveillance and monitoring efforts is a constant challenge (49). For example, voluntary reporting programs such as MSHMP rely on weekly reporting by veterinarians, and adoption of new data standards and sharing of data across organizations requires investment of time, resources, and complicated data-sharing agreements. Even ensuring that all data fields are complete in clinical or diagnostic records (such as location data) requires investment of time and diligence by workers (14, 47). Despite substantial individual and institutional investments, the collective and long-term benefits for big data animal health monitoring at the population, regional, or national level may be murky for the individual practitioner. Thus, sustainability may depend on creating short-term value for participating entities. For companion animal and equine medicine, aggregated health data could be used to research and subsequently deliver "precision" veterinary care that is tailored to the individual (5, 8). For livestock industries, short-term value may focus on research that intends to improve herd and flock management.

The advent of big data has implications for the education of veterinary epidemiologists (6, 13, 17, 51), including technical skills, such as computer programming, that may not be a traditional part of epidemiological training. While epidemiologists may never be responsible for creating complete software applications, the ability to manage relational databases or write simple scripts in a programming language to facilitate preparing data for analysis is critical when datasets become too large to process manually. Further, the analysis of big data often entails the use of supercomputing resources, which usually requires some familiarity with parallel processing and IT systems. To train the current workforce, workshops with hands-on computational activities are needed. Current curricula in graduate education should be expanded to include machine learning as well as traditional statistics, and coding as well as core epidemiological skills. Alternatively, graduate programs in the veterinary sciences could actively recruit students with computer science backgrounds that will readily be able to apply big data thinking to veterinary data. Veterinary epidemiologists with skillsets that allow them to directly engage with, manipulate, and analyze large datasets will be ideally situated to propel veterinary epidemiological research and practice into the coming decades.

The role of big data in veterinary epidemiology, and veterinary medicine more generally, has in some ways been inevitable from the beginning of the digital age, where data have become ever easier and cheaper to generate and store. At this point in time, we are at a turning point in terms of our ability to translate big data, which has existed for well over a decade, into smart data that create meaningful insights for animal health. Forward thinking is required to position our IT systems and workforce to harness the potential of big data. Indeed, from our perspective, big data should not be described as something that exists, but rather a capability. The real promise of big data is to create value out of disparate, chaotic pieces and extract real-time insights from data streams, thus creating a potentially revolutionary opportunity for veterinary epidemiology.

AUTHOR CONTRIBUTIONS

KV wrote the perspective. RM, CN, CV, and AP contributed ideas and reviewed the manuscript. CV contributed figures related with the Swine Health Monitoring Project.

FUNDING

This work was funded in part by the University of Minnesota MnDrive program and the Swine Health Information Center.

REFERENCES

1. Enki DG, Noufaily A, Garthwaite PH, Andrews NJ, Charlett A, Lane C, et al. Automated biosurveillance data from England and Wales, 1991–2011. *Emerg Infect Dis* (2013) 19:35–42. doi:10.3201/eid1901.120493
2. Martin LJ, Lee BE, Yasui Y. Google flu trends in Canada: a comparison of digital disease surveillance data with physician consultations and respiratory virus surveillance data, 2010–2014. *Epidemiol Infect* (2016) 144:325–32. doi:10.1017/S0950268815001478
3. Moss R, Zarebski A, Dawson P, McCaw JM. Forecasting influenza outbreak dynamics in Melbourne from Internet search query surveillance data. *Influenza Other Respir Viruses* (2016) 10:314–23. doi:10.1111/irv.12376
4. Martin LJ, Xu BY, Yasui Y. Improving Google flu trends estimates for the United States through transformation. *PLoS One* (2014) 9:e109209. doi:10.1371/journal.pone.0109209
5. Khoury MJ, Ioannidis JPA. Big data meets public health. *Science* (2014) 346:1054–5. doi:10.1126/science.aaa2709
6. Salathé M, Bengtsson L, Bodnar TJ, Brewer DD, Brownstein JS, Buckee C, et al. Digital epidemiology. *PLoS Comp Biol* (2012) 8:e1002616. doi:10.1371/journal.pcbi.1002616
7. Chan EH, Sahai V, Conrad C, Brownstein JS. Using web search query data to monitor dengue epidemics: a new model for neglected tropical disease surveillance. *PLoS Negl Trop Dis* (2011) 5:e1206. doi:10.1371/journal.pntd.0001206

8. Schneeweiss S. Learning from big health care data. *N Engl J Med* (2014) 370:2161–3. doi:10.1056/NEJMp1401111
9. Escobar LE, Craft ME. Advances and limitations of disease biogeography using ecological niche modeling. *Front Microbiol* (2016) 7:1174. doi:10.3389/fmicb.2016.01174
10. Kao RR, Haydon DT, Lycett SJ, Murcia PR. Supersize me: how whole-genome sequencing and big data are transforming epidemiology. *Trends Microbiol* (2014) 22:282–91. doi:10.1016/j.tim.2014.02.011
11. Hoffmann AR, Proctor LM, Surette MG, Suchodolski JS. The microbiome: the trillions of microorganisms that maintain health and cause disease in humans and companion animals. *Vet Pathol* (2016) 53:10–21. doi:10.1177/0300985815595517
12. van Borm S, Belak S, Freimanis G, Fusaro A, Granberg F, Hoper D, et al. Next-generation sequencing in veterinary medicine: how can the massive amount of information arising from high-throughput technologies improve diagnosis, control, and management of infectious diseases? In: Cunha MV, Inacio J, editors. *Veterinary Infection Biology: Molecular Diagnostics and High-Throughput strategies*. New York: Springer (2015). p. 425–36.
13. Pfeiffer DU, Stevens KB. Spatial and temporal epidemiological analysis in the big data era. *Prev Vet Med* (2015) 133:213–20. doi:10.1016/j.prevetmed.2015.05.012
14. Gates MC, Holmstrom LK, Biggers KE, Beckham TR. Integrating novel data streams to support biosurveillance in commercial livestock production systems in developed countries: challenges and opportunities. *Front Public Health* (2015) 3:74. doi:10.3389/fpubh.2015.00074
15. Dorea FC, Vial F. Animal health syndromic surveillance: a systematic literature review of the progress in the last 5 years (2011–2016). *Dovepress* (2016) 7:157–70. doi:10.2147/VMRR.S90182
16. McAfee A, Brynjolfsson E. Big data: the management revolution. *Harv Bus Rev* (2012) 90:61–8.
17. Mooney SJ, Westreich DJ, El-Sayed AM. Epidemiology in the era of big data. *Epidemiology* (2015) 26:390–4. doi:10.1097/EDE.0000000000000274
18. Andreu-Perez J, Poon CCY, Merrifield RD, Wong STC, Yang G-Z. Big data for health. *IEEE J Biomed Health Inform* (2015) 19:1193–208. doi:10.1109/JBHI.2015.2450362
19. Kass PH, Weng HY, Gaona MA, Hille A, Sydow MH, Lund EM, et al. Syndromic surveillance in companion animals utilizing electronic medical records data: development and proof of concept. *PeerJ* (2016) 4:e1940. doi:10.7717/peerj.1940
20. Kiss IZ, Green DM, Kao RR. The network of sheep movements within Great Britain: network properties and their implications for infectious disease spread. *J R Soc Interface* (2006) 3:669–77. doi:10.1098/rsif.2006.0129
21. Natale F, Giovannini A, Savini L, Palma D, Possenti L, Fiore G, et al. Network analysis of Italian cattle trade patterns and evaluation of risks for potential disease spread. *Prev Vet Med* (2009) 92:341–50. doi:10.1016/j.prevetmed.2009.08.026
22. Stark KDC, Regula G, Hernandez J, Knopf L, Fuchs K, Morris RS, et al. Concepts for risk-based surveillance in the field of veterinary medicine and veterinary public health: review of current approaches. *BMC Health Serv Res* (2006) 6:20. doi:10.1186/1472-6963-6-20
23. VanderWaal KL, Picasso C, Enns EA, Craft ME, Alvarez J, Fernandez F, et al. Network analysis of cattle movements in Uruguay: quantifying heterogeneity for risk-based disease surveillance and control. *Prev Vet Med* (2016) 123:12–22. doi:10.1016/j.prevetmed.2015.12.003
24. Caporale V, Giovannini A, Di Francesco C, Calistri P. Importance of the traceability of animals and animal products in epidemiology. *Rev Sci Tech* (2001) 20:372–7. doi:10.20506/rst.20.2.1279
25. Bajardi P, Barrat A, Savini L, Colizza V. Optimizing surveillance for livestock disease spreading through animal movements. *J R Soc Interface* (2012) 9:2814–25. doi:10.1098/rsif.2012.0289
26. Lal Dutta B, Ezanno P, Vergu E. Characteristics of the spatio-temporal network of cattle movements in France over a 5-year period. *Prev Vet Med* (2014) 117:79–94. doi:10.1016/j.prevetmed.2014.09.005
27. Hardstaff JL, Hasler B, Rushton JR. Livestock trade networks for guiding animal health surveillance. *BMC Vet Res* (2015) 11:82. doi:10.1186/s12917-015-0354-4
28. Green DM, Kiss IZ, Mitchell AP, Kao RR. Estimates for local and movement-based transmission of bovine tuberculosis in British cattle. *Proc Biol Sci* (2008) 275:1001–5. doi:10.1098/rspb.2007.1601
29. Kao RR, Danon L, Green DM, Kiss IZ. Demographic structure and pathogen dynamics on the network of livestock movements in Great Britain. *Proc Biol Sci* (2006) 273:1999–2007. doi:10.1098/rspb.2006.3505
30. Marquetoux N, Heuer C, Wilson P, Ridler A, Stevenson M. Merging DNA typing and network analysis to assess the transmission of paratuberculosis between farms. *Prev Vet Med* (2016) 134:113–21. doi:10.1016/j.prevetmed.2016.09.014
31. Martínez-López B, Perez AM, Sánchez-Vizcaíno JM. Social network analysis. Review of general concepts and use in preventive veterinary medicine. *Transbound Emerg Dis* (2009) 56:109–20. doi:10.1111/j.1865-1682.2009.01073.x
32. Rautureau S, Dufour B, Durand B. Vulnerability of animal trade networks to the spread of infectious diseases: a methodological approach applied to evaluation and emergency control strategies in cattle, France, 2005. *Transbound Emerg Dis* (2011) 58:110–20. doi:10.1111/j.1865-1682.2010.01187.x
33. Gilbert M, Mitchell A, Bourn D, Mawdsley J, Clifton-Hadley R, Wint W. Cattle movements and bovine tuberculosis in Great Britain. *Nature* (2005) 435:491–6. doi:10.1038/nature03548
34. Lee J-G, Kang M. Geospatial big data: challenges and opportunities. *Big Data Res* (2015) 2:74–81. doi:10.1016/j.bdr.2015.01.003
35. Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A. Very high resolution interpolated climate surfaces for global land areas. *Int J Climatol* (2005) 25:1965–78. doi:10.1002/joc.1276
36. Broxton PD, Zeng XB, Sulla-Menashe D, Troch PA, Global Land A. Cover climatology using MODIS data. *J Appl Meteorol Clim* (2014) 53:1593–605. doi:10.1175/JAMC-D-13-0270.1
37. USGS. *National GAP Vertebrate Species Distribution Model*. U.G. Service (2011).
38. FAO. *Global Livestock Densities*. (2005). Available from: http://www.fao.org/geonetwork/
39. FAO. *Global Livestock Production Systems*. (2011). Available from: http://www.fao.org/ag/againfo/resources/en/glw/GLW_prod-sys.html
40. NASA. *MODIS: Moderate Resolution Imaging Spectroradiometer*. (2016). Available from: https://modis.gsfc.nasa.gov
41. Stevens KB, Pfeiffer DU. Sources of spatial animal and human health data: casting the net wide to deal more effectively with increasingly complex disease problems. *Spat Spatiotemporal Epidemiol* (2015) 13:15–29. doi:10.1016/j.sste.2015.04.003
42. FAO. *EMPRES Global Animal Disease Information System (EMPRES-i)*. (2015). Available from: http://www.empres-i-fao.org/
43. Hay SI, George DB, Moyes CL, Brownstein JS. Big data opportunities for global infectious disease surveillance. *PLoS Med* (2013) 10:e1001413. doi:10.1371/journal.pmed.1001413
44. Hampton SE, Strasser CA, Tewksbury JJ, Gram WK, Budden AE, Batcheller AL, et al. Big data and the future of ecology. *Front Ecol Environ* (2013) 11:156–62. doi:10.1890/120103
45. Alkhamis MA, VanderWaal KL. Spatial and temporal epidemiology of lumpy skin disease in the Middle East, 2012–2015. *Front Vet Sci* (2016) 3:19. doi:10.3389/fvets.2016.00019
46. Tourre YM, Lacaux JP, Vignolles C, Lafaye M. Remote sensing in a changing climate and environment: the rift valley fever case. *Int Arch Photogramm* (2012) 39-B8:99–102. doi:10.5194/isprsarchives-XXXIX-B8-99-2012
47. Toh S, Platt R. Is size the next big thing in epidemiology? *Epidemiology* (2013) 24:349–51. doi:10.1097/EDE.0b013e31828ac65e
48. Main R. Supporting VDL messaging and data analysis through data standardization. *Allen D. Leman Swine Conference*. St. Paul, MN (2016).
49. Dorea FC, Sanchez J, Revie CW. Veterinary syndromic surveillance: current initiatives and potential for development. *Prev Vet Med* (2011) 101:1–17. doi:10.1016/j.prevetmed.2011.05.004
50. Willette M, Ponder JB, McRuer DL, Clark EE. Wildlife health monitoring systems in North America: from sentinel species to public policy. In: Aguirre A, Ostfel RS, Daszak P, editors. *Conservation Medicine: Applied Cases of Ecological Health*. New York: Oxford University Press (2013). p. 552–62.
51. Asokan GV, Asokan V. Leveraging "big data" to enhance the effectiveness of "one health" in an era of health informatics. *J Epidemiol Glob Health* (2015) 5:311–4. doi:10.1016/j.jegh.2015.02.001

52. Dórea FC, Muckle CA, Kelton D, McClure JT, McEwen BJ, McNab WB, et al. Exploratory analysis of methods for automated classification of laboratory test orders into syndromic groups in veterinary medicine. *PLoS One* (2013) 8:e57334. doi:10.1371/journal.pone.0057334
53. Anholt RM, Berezowski J, Jamal I, Ribble C, Stephen C. Mining free-text medical records for companion animal enteric syndrome surveillance. *Prev Vet Med* (2014) 113:417–22. doi:10.1016/j.prevetmed.2014.01.017
54. Guernier V, Milinovich GJ, Santos MAB, Haworth M, Coleman G, Magalhaes RJS. Use of big data in the surveillance of veterinary diseases: early detection of tick paralysis in companion animals. *Parasite Vector* (2016) 9:303. doi:10.1186/s13071-016-1590-6
55. Signorini A, Segre AM, Polgreen PM. The use of Twitter to track levels of disease activity and public concern in the US during the influenza A H1N1 pandemic. *PLoS One* (2011) 6:e19467. doi:10.1371/journal.pone.0019467
56. Muellner UJ, Vial F, Wohlfender F, Hadorn D, Reist M, Muellner P. Timely reporting and interactive visualization of animal health and slaughterhouse surveillance data in Switzerland. *Front Vet Sci* (2015) 2:47. doi:10.3389/fvets.2015.00047
57. Bayot B, Sonnonholzner S, Ochoa X, Guerrerro J, Vera T, Calderon J, et al. An online operational alert system for the early detection of shrimp epidemics at the regional level based on real-time production. *Aquaculture* (2008) 277:164–73. doi:10.1016/j.aquaculture.2008.02.035
58. Zeng D. *Infectious Disease Informatics and Biosurveillance*. New York: Springer (2011).
59. Thomas-Bachli AL, Pearl DL, Friendship RM, Berke O. Exploring relationships between whole carcass condemnation abattoir data, non-disease factors and disease outbreaks in swine herds in Ontario (2001–2007). *BMC Res Notes* (2014) 7:185. doi:10.1186/1756-0500-7-185
60. Perez AM, Alba A, Goede D, McCluskey B, Morrison RB. Monitoring the spread of swine enteric coronavirus diseases in the United States in the absence of a regulatory framework. *Front Vet Sci* (2016) 3:18. doi:10.3389/fvets.2016.00018
61. Tousignant SJ, Perez AM, Lowe JF, Yeske PE, Morrison RB. Temporal and spatial dynamics of porcine reproductive and respiratory syndrome virus infection in the United States. *Am J Vet Res* (2014) 61:70–5. doi:10.2460/ajvr.76.1.70
62. Chiolero A. Big data in epidemiology: too big too fail? *Epidemiology* (2013) 24:938–9. doi:10.1097/EDE.0b013e31829e46dc
63. Machado G, Mendoza MR, Corbellini LG. What variables are important in predicting bovine viral diarrhea virus? A random forest approach. *Vet Res* (2015) 46:85. doi:10.1186/s13567-015-0219-7
64. Ortiz-Pelaez A, Pfeiffer DU. Use of data mining techniques to investigate disease risk classification as a proxy for compromised biosecurity of cattle herds in Wales. *BMC Vet Res* (2008) 4:24. doi:10.1186/1746-6148-4-24
65. Peters DPC, Havstad KM, Cushing J, Tweedie C, Fuentes O, Villanueva-Rosales N. Harnessing the power of big data: infusing the scientific method with machine learning to transform ecology. *Ecosphere* (2014) 5:67. doi:10.1890/ES13-00359.1
66. Lazer D, Kennedy R, King G, Vespignani A. The parable of Google flu: traps in big data analysis. *Science* (2014) 343:1203–5. doi:10.1126/science.1248506

Conflict of Interest Statement: The 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.