

'Big Data' in animal health research – opportunities and challenges

cambridge.org/ahr

Janet I. MacInnes 

Department of Pathobiology, Ontario Veterinary College, University of Guelph, Guelph, Canada

Opinion

Cite this article: MacInnes JI (2020). 'Big Data' in animal health research – opportunities and challenges. *Animal Health Research Reviews* **21**, 1–2. <https://doi.org/10.1017/S1466252319000215>

Received: 22 November 2019
Accepted: 22 November 2019
First published online: 20 July 2020

Key words:

Animal health; big data; machine learning; modeling; prediction

Author for correspondence:

Janet I. MacInnes, Department of Pathobiology, Ontario Veterinary College, University of Guelph, Guelph, Canada.
E-mail: macinnes@uoguelph.ca

Abstract

Automated systems for high-input data collection and data storage have led to exponential growth in the availability of information. Such datasets and the tools applied to them have been referred to as 'big data'. Starting with a systematic review of the terms 'informatics, bioinformatics and big data' in animal health this special issue of AHRR illustrates some big-data applications with papers on how the use of various omics methods may be used to facilitate the development of improved diagnostics, therapeutics, and vaccines for foodborne pathogens in poultry and on how a better understanding of rumen microbiota could lead to improved feed absorption while minimizing methane production. Other papers in this issue cover the use of big data modeling in dairy cattle for more effective disease interventions and machine learning tools for livestock breeding. The final two reviews describe the use of big data in better vector-borne pathogen forecasts with canine seroprevalence maps and modeling approaches to understand the transmission of avian influenza virus. Although a lot of technical and ethical issues remain with the use of big data, these reviews illustrate the tremendous potential that big-data systems have to revolutionize animal health research.

As the world's population grows and the planet is under increasing stress, there is mounting urgency to address animal health issues in order to ensure the continuation of essential food systems and life-sustaining ecological services. To do this, researchers are increasingly taking advantage of very large data sets and the tools that can be used to analyze them. Automated systems for high-input data collection are becoming commonplace, leading to exponential growth in the availability of information regarding a variety of host, pathogen, and environmental factors that affect animal health. For example, the capture of whole genome sequences and gene expression data from both hosts and pathogens is becoming routine, as is the keeping of electronic medical records and the use of digitized images. In addition, vast quantities of data are routinely being collected by sensors, drones, and satellites. The increasing availability of such electronic data has led to the creation of massive databases that are too large and complex to be managed using traditional data analysis tools. In these cases, specialized tools are required to collect, organize, and analyze the data. Collectively, these datasets and the tools applied to them have been referred to as 'big data.' This special issue of AHRR illustrates a number of applications of big data that are being used in animal health research.

In the first article, 'A scoping review of "big data," "informatics" and "bioinformatics" in the animal health and veterinary medical literature' by Ouyang *et al.* the authors define and describe the ways in which these terms are used in the animal health and veterinary literature (Ouyang *et al.*, 2019). The authors found a steady increase in the number of published 'bioinformatics' articles, and noted the number of 'informatics' articles increased up to 2012 and then declined. Since a seminal article focusing on 'big data' was published in 2012, the authors suggest that the rapidly evolving definition of big data may have led animal health researchers to avoid using the term.

The next two articles explore some aspects of the growing literature involving 'omics' – studies which permit a detailed molecular analysis of various biological molecules – for example at the gene level (genomics), the ribonucleic acid level (transcriptomics), the protein level (proteomics), or the metabolic level (metabolomics). The first of these papers, 'Translating big data: better understanding of host-pathogen interaction to control bacterial foodborne pathogens in poultry' by Deblais *et al.* (2019) discusses how various omics techniques might facilitate the development of improved diagnostics, therapeutics, and vaccines for foodborne pathogens (Deblais *et al.*, 2019). Focusing on *Campylobacter* spp., *Salmonella* spp., and *Escherichia coli* in poultry, the authors describe, among other things, the ways in which whole genome sequencing has revolutionized our ability to track these pathogens and their associated resistance genes and how the development of novel proteomic and metabolomic methods could allow for more rapid and accurate detection of a variety of foodborne pathogens in poultry.

In the report by Cholewinska *et al.* entitled 'The microbiome of the digestive system of ruminants – a review,' the authors describe the composition and function of the microbiota

© Cambridge University Press 2020

CAMBRIDGE
UNIVERSITY PRESS

of the rumen (Cholewinska *et al.*, 2019). They assert that an understanding of rumen microbiota is an essential step towards the eventual ability to manipulate the rumen microbiome to enable better feed absorption while minimizing methane production. Because of its important role in animal health, livestock production parameters and methane production, Cholewinska *et al.* advocate additional research to fully understand the complex interactions of rumen microbiota.

The next two papers describe modeling approaches for the prospective use of big data. The first of these papers, 'Prospects for predictive modelling of transition cow diseases' by Wisnieski *et al.* discusses how various predictive modeling methods and candidate variables can be used to predict transition cow diseases (Wisnieski *et al.*, 2019). The authors explain that traditional biomarker test results (health records, feed intake, body condition scores, and milk production) in dairy cattle are often collected and analyzed in isolation and suggest that using predictive models with relevant predictors could aid in the development of more effective disease reduction interventions.

The next paper in this issue, 'A review of traditional and machine learning methods applied to animal breeding' by Nayeri *et al.* discusses both new and traditional prediction methods that are used in the livestock breeding field and provide insight into the potential for machine learning techniques to surpass traditional breeding prediction methods (Nayeri *et al.*, 2019).

The final two papers in this issue consider the use of big data in understanding the spread of important human and animal pathogens such as heartworm, Lyme spirochetes, and avian influenza (AI) virus. In their article, 'Canine vector-borne disease: mapping and the accuracy of forecasting using big data from the veterinary community,' Watson Self *et al.* discuss the reliability of vector-borne disease forecasts and illustrate the ways in which pathogen prevalence maps and forecast data from the Companion Animal Parasite Council can be used by scientists and practitioners for evidence-based decision making and for client education (Watson Self *et al.*, 2019).

In the final paper by Yousefi Naghani *et al.* 'A review of knowledge discovery process in control and mitigation of avian influenza,' the authors remind us of the many threats that AI virus poses to the poultry industry and to public health, and provide a critical review of modeling approaches that draw on diverse data sources to better understand AI virus transmission (Yousefi Naghani *et al.*, 2019). From this far-reaching review of approaches that have been used to model AI infections, in the context of Knowledge Discovery in Data, the authors identify gaps in the literature and make a number of recommendations.

Studies using big data have tremendous promise to revolutionize the ways in which we address crucial issues in animal health – ranging from genome manipulations to disease diagnosis. However, it will be important to establish common standards for data handling to optimize sharing and analysis of data. Moreover, the same level of critical appraisal that is demanded by any research method or approach must be practiced – applications of big data require sound judgment if they are to be turned into 'smart data' (i.e. smaller sets of valuable and actionable information). Animal health researchers need to remember that the automated collection and analysis of data is only as unbiased as

the underlying assumptions and financial limitations associated with their design. As the use and importance of big data grows, it may become increasingly necessary that veterinarians and other animal health researchers receive training in the area of big data – at least to the level where they can work effectively in teams with other researchers who have backgrounds in computing and artificial intelligence.

We also need to be cognizant of the ethical issues that may be associated with the use of big data in the animal health field. For example, decisions regarding who has the rights to study, to distribute or to financially gain from such data may require value-laden judgments, as might decisions regarding the security and storage of the datasets. These ethical considerations become increasingly complex as various jurisdictions develop and implement different regulations. Moreover, when collecting data, there is often tension between common good versus individual good – careful consideration needs to be given to balance the rights of both individuals and society. The growing use of complex networks and the development of artificial intelligence add further levels of complexity to these ethical issues. Additionally, there are environmental costs associated with the use and maintenance of huge data sets that should be considered when weighing the benefits of big data approaches, though it is possible that future technological developments may mitigate some of these environmental challenges.

In conclusion, big data systems will likely play a critically important role in our future understanding of animal health and disease, but we need to remember, at all times, they should be 'handled with care.'

References

- Cholewinska P, Czyz K, Nowakowski P and Wrostek A (2019) The microbiome of the digestive system of ruminants – a review. *Animal Health Research Reviews* 20, 42–53.
- Deblais L, Kathayat D, Helmy YA, Closs Jr G and Rajashekara G (2019) Translating 'big data': better understanding of host-pathogen interactions to control bacterial foodborne pathogens in poultry. *Animal Health Research Reviews* 20, 21–41.
- Nayeri S, Sargolzaei M and Tulpan D (2019) A review of traditional and machine learning methods applied to animal breeding. *Animal Health Research Reviews* 20, 66–81.
- Ouyang ZB, Sargeant J, Thomas A, Wycherley K, Ma R, Esmalbeigi R, Versluis A, Stacey D, Stone E, Poljak Z and Bernardo TM (2019) A scoping review of 'big data', 'informatics', and 'bioinformatics' in the animal health and veterinary medical literature. *Animal Health Research Reviews* 20, 3–20.
- Watson Self SC, Liu Y, Nordone SK, Yabsley MJ, Stockdale Walden H, Lund RB, Bowman DD, Carpenter C, McMahan CS and Gettings JR (2019) Canine vector-borne disease: mapping and the accuracy of forecasting using big data from the veterinary community. *Animal Health Research Reviews* 20, 82–95.
- Wisnieski L, Norby B, Pierce SJ, Becker T and Sordillo LM (2019) Prospects for predictive modeling of transition cow diseases. *Animal Health Research Reviews* 20, 54–65.
- Yousefi Naghani S, Pljak Z, Sharif S and Dara R (2019) A review of knowledge discovery process in control and mitigation of avian influenza. *Animal Health Research Reviews* 20, 96–106.