Sprechen Sie Statistik?

KATHRINE FREY FRØSLIE

E-mail: kathrine.frey.froslie@nmbu.no
Kathrine Frey Frøslie, associate professor in biostatistics at the Norwegian University of Life Sciences. She runs the popular science knitting blog statistrikk.no.
The author has completed the ICMJE form and declares no conflicts of interest.

JO RØISLIEN

Jo Røislien, professor of medical statistics at the Faculty of Health Sciences at the University of Stavanger, and prominent television personality and presenter.
The author has completed the ICMJE form and declares no conflicts of interest.

Is statistics a foreign language, like German, Latin or Mandarin?
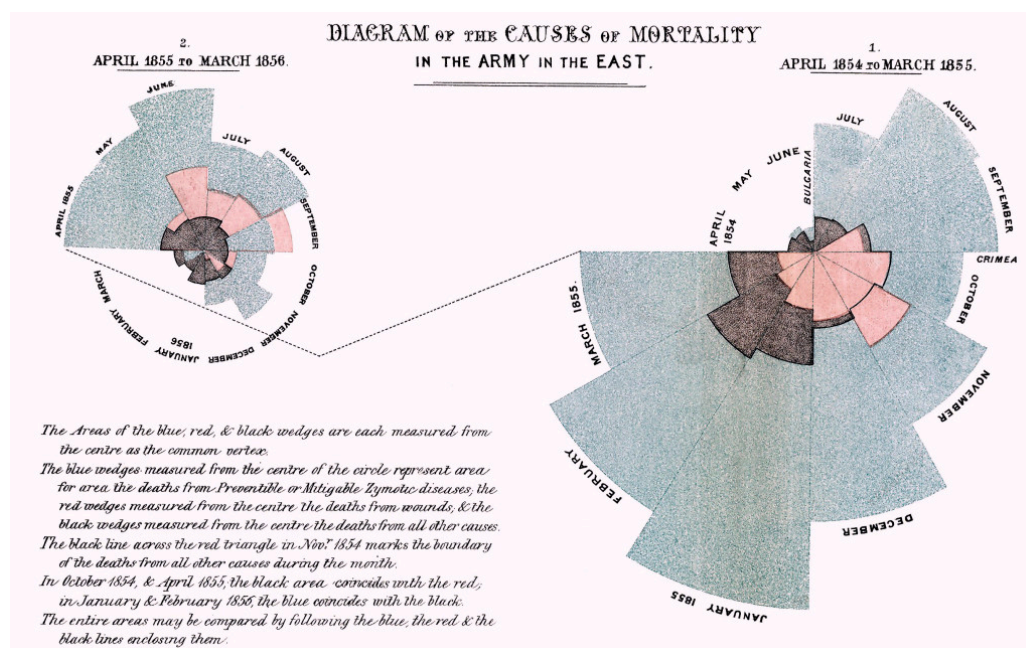
Figure 1 Florence Nightingale changed the British healthcare system by using numbers to tell stories about the many. Illustration: Florence Nightingale / Wikimedia Commons

Using numbers to tell stories is a cornerstone of medical research. Clinical researchers translate observations of blood, sugar, sweat and tears into numbers and categories that can be analysed using statistical methods. They exchange experiences, empirical data and results with their peers through tables, figures and statistical analyses in scientific articles, at conferences and in textbooks.

But although statistics is central to acquiring and disseminating medical knowledge, few clinicians have statistics as their primary field of interest. The numbers and analyses can often seem foreign, formulated in a language that they weren't trained how to use. Clinicians who want to conduct research must perform and communicate at an elite international level – in a language that is not their own.

Statistics as language

A language is a tool for communicating thoughts, ideas and knowledge. The same is true for statistics. Both language and statistics have a technical grammar. While linguists talk of subjects, predicates and cases, statisticians talk of standard deviations, p-values and regression coefficients. Both language and statistics have alphabets and punctuation. Where languages use letters and characters, statistics is based on mathematical notation, with mathematical formulas for mean, standard deviation and regression equations.

To master a language, you need to gain an understanding of a number of concepts. The same is true for statistics. A noun is a word for a thing. A standard deviation is a number for variation. The subject of a sentence is the doer. And the odds is the probability of something happening divided by the probability of it not happening.

For both language and statistics, context is key. Knowledge of metre and rhyme is for the most part of little relevance – or interest – until this knowledge is put into context by a specific poem, for example where the author uses them as a means to evoke the sorrow of losing a child. Knowledge of how the SPSS software package is used to calculate a regression coefficient has little clinical relevance before it is put into context by a specific research project and is used to say that the risk of stillbirth increases after 42 weeks of pregnancy.

Both language and statistics can be used to tell important stories about the world we live in. Language can use words to draw the reader into stories of the known and the unknown, to convey emotions and understanding. Similarly, numbers can summarise a collection of individual patient histories so well that the clinical insight becomes obvious.

Statistics as non-language

Even though statistics bears similarities to language, it is not a true language. A language has subjects that do something, verbs that describe action; it has adjectives and superlatives. Statistics has none of these things. You cannot say 'I'm really looking forward to the summer holidays!' in 'statisticsish'.

Statistics is a special tool for communicating a very specific type of information: quantitative information. Statistics is based on mathematics, and both the calculations and the language used to describe the methods and the numbers must be precisely formulated for the message to be correct. You cannot change or adapt mathematics to achieve communication – both the sender and the receiver must learn the maths. This is different from the modus operandi of all linguistic communication.

In any language, you will be understood even if you use the wrong word and there are major grammatical shortcomings in the way you speak or write. If you say you are going to visit 'gan-gan', then we understand who you are talking about, even though you will not find that word in any dictionary. If you use 'affect' instead of 'effect' or 'there' instead of 'their' in a text message, you will still get your point across. That is not the case with statistics. If you use the wrong words or the wrong numbers to summarise your data, or you present an incorrect statistical analysis, not only will your communication fail – you could jeopardise your entire research project.

The story of the many

Individual stories from everyday clinical work can be compelling. They are about bleeding, eating disorders, adverse effects and death. There are many individual stories, but they are still individual stories. In order to offer the best possible health care, it is not sufficient to tell stories about individual patients. We must tell the story of the many. And to tell that story, we need statistics.

Statistics is built on invariable absolutes and right and wrong. But statistics is not *just* mathematics. Statistics is about communicating what the data collected by a research project tells us about everyday clinical work. To understand the stories about the many, to be understood when you want to communicate your experiences, empirical data and results, you must master the language of the many. You must know 'statisticsish'.

This article is based on the lead author's lecture at the ADVICE2018 conference.