

## Introduction

### Les données et l'indispensable quantification : un point de vue personnel

S.W.J. Reid

Centre collaborateur de l'Organisation mondiale de la santé animale pour l'analyse du risque et la modélisation, c/o Royal Veterinary College, Hawkshead Lane, North Mymms, Hatfield, Hertfordshire, AL9 7TA, Royaume-Uni

Courriel : [principal@rvc.ac.uk](mailto:principal@rvc.ac.uk)

#### Mots-clés

Lutte contre les maladies animales – Médecine fondée sur le niveau de preuve – Mégadonnées – Modélisation – Quantification.

« Je dis souvent que si vous pouvez mesurer ce dont vous parlez et l'exprimer par un nombre, vous savez quelque chose de votre sujet ; mais si vous ne pouvez pas le mesurer, si vous ne pouvez pas l'exprimer en nombres, vos connaissances sont d'une pauvre espèce et bien peu satisfaisantes. » Lord Kelvin (1824–1907)

On peut faire valoir que le seul langage véritablement unificateur de notre monde moderne est celui des *un* et des *zéro* de l'univers numérique, aujourd'hui omniprésent ; c'est dans cet écosystème que règne le concept des « big data » (ou « mégadonnées »). Toutefois, si l'on veut extraire des éléments exploitables à partir de ces données (collectées aujourd'hui par téraoctets), il reste indispensable de mobiliser des capacités d'analyse, d'évaluation et d'établissement des priorités – de même qu'il est nécessaire et essentiel de quantifier.

Cela fait près de 30 ans que Hiatt et Goldman [1], tous deux professeurs de médecine à la faculté de Harvard, ont publié dans *Nature* un commentaire intitulé « *Making medicine more scientific* » [Pour une médecine plus scientifique]. L'article réfutait l'opinion largement répandue d'une médecine qui serait enfin entrée dans l'ère scientifique

grâce au nombre croissant de médecins ayant reçu une formation de base en biologie en plus de leur formation médicale. Hiatt et Goldman ne partageaient pas cette opinion car, d'après eux, si les médecins avaient effectivement intégré certaines connaissances en biologie, notamment moléculaire, la prise de décision médicale reposait toujours sur une base subjective tandis que les sciences cliniques fondées sur les méthodes d'évaluation (telles que la statistique, l'épidémiologie, l'analyse des décisions, l'économie, l'analyse coûts-avantages et les sciences de la donnée) étaient pour l'essentiel ignorées. Ils y voyaient deux conséquences : d'une part, les études fondées sur les populations et les travaux sur les retombées ou l'impact des diagnostics et des traitements étaient menés en marge de la recherche médicale universitaire ; d'autre part, les considérations quantitatives, d'efficacité, et d'éthique vis-à-vis des patients individuels, étaient toujours largement négligées, sinon inexistantes.

L'année suivante, Wain-Hobson [2] publiait dans le même journal un article sur les avancées de la recherche sur le virus de l'immunodéficience humaine ; il y faisait observer que les résultats obtenus étaient le fruit de « biologistes, en l'occurrence d'immunologistes travaillant en équipe avec des mathématiciens ». Huit jours plus tard, l'éditorial de *Nature* reprenait cette observation, mais faisait aussi remarquer que si les biologistes moléculaires et les virologues étaient spécifiquement cités comme étant confrontés à un manque criant de données quantitatives, en réalité les besoins en la matière allaient bien au-delà de ces domaines de recherche.

Il est difficile de croire qu'à la fin du XX<sup>e</sup> siècle, l'une des revues scientifiques les plus prestigieuses du monde alertait la communauté sur le déficit de quantification dans la recherche médicale fondamentale ainsi que dans la médecine clinique. S'agissant de la lutte contre les maladies, un tel déficit de la capacité de quantifier (pour reprendre l'appréciation de Kelvin) a été constaté dans deux des trois niveaux d'investigation qu'on peut reconnaître dans ce domaine, à savoir le niveau moléculaire, le niveau du patient et le niveau de la population.

À l'échelle d'une population, la santé publique et l'épidémiologie sont des disciplines qui depuis leur origine reposent fortement sur les capacités de calcul, du moins à des fins descriptives, et les concepts de l'épidémiologie moderne se sont ainsi imposés avec brio. Si la définition des facteurs de risque d'apparition des maladies et la recherche d'associations statistiques entre exposition et maladie sont des pratiques bien établies dans le domaine médical, en revanche les méthodes de modélisation mathématique, bien qu'appliquées depuis de nombreuses années, sont surtout l'apanage du secteur non clinique, en particulier la biologie des micro-parasites et des macro-parasites, l'écologie, et la biologie de l'évolution. C'est à Anderson et May, respectivement zoologiste et physicien et eux-mêmes dans une filiation rarement perçue avec les travaux de Bailey [3], que nous devons l'introduction des méthodes de modélisation des maladies infectieuses dans le monde plus large de la médecine humaine et vétérinaire [4].

Dans une tout autre école de pensée, mais toute aussi brillante, un épidémiologiste moderne citerait sans doute Popper [5], ainsi que Rothman *et al.* [6] parmi les auteurs ayant contribué de manière décisive à façonner de nouvelles approches quantitatives et à en rechercher les applications concrètes. Dans cette école, c'est l'inférence causale, l'approche statistique et la prise en compte des biais ainsi que des facteurs de confusion et de risques qui ont jeté les bases de travaux essentiels dans les domaines de la recherche médicale et vétérinaire.

À l'autre extrémité du spectre – c'est-à-dire là où les données ont un impact sur le niveau de preuve susceptible d'avoir des conséquences pour un individu – le commentaire de *Nature* mentionné ci-dessus a été éclipsé par l'article publié deux ans auparavant par Guyatt *et al.* [7]. La médecine moderne fondée sur le niveau de preuve (en anglais, *evidence-based medicine*, EBM), est née de la nécessité de formaliser les méthodes d'évaluation des meilleures pratiques actuelles, tout en reconnaissant que les principes sous-jacents de ces méthodes sont aussi anciens que la médecine elle-même. Mais qu'est-ce qu'une preuve ? Quelle est vraiment la fréquence des maladies que nous traitons ? Nos techniques de diagnostic sont-elles vraiment efficaces ? Nos thérapeutiques sont-elles de qualité ? Et quel est le succès de nos

interventions ? Ces questions ont toute leur place dans un texte épidémiologique et il est clair que la pratique de l'EBM repose sur les théories et les principes de l'épidémiologie et de la biostatistique – d'où, ici encore, la nécessité prépondérante d'une quantification. Sackett *et al.* [8] ont proposé une hiérarchie des niveaux de preuve, modifiée ensuite par Yusuf *et al.* [9], qui recouvre en quatre étapes la méthode de l'EBM [10], à savoir :

- 1) examens systématiques d'essais multiples, randomisés, en aveugle et contrôlés par placebo, conçus pour répondre à des questions cliniques spécifiques
- 2) essais cliniques non randomisés, avec des contrôles historiques
- 3) séries de cas non contrôlées
- 4) opinion d'expert et/ou éléments probants extrapolés à partir de travaux de recherche publiés.

Même si cette hiérarchie n'est pas universellement acceptée, ce qui importe c'est que l'abondance des données et l'existence de techniques quantitatives y soient déterminantes, et ce dans chacune des catégories de preuve. Le monde vétérinaire a suivi sur cette lancée [11].

### **Continuer le débat dans plusieurs perspectives**

Vingt années ont passé et nous disposons désormais d'une abondance de données tandis que les applications des mégadonnées se multiplient tant en médecine humaine qu'en médecine vétérinaire [12, 13, 14, 15] ; quant au champ d'investigation, il recouvre tout, des données météorologiques et spatiales à celles générées par le séquençage du génome entier et la bio-informatique. Les études et analyses rassemblées dans ce numéro de la *Revue scientifique et technique* de l'Organisation mondiale de la santé animale (OMSA), suivant un projet germé lors de la création du Centre collaborateur de l'OMSA pour l'analyse des risques et la modélisation (un consortium constitué par le Royal Veterinary College et l'Agence de la santé animale et végétale du Royaume-Uni), sont autant d'exemples montrant quand, pourquoi et comment utiliser les données et appliquer les méthodes quantitatives.

Les études vont du niveau macro (population) au niveau microbiologique et moléculaire ; à chaque résolution, l'examen de la question des données et l'approche quantitative adoptée sont ce qui donne au discours et aux déductions scientifiques leur solidité.

Ce numéro aborde plusieurs aspects importants, depuis les considérations de gouvernance, les questions spécifiquement liées à des espèces particulières, les exemples tirés de domaines différents, jusqu'à la pertinence et le leadership de l'OMSA concernant les aspects essentiels des données utilisées à des fins de surveillance, de lutte contre les maladies et d'élaboration des politiques. Les études réunies ici émanent du secteur des productions animales destinées à l'alimentation humaine ou des programmes sur la faune sauvage ou les animaux de compagnie, et couvrent des thèmes allant du séquençage du génome entier aux systèmes d'information géographique, en faisant appel aux sciences statistiques, mathématiques, de l'information et de la donnée ; pour toutes ces raisons, il y aura des enseignements à tirer de la comparaison, voire la confrontation, entre ces nombreuses contributions.

La véritable beauté de l'approche suivie dans ce numéro réside dans le fait qu'elle contient l'ingrédient indispensable pour qu'une recherche devienne un récit cohérent : dans toutes les études publiées ici, si la motivation scientifique fournit l'intrigue et que les observations biologiques constituent les substantifs et les verbes, alors les éléments quantitatifs sont les adverbess, les adjectifs, les pronoms et les nuances grammaticales qui ensemble configurent un énoncé accompli et compréhensible.

Chacune de ces études illustre la place centrale des données et de la quantification en santé animale et, malgré les nombreuses maximes « quantitatives » bien peu élogieuses faisant surface dans la littérature, depuis l'observation selon laquelle « tous les modèles sont faux, mais certains sont utiles » [16] ou l'affirmation que « les erreurs liées à des données inadéquates sont moindres que celles liées à l'absence de données » (Babbage, 1792–1871), jusqu'à la citation sur « les trois types de mensonges : les mensonges, les maudits mensonges et les

statistiques » (attribuée à Mark Twain et à Benjamin Disraeli), le fait est que s'il est effectivement possible de faire des déductions fausses ou erronées à partir des données en appliquant des techniques quantitatives [17], il est beaucoup plus facile de tirer des conclusions inappropriées lorsqu'on n'applique pas ces techniques ni ne recourt aux données.

## Le défi devant nous

Le prochain défi pour le secteur sera de procéder à la synthèse des données dans un contexte qui soit véritablement « Une seule santé » : ceux d'entre nous qui travaillons dans le domaine de la santé animale, comment procéderons-nous pour combiner efficacement nos données avec celles d'autres secteurs et organisations, en particulier l'Organisation des Nations unies pour l'alimentation et l'agriculture et l'Organisation mondiale de la santé [18], dont les motivations pour mener leurs enquêtes et collecter des données sont probablement différentes des nôtres, mais avec lesquelles le recoupement des jeux de données constituerait une réelle valeur et se traduirait par une véritable synergie ? La collaboration mise en place sur le thème de la résistance aux antimicrobiens [19] en est un exemple, qui a mis en avant la nécessité d'exploiter un grand nombre de sources afin d'accomplir des progrès substantiels et de s'assurer que nous allons de l'avant en fondant nos processus décisionnaires sur les données et sur des éléments probants. Cette question fera d'ailleurs l'objet d'un autre numéro de la *Revue* de l'OMSA.

Après cette plaidoirie en faveur des données et de la quantification, une précision importante s'impose car, bien que nécessaires, à elles seules ces données et cette quantification ne suffisent pas ; une telle approche réductionniste serait tout aussi détestable pour faire passer un message scientifique complexe. Les mots, eux aussi, importent [20], comme en témoignent, nous l'espérons, les articles réunis ici.

## Références

- [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. 2<sup>e</sup> éd. Griffin, Londres, Royaume-Uni, 413 pp.
- [4] Anderson R.M. & May R.M. (1992). – Infectious diseases of humans: dynamics and control. Oxford University Press, Oxford, Royaume-Uni, 766 pp.
- [5] Popper K.R. (1935). – Logik der Forschung. Zur Erkenntnistheorie der modernen Naturwissenschaft. Traduction anglaise: The logic of scientific discovery. Routledge Classics, Londres, Royaume-Uni, 1959, 513 pp. Disponible en ligne : <http://philotextes.info/spip/IMG/pdf/popper-logic-scientific-discovery.pdf> (consulté le 13 mars 2023). Traduction française : La logique de la découverte scientifique (Ph. Devaux, tr.), Payot, Paris, France, 1974, 496 pp.
- [6] Rothman K.J., Greenland S. & Lash T.L. (2008). – Modern epidemiology. 3<sup>e</sup> éd. Lippincott Williams and Wilkins, Philadelphie, États-Unis d'Amérique, 758 pp. Disponible en ligne : [http://students.aiu.edu/submissions/profiles/resources/onlineBook/a9c7D5\\_Modern\\_Epidemiology\\_3.pdf](http://students.aiu.edu/submissions/profiles/resources/onlineBook/a9c7D5_Modern_Epidemiology_3.pdf) (consulté le 13 mars 2023).
- [7] Guyatt G.H. & Groupe de travail de la médecine fondée sur les preuves (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. 2<sup>e</sup> éd. Lippincott Williams and Wilkins, Philadelphie, États-Unis d'Amérique, 466 pp.
- [9] Yusuf S., Cairns J.A., Camm A.J., Fallen E.L. & Gersh B.J. (2008). – Evidence-based cardiology. 2<sup>e</sup> éd. BMJ Publishing Group, Londres, Royaume-Uni, 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. 2<sup>e</sup> éd. Churchill Livingstone, Toronto, Canada, 280 pp.
- [11] Cockcroft P.D. & Holmes M.A. (2003). – Handbook of evidence-based veterinary medicine. Blackwell Publishing, Oxford, Royaume-Uni, 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, coord.). Academic Press, New York, États-Unis d'Amérique, 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] Organisation des Nations unies pour l'alimentation et l'agriculture (FAO), Organisation mondiale de la santé animale (OMSA) & Organisation mondiale de la santé (OMS) (2017). – The Tripartite's commitment: providing multi-sectoral, collaborative leadership in addressing health challenges. FAO, OMSA & OMS, Rome, Italie, 4 pp. Disponible en ligne : <https://doc.woah.org/dyn/portal/index.xhtml?page=alo&aloId=34824&espaceId=100> (consulté le 13 mars 2023).

[19] Organisation mondiale de la santé (OMS), Organisation des Nations unies pour l'alimentation et l'agriculture (FAO), Organisation mondiale de la santé animale (OMSA) & Programme des Nations unies pour l'environnement (PNUE) (2022). – Cadre stratégique de collaboration sur la résistance aux antimicrobiens : ensemble pour une seule santé. OMS, FAO, OMSA & PNUE, Genève, Suisse, 19 pp. Disponible en ligne : <https://www.who.int/fr/publications/i/item/9789240045408> (consulté le 13 mars 2023).

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

© 2023 Reid S.W.J.; licensee the World Organisation for Animal Health. This is an open access article distributed under the terms of the Creative Commons Attribution IGO Licence (<https://creativecommons.org/licenses/by/3.0/igo/legalcode>), which permits unrestricted use, distribution and reproduction in any medium, provided the original work is properly cited. In any reproduction of this article there should not be any suggestion that WOAHA or this article endorses any specific organisation, product or service. The use of the WOAHA logo is not permitted. This notice should be preserved along with the article's original URL.

Pre-print