

# Introduction to epidemiology

Solymosi Norbert

*Quantitative veterinary epidemiology*

Department of Microbiology and Infectious Diseases

University of Veterinary Medicine Budapest



A 21st century clinician who cannot critically read a study is as unprepared as one who cannot take a blood pressure or examine the cardiovascular system.

## SHARE



The world may never learn the origin of the novel coronavirus, but intriguing information keeps materializing even in the pandemic's third year. A pair of scientists from Hungary may have stumbled into more evidence supporting the theory that Covid-19 emerged from a laboratory.



OPINION: POTOMAC WATCH



**The Revolt of the Parents Is Spreading**



00:00

1x



SUBSCRIBE



A paper [posted](#) recently suggests that genetic data from soil samples sent to a Chinese firm for sequencing could have been contaminated by an ancestor of the Covid-19 virus. The Wuhan Institute of Virology (WIV), where some scientists believe the pandemic likely began, has used the same firm.

If the samples were analyzed in early 2020, it's possible they became contaminated when other scientists were studying Covid-19 as it began to spread. But it would be suspicious if the analysis was done before the end of December 2019, when China says it first identified the virus.



## Findings

**Seroprevalence** was 5.0% (95% CI 4.7–5.4) by the point-of-care test and 4.6% (4.3–5.0) by immunoassay, with a **specificity–sensitivity** range of 3.7% (3.3–4.0; both tests positive) to 6.2% (5.8–6.6; either test positive), with no differences by sex and lower seroprevalence in children younger than 10 years (<3.1% by the point-of-care test). There was substantial **geographical variability**, with higher **prevalence** around Madrid (>10%) and lower in coastal areas (<3%). Seroprevalence among 195 participants with positive PCR more than 14 days before the study visit ranged from 87.6% (81.1–92.1; both tests positive) to 91.8% (86.3–95.3; either test positive). In 7273 individuals with anosmia or at least three symptoms, seroprevalence ranged from 15.3% (13.8–16.8) to 19.3% (17.7–21.0). Around a third of seropositive participants were asymptomatic, ranging from 21.9% (19.1–24.9) to 35.8% (33.1–38.5). Only 19.5% (16.3–23.2) of symptomatic participants who were seropositive by both the point-of-care test and immunoassay reported a previous PCR test.

## Results

Of the 8910 patients with Covid-19 for whom discharge status was available at the time of the analysis, a total of 515 died in the hospital (5.8%) and 8395 survived to discharge. The factors we found to be independently associated with an increased risk of in-hospital death were an age greater than 65 years (mortality of 10.0%, vs. 4.9% among those  $\leq$ 65 years of age; **odds ratio**, 1.93; 95% **confidence interval [CI]**, 1.60 to 2.41), coronary artery disease (10.2%, vs. 5.2% among those without disease; odds ratio, 2.70; 95% CI, 2.08 to 3.51), heart failure (15.3%, vs. 5.6% among those without heart failure; odds ratio, 2.48; 95% CI, 1.62 to 3.79), cardiac arrhythmia (11.5%, vs. 5.6% among those without arrhythmia; odds ratio, 1.95; 95% CI, 1.33 to 2.86), chronic obstructive pulmonary disease (14.2%, vs. 5.6% among those without disease; odds ratio, 2.96; 95% CI, 2.00 to 4.40), and current smoking (9.4%, vs. 5.6% among former smokers or nonsmokers; odds ratio, 1.79; 95% CI, 1.29 to 2.47). No increased risk of in-hospital death was found to be associated with the use of ACE inhibitors (2.1% vs. 6.1%; odds ratio, 0.33; 95% CI, 0.20 to 0.54) or the use of ARBs (6.8% vs. 5.7%; odds ratio, 1.23; 95% CI, 0.87 to 1.74).



## Covid-19 antibody testing

Diagnostic accuracy of three serological tests

**Summary**

Current evidence does not support the continued use of existing point-of-care covid-19 serology tests with the LFIA method.  
Only two studies evaluated performance at the point of care

**Study design**

Systematic review and meta-analysis

Sensitivity estimated from people with confirmed covid-19

Specificity estimated mostly using pre-epidemic samples, or low risk individuals

**Data sources**

40 studies

29 842 tests

More than one sample may have originated from the same participant

**Comparison****ELISA**

Enzyme Linked Immunosorbent Assays

**LFIA**

Lateral Flow Immunoassays

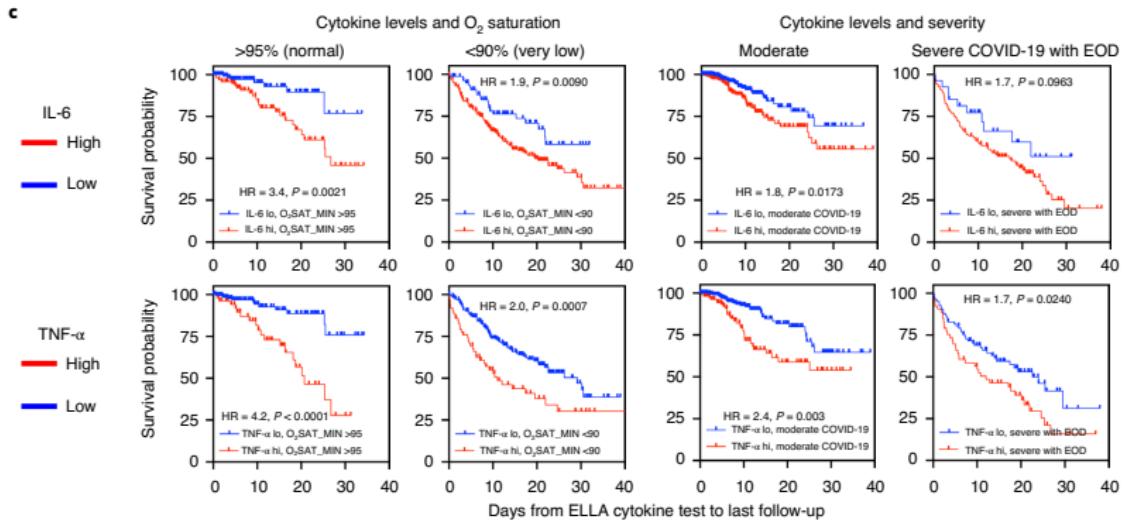
**CLIA**

Chemiluminescent Immunoassays

**Results**

Under 10% prevalence

**TRUE result****FALSE result**



**Fig. 4 | Cytokine levels correlate with severity and independently predict survival.** Correlation of cytokine levels with established inflammatory and severity measurements. **a**, Correlation of each cytokine with each metric ( $n=1,106$  for fever,  $n=1,112$  for  $O_2$  saturation,  $n=1,023$  for CRP,  $n=926$  for D-dimer,  $n=1,017$  for ferritin,  $n=1,038$  for platelets and  $n=1,023$  for disease severity score), using the same univariate and multivariate analyses as in the Fig. 2 legend. Error bar indicates the median  $\pm$  95% CI. **b**, Competing risk analysis ( $n=671$ ) showing survival differences by IL-6 and TNF- $\alpha$  levels, after adjusting the following variables: IL-6, IL-8, TNF- $\alpha$ , IL-1 $\beta$ , age, sex, race/ethnicity, smoking status, asthma, atrial fibrillation, cancer, CHF, CKD, COPD, diabetes, hypertension, sleep apnea, severity, systolic blood pressure max,  $O_2$  saturation min, D-dimer, albumin, calcium, chloride and platelet count. **c**, Kaplan-Meier univariate analyses of survival by IL-6 and TNF- $\alpha$  levels in patients with normal ( $n=257$ ), low ( $n=258$ ) or very low ( $n=287$ )  $O_2$  saturation, or in patients with moderate ( $n=588$ ) versus severe COVID-19 with end organ damage ( $n=136$ ), as measured at the first available test.

## EDITORIAL

# Will COVID-19 be evidence-based medicine's nemesis?

Trisha Greenhalgh \*

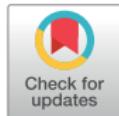
Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, United Kingdom

\* [trish.greenhalgh@phc.ox.ac.uk](mailto:trish.greenhalgh@phc.ox.ac.uk)

Once defined in rhetorical but ultimately meaningless terms as “the conscientious, judicious and explicit use of current best evidence in making decisions about the care of individual patients” [1], evidence-based medicine rests on certain philosophical assumptions: a singular truth, ascertainable through empirical enquiry; a linear logic of causality in which interventions have particular effect sizes; rigour defined primarily in methodological terms (especially, a hierarchy of preferred study designs and tools for detecting bias); and a deconstructive approach to problem-solving (the evidence base is built by answering focused questions, typically framed as ‘PICO’—population-intervention-comparison-outcome) [2].

The trouble with pandemics is that these assumptions rarely hold. A pandemic-sized problem can be framed and contested in multiple ways. Some research questions around COVID-19, most notably relating to drugs and vaccines, are amenable to randomised controlled trials (and where such trials were possible, they were established with impressive speed and efficiency [3, 4]). But many knowledge gaps are broader and cannot be reduced to PICO-style questions. Were care home deaths avoidable [5]? Why did the global supply chain for personal protective equipment break down [6]? What role does health system resilience play in controlling the pandemic [7]? And so on.

Against these—and other—wider questions, the neat simplicity of a controlled, interven-



Check for  
updates

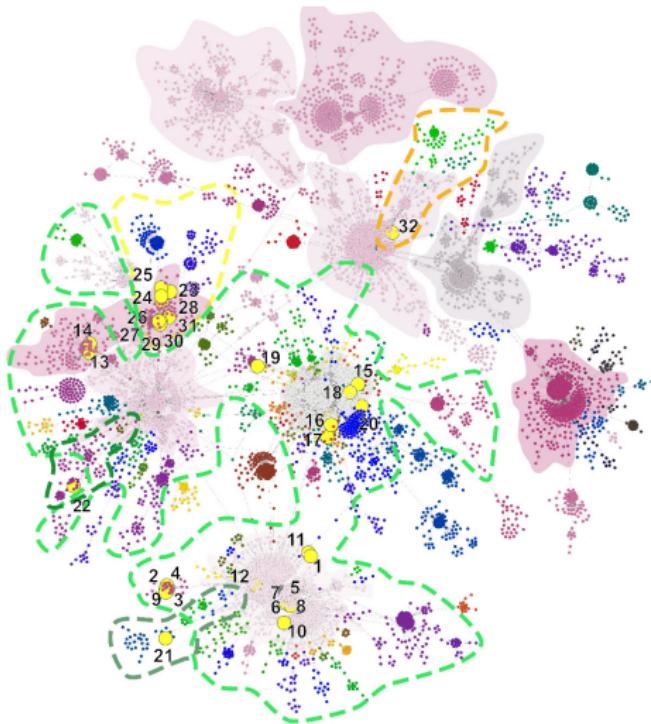
---

 OPEN ACCESS

Greenhalgh (2020)

QEpi (13/9/2022)

Intro



**Figure 1.** Genetic network analysis of 7864 SARS-CoV-2 complete genomic sequences. Hungarian strains are indicated with numbered yellow dots—numbers referring to Table 1. The nine major clades are represented by a solid color. Genetic lineages are marked with colored dotted lines, where green lines are bordering B 1, B 1.1 and B 1.11; yellow and orange lines mark B 1.5 and B 3, respectively.

**Epidemiology** is the study of disease in populations and of factors that determine its occurrence; the key word being **populations**.

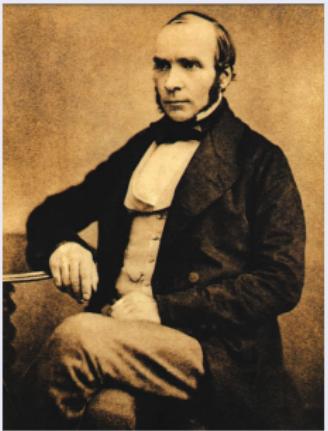
Epidemiology is concerned with the prevention and control of disease in human and animal **populations**. Veterinary epidemiology additionally includes the investigation and assessment of other health-related events, notably **productivity**.

Objectives of epidemiology:

- determination of the origin of a disease whose cause is known
- investigation and control of a disease whose cause is either unknown or poorly understood
- acquisition of information on the ecology and natural history of a disease
- planning, monitoring and assessment of disease control programmes
- assessment of the economic effects of a disease, and analysis of the costs and economic benefits of alternative control programmes

- **Causality models**
- **Measures of health**
- **Diagnostic tests**
- **Measures of associations**
- **EBVM**
- **Spatial epidemiology**
- Outbreak, spreading models
- Sample size estimations
- Monitoring & surveillance systems (MOSS)





John Snow  
1813-1858



<https://www.youtube.com/watch?v=2GhQ7AYHLr4>

Epidemiology studies the occurrence, associations of disease and factors, causes.

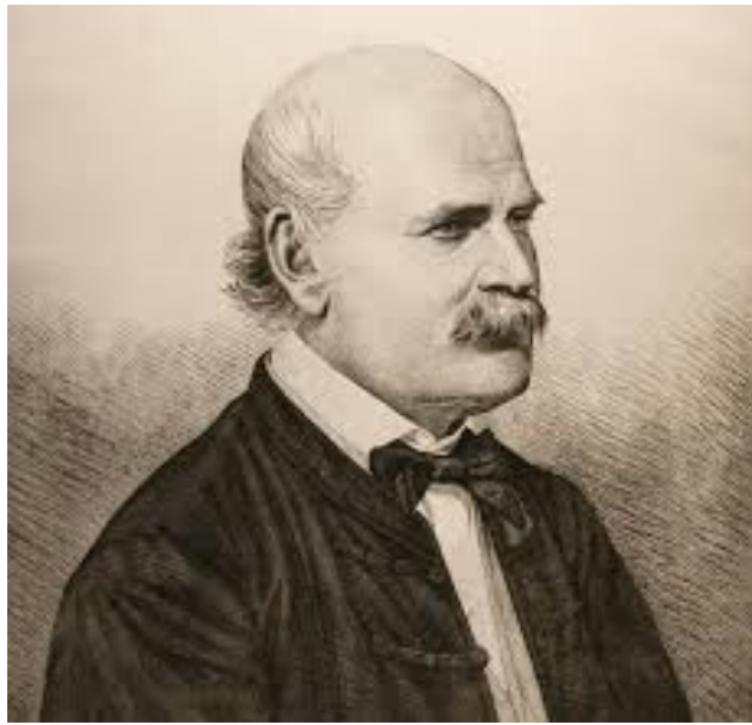
The study of causality try to establish the connection of cause and consequence, in natural science or philosophy.

In natural science the main goal to identify the causes, explaining certain natural phenomena.

association or cause?

*„Fortis imaginatio generat casum“*  
Michel de Montaigne (1533–1592)

A serological survey of 213 randomly chosen dairy farm residents in the Manawatu showed that 66 (34%) of the people who milked cows had leptospiral titres  $\leq 1:24$  by the Microscopic Agglutination Test. Forty-eight (72.7%) of these people had titres to hardjo, while 29 (43.9%) had titres to pomona. Dual hardjo/pomona titres occurred in 12 people. Ballum and Copenhageni accounted for 8% of the titres found. Women milkers and farm residents who did not milk were all serologically negative. A third of the seropositive milkers had a history of clinical leptospirosis. Other factors which significantly correlated with leptospiral titres included the time spent in the dairy shed during milking, the **wearing of shorts**, the keeping of pigs for sale, and the number of years the individual had been working on a dairy farm. The type of milking shed and the size of the herd were interrelated and both showed strong trends towards a correlation with serological prevalence.



Ignaz Philipp Semmelweis (1818–1865)

- 1847: chlorinated lime for hand washing
- 1879: Pasteur presented *streptococcus* from childbed fever
- causality? association? efficiency

### *Leviticus 13,4:*

*"If the shiny spot on the skin is white but does not appear to be more than skin deep and the hair in it has not turned white, the priest is to isolate the affected person for seven days."*

In the case of "simple cause" infectious disease postulates of *Koch* (1884) are sufficient to explain the causality.

---

An organism is causal if

---

- 1 it is present in all cases of the disease
  - 2 it does not occur in another disease as a fortuitous and non-pathogenic parasite
  - 3 it is isolated in pure culture from an animal, is repeatedly passed, and induces the same disease in other animals
- 

In the case of multifactorial diseases the Koch's model is not applicable.

## The top ten causes of death

Year	Rank	Disease	Proportion (%)
1860	1	Tuberculosis	19.8
	2	Diarrhoea, enteritis	15.0
	3	Cholera	6.4
	4	Pneumonia/influenza/bronchitis	6.1
	5	Infantile convulsions	5.9
	6	Diphtheria, croup	2.7
		Disentery	2.7
		Stroke	2.7
	9	Scarlet fever	2.5
	10	Nephritis	2.4

## The top ten causes of death

Year	Rank	Disease	Proportion (%)
1900	1	Pneumonia/influenza/bronchitis	14.4
	2	Tuberculosis	11.3
	3	Diarrhoea, enteritis	8.1
	4	Heart diseases	8.0
	5	Nephritis	4.7
	6	Accidents	4.5
	7	Stroke	4.2
		Diseases of early infancy	4.2
	9	Cancer	3.7
	10	Diphtheria	2.3

## The top ten causes of death

Year	Rank	Disease	Proportion (%)
1970	1	Heart diseases	38.3
	2	Cancer	17.2
	3	Stroke	10.8
	4	Pneumonia/influenza/bronchitis	3.6
	5	Accidents, suicide	3.1
	6	Motor vehicle accidents	2.8
	7	Diseases of early infancy	2.3
	8	Diabetes	2.0
	9	Arteriosclerosis	1.7
	10	Cirrhosis	1.6

## Hill's criteria for causation (1965):

### ① Strength of association

- strong associations are more likely to be causal
- cannot infer that weak association is not causal

### ② Consistency

- has the cause and effect relationship identified by a number of different researchers?

### ③ Specificity

- a single exposure should cause a single disease
- when present, specificity does provide causality, but its absence does not preclude causation

### ④ Temporality

- cause must precede effect

### ⑤ Dose-response relationship

- as the level of exposure is increased, the rate of disease also increases

### ⑥ Plausibility and coherence

- biological mechanism

### ⑦ Experimental evidence

### ⑧ Analogy (BSE/scrapie)

## *Evans's postulates (1976)*

---

- 1 the proportion of individuals with disease should be significantly higher in those **exposed** to the supposed cause than in those who are not
  - 2 exposure to the supposed cause should be present more commonly in those with than those without the disease, when all other risk factors are held constant
  - 3 the number of new cases of disease should be significantly higher in those exposed to the supposed cause than in those not so exposed, as shown in **prospective** studies
  - 4 temporally, the disease should follow exposure to the supposed cause with a distribution of incubation periods on a bell-shaped curve
-

## *Evans's postulates (1976)*

---

- 5 a spectrum of host responses, from mild to severe, should follow exposure to the supposed cause along a logical biological gradient
  - 6 a measurable host response (e.g. antibody) should appear regularly following exposure to the supposed cause in those lacking this response before exposure, or should increase in magnitude if present before exposure, this pattern should not occur in individuals not so exposed
  - 7 experimental reproduction of disease should occur with greater frequency in animals or man appropriately exposed to the supposed cause than in those not so exposed; this exposure may be deliberate in volunteers, experimentally induced in the laboratory, or demonstrated in a controlled regulation of natural exposure
-

## *Evans's postulates (1976)*

---

- 8 elimination (e.g. removal of a specific infectious agent) or modification (e.g. alteration of a deficient diet) of supposed cause should decrease the frequency of occurrence of the disease
  - 9 prevention or modification of the host's response (e.g. by immunization) should decrease or eliminate the disease that normally occurs on exposure to the supposed cause
  - 10 all relationships and associations should be biologically and epidemiologically credible
-

No. 4356 April 25, 1953

NATURE

737

equipment, and to Dr. G. E. R. Deacon and the captain and officers of R.R.S. *Discovery II* for their part in masking the observations.

\* Young, F. B., Gerrard, H., and Jevons, W., *Phil. Mag.*, **40**, 149 (1931).

<sup>1</sup> Longuet-Higgins, M. S., *Mon. Natl. Roy. Astr. Soc., Geophys. Suppl.*, **5**, 285 (1949).

<sup>2</sup> Von Arx, W. S., Woods Hole Papers in Phys. Oceanogr. Meteor., **11**, 103 (1951).

<sup>3</sup> Ekman, V. W., *Actua. Mat. Astron. Fysik* (Stockholm), **9**(11) (1905).

## MOLECULAR STRUCTURE OF NUCLEIC ACIDS

### A Structure for Deoxyribose Nucleic Acid

WE wish to suggest a structure for the salt of deoxyribose nucleic acid (D.N.A.). This structure has novel features which are of considerable biological interest.

A structure for nucleic acid has already been proposed by Pauling and Corey<sup>1</sup>. They kindly made their manuscript available to us in advance of publication. Their model consists of three intertwined chains, with the phosphates near the fibre axis, and the bases on the outside. In our opinion, this structure is unsatisfactory for two reasons: (1) We believe that the material which gives the X-ray diagrams is the salt, not the free acid. Without the acidic hydrogen atoms it is not clear what forces would hold the structure together, especially as the negatively charged phosphates near the axis will repel each other. (2) Some of the van der Waals distances appear to be too small.

Another three-chain structure has also been suggested by Fraser (in the press). In his model the phosphates are on the outside and the bases on the inside, linked together by hydrogen bonds. This structure as described is rather ill-defined, and for this reason we shall not comment on it.

We wish to put forward a radically different structure for the salt of deoxyribose nucleic acid. This structure has two helical chains each coiled round the same axis (see diagram). We have made the usual chemical assumptions, namely, that each chain consists of phosphate diester groups joining  $\beta$ -D-deoxyribofuranose residues with 3',5' linkages. The two chains (but not their bases) are related by a dyad perpendicular to the fibre axis. Both chains follow right-handed helices, but owing to the dyad the sequences of the atoms in the two chains run in opposite directions. Each chain closely resembles Furberg's<sup>2</sup> model No. I, that is, the bases are on the inside of the helix and the phosphates on the outside. The configuration

is a residue on each chain every 3.4 Å. in the z-direction. We have assumed an angle of 36° between adjacent residues in the same chain, so that the structure repeats after 10 residues on each chain, that is, after 34 Å. The distance of a phosphorus atom from the fibre axis is 10 Å. As the phosphates are on the outside, cations have easy access to them.

The structure is an open one, and its water content is rather high. At lower water contents we would expect the bases to tilt so that the structure could become more compact.

The novel feature of the structure is the manner in which the two chains are held together by the purine and pyrimidine bases. The planes of the bases are perpendicular to the fibre axis. They are joined together in pairs, a single base from one chain being hydrogen-bonded to a single base from the other chain, so that the two lie side by side with identical z-coordinates. One of the pair must be a purine and the other a pyrimidine for bonding to occur. The hydrogen bonds are made as follows: purine position 1 to pyrimidine position 1; purine position 6 to pyrimidine position 6.

If it is assumed that the bases only occur in the structure in the most plausible tautomeric forms (that is, with the keto rather than the enol configurations) it is found that only specific pairs of bases can bond together. These pairs are: adenine (purine) with thymine (pyrimidine), and guanine (purine) with cytosine (pyrimidine).

In other words, if an adenine forms one member of a pair, on either chain, then on these assumptions the other member must be thymine; similarly for guanine and cytosine. The sequence of bases on a single chain does not appear to be restricted in any way. However, if one specific pair of bases can be formed, it follows that the sequence of bases on one chain is given, then the sequence on the other chain is automatically determined.

It has been found experimentally<sup>3,4</sup> that the ratio of the amounts of adenine to thymine, and the ratio of guanine to cytosine, are always very close to unity for deoxyribosy nucleic acid.

It is probably impossible to build this structure with a ribose sugar in place of the deoxyribose, as the extra oxygen atom would make too close a van der Waals contact.

The previously published X-ray data<sup>5,6</sup> on deoxyribose nucleic acid are insufficient for a rigorous test of our structure. So far as we can tell, it is roughly compatible with the experimental data, but it must be regarded as unproved until it has been checked against more exact results. Some of these are given in the following communications. We were not aware of the details of the results presented there when we devised our structure, which rests mainly though not entirely on published experimental data and stereochemical arguments.

It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material.

Full details of the structure, including the conditions assumed in building it, together with a set of co-ordinates for the atoms, will be published elsewhere.

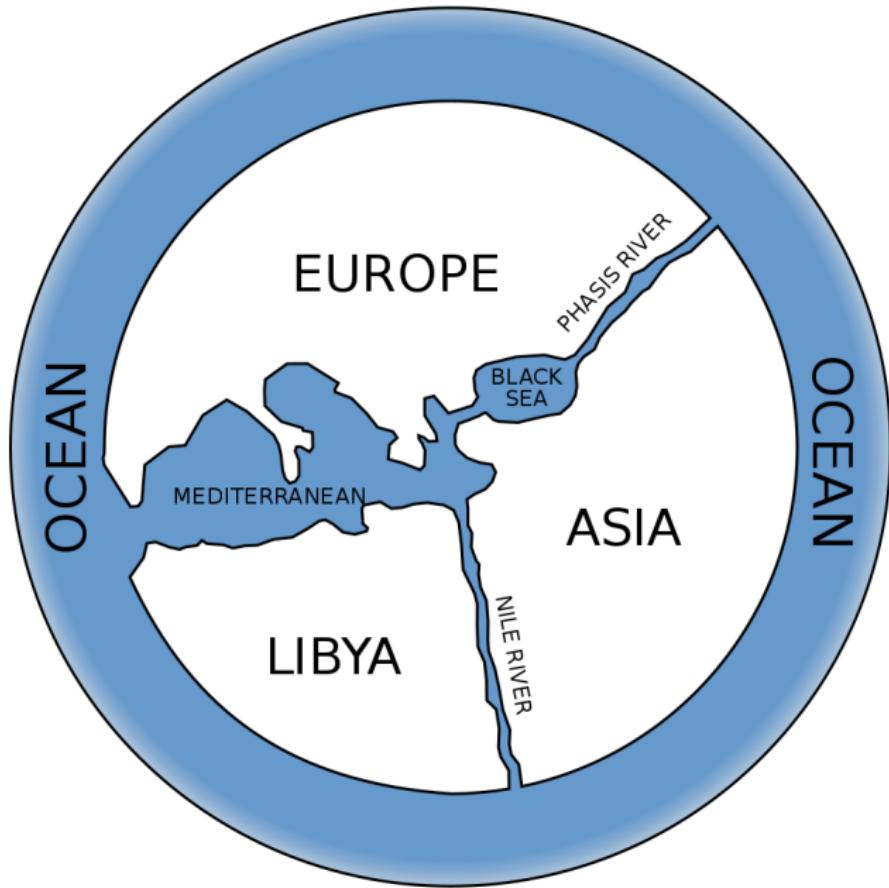
We are much indebted to Dr. Jerry Donohue for



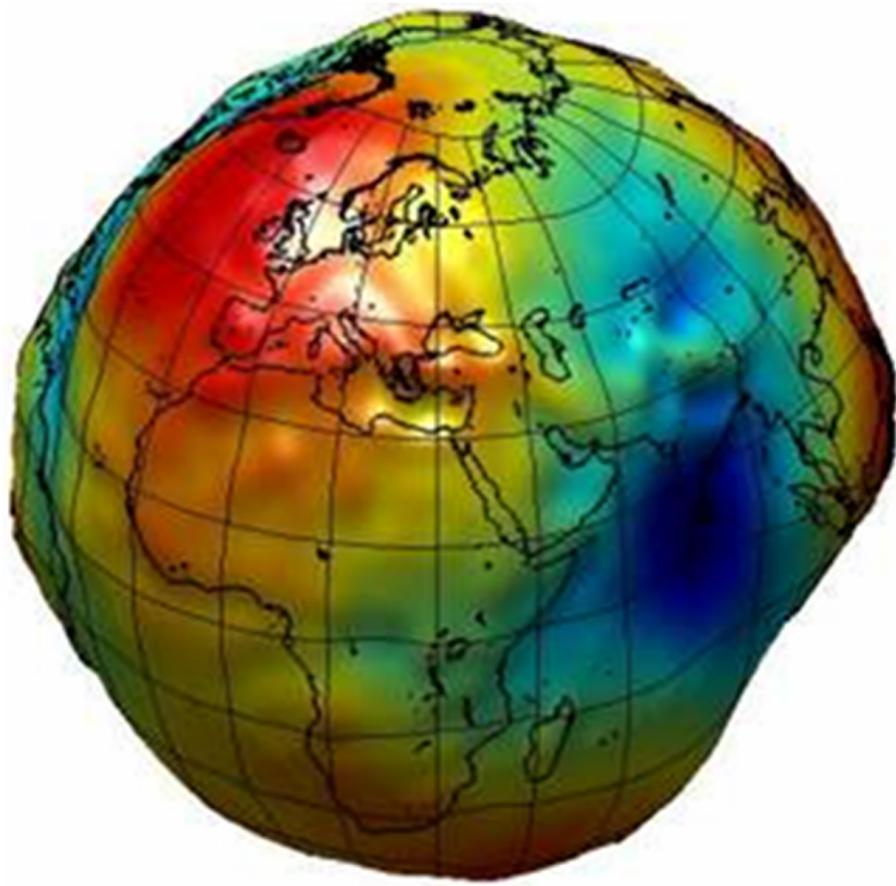
"All models are wrong  
but some are useful"

George E. P. Box

1919–2013

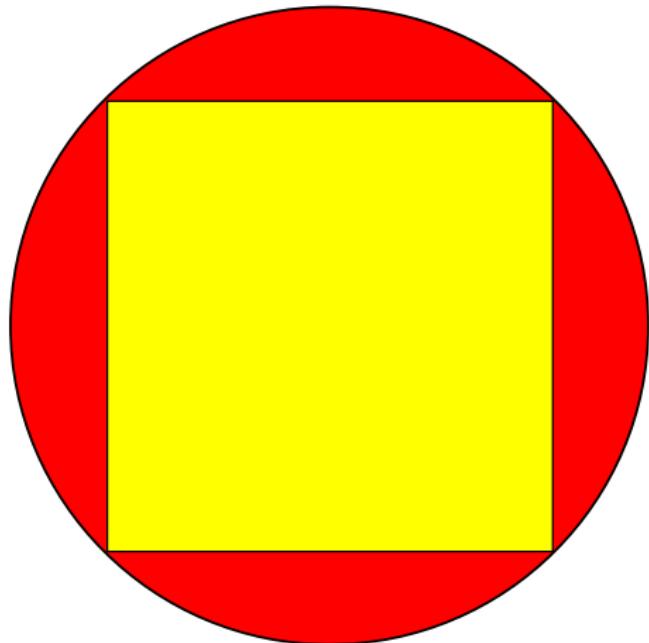






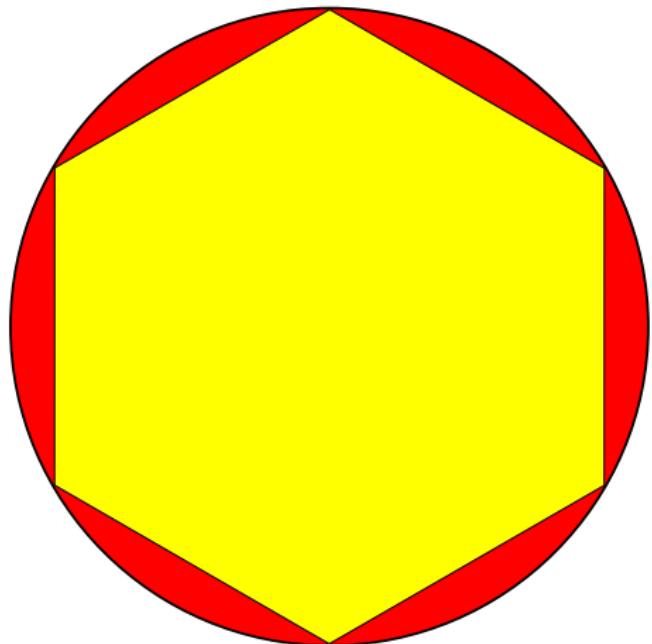


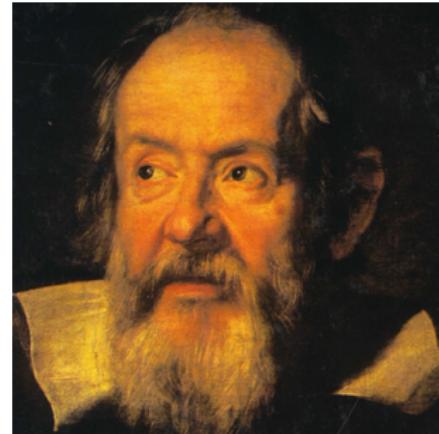
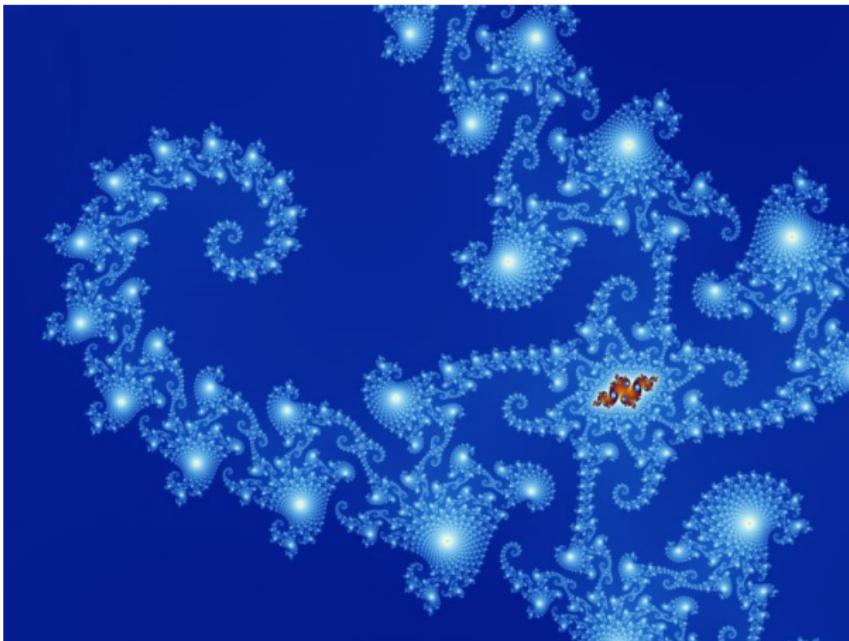
Nicolaus Cusanus  
1401–1464





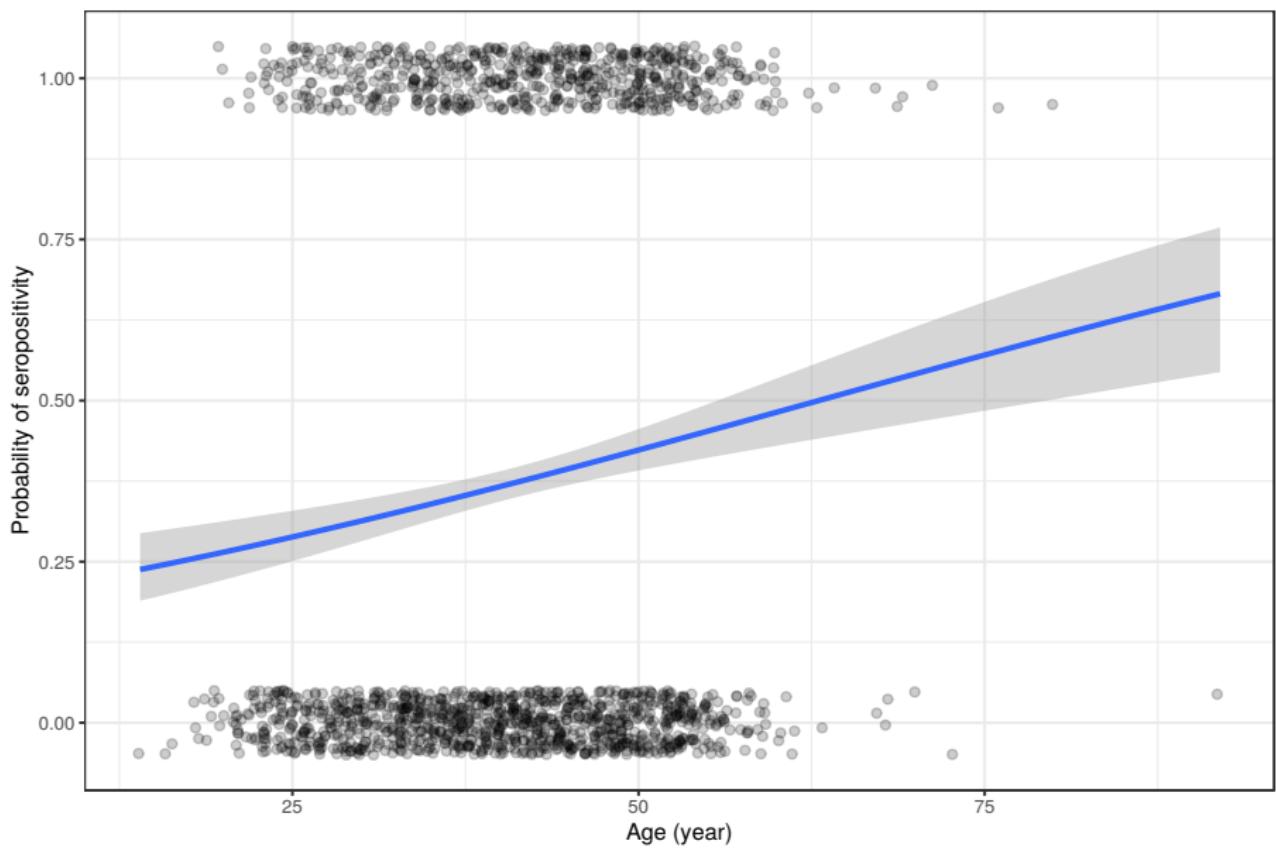
Nicolaus Cusanus  
1401–1464

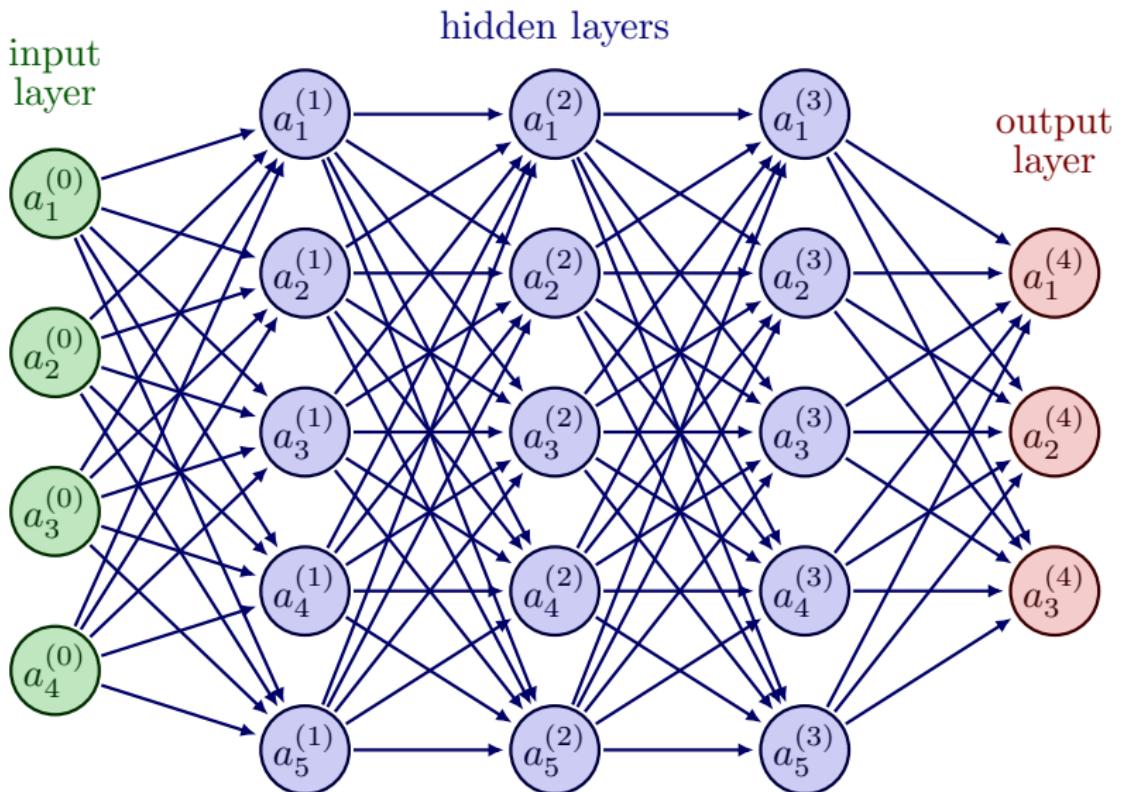


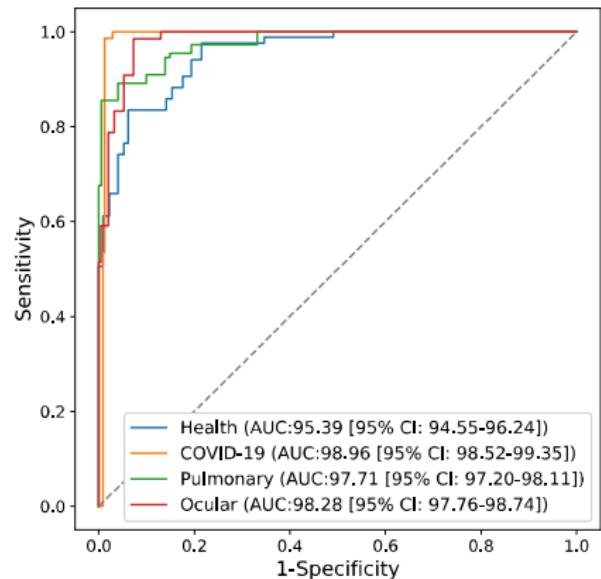
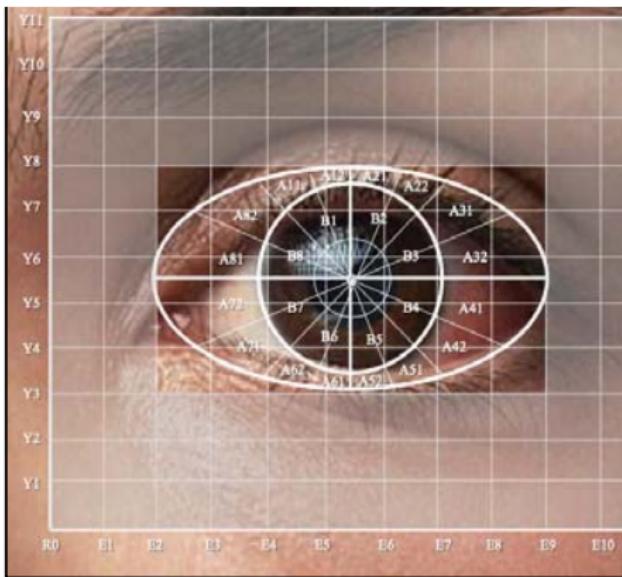


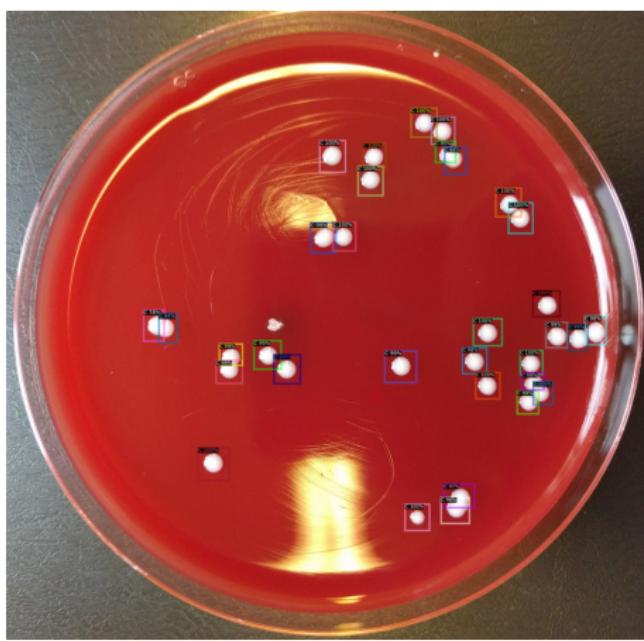
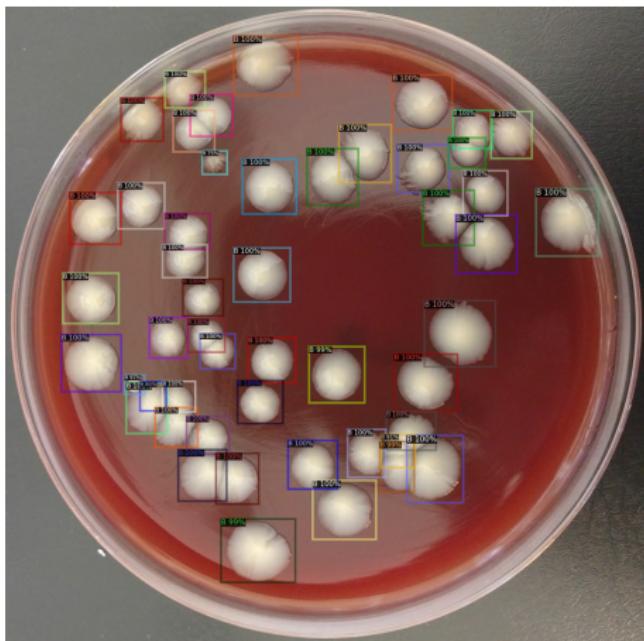
Galileo Galilei  
1564–1642

$$Z_{n+1} = z_n^2 + C$$









- Bastos, M. L., G. Tavaziva, S. K. Abidi, J. R. Campbell, L.-P. Haraoui, J. C. Johnston, Z. Lan, S. Law, E. MacLean, A. Trajman, et al. (2020). Diagnostic accuracy of serological tests for covid-19: systematic review and meta-analysis. *BMJ* 370.
- Del Valle, D. M., S. Kim-Schulze, H.-H. Huang, N. D. Beckmann, S. Nirenberg, B. Wang, Y. Lavin, T. H. Swartz, D. Madduri, A. Stock, et al. (2020). An inflammatory cytokine signature predicts covid-19 severity and survival. *Nature medicine*, 1–8.
- Dohoo, I., W. Martin, and H. Stryhn (2002). *Veterinary Epidemiologic Research* (2nd ed.). Charlottetown, Prince Edward Island, Canada: VER Inc.
- Fu, Y., F. Li, W. Wang, H. Tang, X. Qian, M. Gu, and X. Xue (2020). A new screening method for covid-19 based on ocular feature recognition by machine learning tools. *medRxiv*.
- Glasziou, P., A. Burls, and R. Gilbert (2008). Evidence based medicine and the medical curriculum. *BMJ (Clinical Research Ed.)* 337(7672), 704–705.
- Greenhalgh, T. (2020, 06). Will COVID-19 be evidence-based medicine's nemesis? *PLOS Medicine* 17(6), 1–4.
- Kemenesi, G., S. Zeghbib, B. A. Somogyi, G. E. Tóth, K. Bánya, N. Solymosi, P. M. Szabo, I. Szabó, Á. Bálint, P. Urbán, et al. (2020). Multiple sars-cov-2 introductions shaped the early outbreak in central eastern europe: comparing hungarian data to a worldwide sequence data-matrix. *Viruses* 12(12), 1401.
- Mackintosh, C., L. M. Schollum, R. Harris, D. Blackmore, A. Willis, N. Cook, and J. Stoke (1980). Epidemiology of leptospirosis in dairy farm workers in the Manawatu Part I: A cross-sectional serological survey and associated occupational factors. *New Zealand Veterinary Journal* 28, 245–250.
- Mehra, M. R., S. S. Desai, S. Kuy, T. D. Henry, and A. N. Patel (2020). Cardiovascular disease, drug therapy, and mortality in covid-19. *New England Journal of Medicine*.
- Noordhuizen, J. P. T. M., K. Frankena, M. Thrusfield, and E. A. M. Graat (2001). *Application of Quantitative Methods in Veterinary Epidemiology*. Wageningen, The Netherland: Wageningen Pers.
- Pfeiffer, D. (2002). Veterinary epidemiology: An introduction. [ww3.panaftosa.org.br/Comp/MAPA/431857.pdf](http://ww3.panaftosa.org.br/Comp/MAPA/431857.pdf).
- Pfeiffer, D. (2010). *Veterinary Epidemiology: An Introduction*. Oxford, UK: Wiley.
- Pollán, M., B. Pérez-Gómez, R. Pastor-Barriuso, J. Oteo, M. A. Hernán, M. Pérez-Olmeda, J. L. Sanmartín, A. Fernández-García, I. Cruz, N. F. de Larrea, et al. (2020). Prevalence of SARS-CoV-2 in Spain (ENE-COVID): a nationwide, population-based seroepidemiological study. *The Lancet* 396(10250), 535–544.
- Smith, R. D. (2005). *Veterinary Clinical Epidemiology: From Patient to Population* (3rd ed.). Boca Raton, Florida, USA: CRC Press.
- Thrusfield, M., R. Christley, H. Brown, P. J. Diggle, N. French, K. Howe, L. Kelly, A. O'Connor, J. Sargeant, and H. Wood (2018). *Veterinary Epidemiology* (4th ed.). Oxford, UK: Wiley.