Introduction

Data and the need to quantify: a personal perspective

S.W.J. Reid

World Organisation for Animal Health Collaborating Centre for Risk Analysis and Modelling, c/o Royal Veterinary College, Hawkshead Lane, North Mymms, Hatfield, Hertfordshire, AL9 7TA, United Kingdom

E-mail: principal@rvc.ac.uk

Keywords

Big data – Disease control – Evidence-based medicine – Modelling – Quantification.

> 'I often say that when you can measure what you are speaking about, and express it in numbers, you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind.' Lord Kelvin (1824–1907)

There is an argument to be made that the only truly unifying language of our modern world lies in the ones and zeros of our pervasive digital universe, and it is in this ecosystem that the concept of big data resides. Yet remaining central to the utility we extract from the data we now collect by the terabyte is our ability to analyse, assess and prioritise – and importantly, the need to quantify.

It has been nearly 30 years since Hiatt and Goldman [1], both professors of medicine at Harvard Medical School, published a commentary in the journal *Nature* titled 'Making medicine more scientific'. The article was set against the broadly based view that, with increasing numbers of physicians having basic biology training as well as medical training, the era of scientific medicine had arrived. Hiatt and Goldman challenged this view, arguing that whilst some basic biological and molecular

sciences had been adopted, decision-making in medicine was still based on subjective grounds, and the evaluative clinical sciences – such as statistics, epidemiology, decision analysis, economics, cost effectiveness analysis and data sciences – had essentially been ignored. The consequences of this omission, they argued, were two-fold: first, population-based studies and research into diagnostic and therapeutic outcomes or impacts were essentially peripheral to medical academic research; and second, quantitative, efficacy and ethical considerations related to the individual patient were still neglected, if emerging, concepts.

The following year, in the same journal, an article by Wain-Hobson [2] on advances in human immunodeficiency virus research remarked that the reported findings were the result of 'biologists, in this case immunologists, teaming up with mathematicians'. The next week, *Nature*'s editorial built on this observation but commented that molecular biologists and virologists were specifically cited as lacking in quantitative regard and hinted at a much wider need.

It is perhaps difficult to believe that at the end of the 20th century, one of the world's leading scientific journals was alerting the community to the fact that both basic medical research and clinical medicine were quantitatively wanting. In terms of disease control, a field in which one might consider three levels of investigation, *viz.* the molecular, the patient and the population, two were found to be deficient with regard to Kelvin's aforementioned assessment.

At the scale of the population, in public health and epidemiology, numeracy has historically been strong, at least descriptively, and the concepts of modern epidemiology have a long and distinguished record. Whilst identifying risk factors for disease occurrence and defining statistical associations between exposures and disease were practices widely applied in the medical domain, mathematical modelling methods, in use for many years, were largely the preserve of the non-clinical sector, particularly in micro and macro parasite biology, ecology and evolutionary biology. With rarely acknowledged roots in the work of Bailey [3], it was the contribution of a zoologist and a

physicist, Anderson and May [4], that was arguably responsible for bringing modelling approaches in infectious diseases to the wider medical and veterinary world.

From another school, but equally distinguished, the modern epidemiologist would undoubtedly cite Popper [5] and Rothman *et al.* [6], amongst others, as key players in shaping the way in which quantitative approaches evolved and found application. In this school, causal inference, a statistical approach and a focus on bias, confounding and risk have laid the foundations for key investigations in both medical and veterinary domains.

At the other end of the spectrum – the impact of data on evidence impacting the individual – the aforementioned *Nature* commentary was overshadowed by an article written by Guyatt *et al.* some two years earlier [7]. The need for a formal approach to the assessment of best current practice led to the establishment of modern evidence-based medicine (EBM), albeit acknowledging that the underlying principles are as old as medicine itself. But what is the evidence? How common are the diseases we treat, how effective are our diagnostic techniques, how good are our therapeutics, how successful are our interventions? These questions sit comfortably in an epidemiological text and it is clear that the practice of EBM relies on the theories and principles of epidemiology and biostatistics – and, above all, the need to quantify is again pivotal. Sackett *et al.* [8] proposed a hierarchy of evidence, subsequently modified by Yusuf *et al.* [9], that is required to address a four-step approach [10] to EBM, *viz.*:

1) systematic reviews of multiple, randomised, blinded, placebo-controlled trials designed to address specific clinical questions

2) non-randomised clinical trials using historical controls

3) uncontrolled case series

4) expert opinion and/or extrapolated evidence from published research.

Although this hierarchy is not ubiquitously accepted, importantly, data in abundance and quantitative techniques are central to each evidence type. The veterinary world has followed suit [11].

Advancing the discussion with a range of perspectives

Twenty years later, we have data in abundance and a proliferation of big data applications in medicine and veterinary medicine [12, 13, 14, 15], where the scope of investigation embraces everything from meteorological and spatial data to that generated by whole genome sequencing and bioinformaticians. The collection of studies and reviews in this issue of the World Organisation for Animal Health (WOAH)'s *Scientific and Technical Review*, spawned by the creation of the WOAH Collaborating Centre for Risk Analysis and Modelling (a joint venture between the Royal Veterinary College and the Animal and Plant Health Agency in the United Kingdom), illustrates the when, why and how of the use of data and the application of quantitative techniques. The studies range from the macro (population) to the microbiological and molecular and, at each resolution, issues relating to data and the application of a quantitative approach are pivotal to sound scientific discourse and inference.

The issue addresses several important features ranging from governance considerations to species-specific issues through to examples from several domains and, importantly, the relevance and leadership of WOAH in key aspects of data in surveillance, disease control and policy. With studies from the food animal sector and from wildlife species and companion animal programmes, spanning from whole genome sequencing to geographic information systems, and with statistical, mathematical, information and data sciences, there will be lessons to be learned from comparing and contrasting these many contributions.

The real beauty of the approach in this issue is that it provides the essential ingredient for an investigation to become a coherent story: for all the studies published here, if the scientific motivation provides the plot, and biological observations are nouns and verbs, then the

quantitative elements are the adverbs, adjectives, pronouns and nuances of grammar that lead to accomplished and comprehensible prose.

The studies are all examples of the pivotal place for data and quantification in animal health, and, while there are many less than complimentary 'quantitative' maxims that surface in the literature, ranging from the observation that 'all models are wrong, but some are useful' [16], through 'errors associated with inadequate data are smaller than errors associated with no data' (Babbage, 1792–1871), to 'lies, damned lies and statistics' (attributed to Mark Twain and Benjamin Disraeli), the reality is that whilst it is possible to make false or flawed inferences from data using quantitative techniques [17], it is much easier to draw inappropriate conclusions when such measures are not applied or when data are absent.

The challenge ahead

The next challenge for the sector will be in addressing the synthesis of data in a truly One Health context – how we, working in animal health, can ensure that our data can be efficiently combined with data from other sectors and organisations, such as the Food and Agriculture Organization of the United Nations and the World Health Organization [18], where the motivation for survey and collection may be different but where value and synergy lie in the intersectionality of datasets. One example is in addressing antimicrobial resistance [19] and the necessity to exploit multiple sources for real progress and to ensure we move forward with a data-focused, evidenced-based approach to decision-making. But that will be for another issue of WOAH's *Review*.

Having made the case for data and quantification, there is an important footnote as, while they are necessary, they are not alone sufficient; a reductionist approach is equally abhorrent in conveying a complex scientific message. Words matter, too [20], as one hopes this collection of papers illustrates.

References

[1] Hiatt H. & Goldman L. (1994). – Making medicine more scientific. *Nature*, **371** (6493), 100. https://doi.org/10.1038/371100a0

[2] Wain-Hobson S. (1995). – Virological mayhem. *Nature*, **373** (6510), 102. https://doi.org/10.1038/373102a0

[3] Bailey N.T.J. (1975). – The mathematical theory of infectious diseases and its application. 2nd Ed. Griffin, London, United Kingdom, 413 pp.

[4] Anderson R.M. & May R.M. (1992). – Infectious diseases of humans: dynamics and control. Oxford University Press, Oxford, United Kingdom, 766 pp.

[5] Popper K.R. (1935). – Logik der Forschung. Zur Erkenntnistheorie der modernen Naturwissenshcaft. English translation: The logic of scientific discovery. Routledge Classics, London, United Kingdom, 1959, 513 pp. Available at: http://philotextes.info/spip/IMG/pdf/popper-logic-scientific-discovery.pdf (accessed on 13 March 2023).

[6] Rothman K.J., Greenland S. & Lash T.L. (2008). – Modern epidemiology. 3rd Ed. Lippincott Williams and Wilkins, Philadelphia, United States of America, 758 pp. Available at: http://students.aiu.edu/submissions/profiles/resources/onlineBook/a9c7D5_Modern_Epidemiology_3.pdf (accessed on 13 March 2023).

[7] Guyatt G.H. & Evidence-Based Medicine Working Group (1992). – Evidence-based medicine: a new approach to teaching the practice of medicine. *JAMA*, **268** (17), 2420–2425. https://doi.org/10.1001/jama.1992.03490170092032

[8] Sackett D.L., Haynes R.B., Guyatt G.H. & Tugwell P. (1991). – Clinical epidemiology: a basic science for clinical medicine. 2nd Ed. Lippincott Williams and Wilkins, Philadelphia, United States of America, 466 pp.

[9] Yusuf S., Cairns J.A., Camm A.J., Fallen E.L. & Gersh B.J. (2008). – Evidence-based cardiology. 2nd Ed. BMJ Publishing Group, London, United Kingdom, 1024 pp.

[10] Sackett D.L., Straus S.E., Richardson W.S., Rosenberg W. & Haynes R.B. (2000). – Evidence-based medicine: how to practice and teach EBM. 2nd Ed. Churchill Livingstone, Toronto, Canada, 280 pp.

[11] Cockcroft P.D. & Holmes M.A. (2003). – Handbook of evidence-based veterinary medicine. Blackwell Publishing, Oxford, United Kingdom, 210 pp. https://doi.org/10.1002/9780470690833

[12] Ristevski B. & Chen M. (2018). – Big data analytics in medicine and healthcare. *J. Integr. Bioinform.*, **15** (3), 20170030. https://doi.org/10.1515/jib-2017-0030

[13] Kao R.R., Haydon D.T., Lycett S.J. & Murcia P.R. (2014). – Supersize me: how whole-genome sequencing and big data are transforming epidemiology. *Trends Microbiol.*, **22** (5), 282–291. https://doi.org/10.1016/j.tim.2014.02.011

[14] Ouyang Z., Sargeant J., Thomas A., Wycherley K., Ma R., Esmaeilbeigi R., Versluis A., Stacey D., Stone E., Poljak Z. & Bernardo T.M. (2019). – A scoping review of 'big data', 'informatics', and 'bioinformatics' in the animal health and veterinary medical literature. *Anim. Health Res. Rev.*, **20** (1), 1–18. https://doi.org/10.1017/S1466252319000136

[15] Gulyaeva M., Huettmann F., Shestopalov A., Okamatsu M., Matsuno K., Chu D.-H., Sakoda Y., Glushchenko A., Milton E. & Bortz E. (2020). – Data mining and model-predicting a global disease reservoir for low-pathogenic Avian Influenza (AI) in the wider pacific rim using big data sets. *Sci. Rep.*, **10** (1), 16817. https://doi.org/10.1038/s41598-020-73664-2

[16] Box G.E.P. (1979). – Robustness in the strategy of scientific model building. *In* Robustness in statistics (R.L. Launer & G.N. Wilkinson, eds). Academic Press, New York, United States of America, 201–236. https://doi.org/10.1016/B978-0-12-438150-6.50018-2

[17] Ioannidis J.P.A. (2005). – Why most published research findings are false. *PLoS Med.*, **2** (8), e124. https://doi.org/10.1371/journal.pmed.0020124

[18] Food and Agriculture Organization of the United Nations (FAO), World Organisation for Animal Health (WOAH) & World Health Organization (WHO) (2017). – The Tripartite's commitment: providing multi-sectoral, collaborative leadership in addressing health challenges. FAO, WOAH & WHO, Rome, Italy, 4 pp. Available at: https://doc.woah.org/dyn/portal/index.xhtml?page=alo&aloId=34824&espaceId=100 (accessed on 13 March 2023).

[19] World Health Organization (WHO), Food and Agriculture Organization of the United Nations (FAO), World Organisation for Animal Health (WOAH) & United Nations Environment Programme (UNEP) (2022). – Strategic framework for collaboration on antimicrobial resistance – together for One Health. WHO, FAO, WOAH & UNEP, Geneva, Switzerland, 15 pp. Available at: https://www.who.int/publications/i/item/9789240045408 (accessed on 13 March 2023).

[20] Boyle D. (2002). – The tyranny of numbers: why counting can't make us happy. Harper Perennial GB, London, United Kingdom, 256 pp.