



# Jane and John do not exist

---

## MEDISIN OG TALL

### JO RØISLIEN

E-mail: [jo@joroislien.no](mailto:jo@joroislien.no)

Jo Røislien, professor of medical statistics at the Faculty of Health Sciences at the University of Stavanger, and prominent science communicator and television presenter.

The author has completed the ICMJE form and reports no conflicts of interest.

### KATHRINE FREY FRØSLIE

Kathrine Frey Frøslie, statistician at the Norwegian National Advisory Unit on Women's Health, Oslo University Hospital, and runs the popular science knitting blog [Statistrikk.no](http://Statistrikk.no).

The author has completed the ICMJE form and reports no conflicts of interest.

---

The average person is a ghost from the past.

An important use of statistics is to condense many individual observations into a few summary figures, so-called descriptive statistics. Naturally, there are certain principles to follow to ensure this is done properly, but some mistakes have been made over the years that have proven hard to shake off. The average person is just that.

## A wave of statistics

In the mid-19th century, a wave of statistics washed over the world. The natural sciences had long been successful in using numbers to describe nature, and other branches of science followed suit. As a social parallel to physics, there was a movement to create a *social physics*, now termed sociology, the goal of which was to describe society using numbers. Country after country set up national statistics offices, and with the vast quantity of numerical information about society and its inhabitants, there was a growing need to summarise and present these figures in a meaningful and understandable way.

One of the pioneers in this work was the Belgian mathematician Adolphe Quetelet (1796–1874), with his groundbreaking contributions in statistics and quantitative analysis. In addition to developing the body mass index (BMI) to describe body composition, one of Quetelet's major scientific inventions was *l'homme moyenne* – the average man: a person with average traits (1). The idea was captivating. Instead of rows and columns of numerical information, the summary figures took shape in front of our eyes, manifested as an actual human being. The idea spread, and eventually many countries even gave names to their average person. In Norway, we have Kari and Ola Nordmann, the Swedes have their Medel Svensson, while the British have Jane and John Doe.

## The average man meets resistance

But Quetelet's average man met resistance. It was hard to imagine that a human being with

average characteristics actually existed. The human body is a fine-tuned system, where everything is interconnected, and average numbers for various human traits cannot just be assembled into a new human being. As a principle, the average person quite simply did not make sense.

Nevertheless, the average man lived on. Then, as now, the dissemination of scientific results was considered important, and in 1825, the French mathematician Joseph Diaz Gergonne introduced the 'man in the street' as the final litmus test for explaining a theory of any merit (2). From 1900 onwards, the 'man in the street' was popularised by the German mathematician David Hilbert (1862–1943), and also found his way into American politics, where he assumed the meaning of 'the average voter'. In Norway, the 'man in the street' entered society in full force in connection with the 1972 referendum on EU membership. The idea of one person as a representative of the many took hold once more.

## The sufficiency principle

The question that Quetelet and his contemporaries were poking at is one of the basic questions in statistics. How do we summarise numerical data without losing essential information in the process? It may seem like an impossible question, but when formulated mathematically, it becomes surprisingly manageable. The answer is what is known as the 'sufficiency principle' in mathematical statistics (3). The sufficiency principle gives us valuable insights into what is actually needed in order to summarise data properly.

## The average is not enough

There are several ways to find the middle of a data set. The most widely used numbers are the median, which is the middle number, and the average, which is a mathematical calculation of the geometrical centre of the numbers. Statisticians love the average because of its mathematical properties. It is easier to use in calculations than the median, and it appears naturally in many contexts. However, while the median is often an actual observation, the average is not part of the data set. The average is not typical, it is not common, it is not normal. In fact, the average is so atypical that it does not even exist among the observations.

However, letting one number represent the many, regardless of whether this number is a mathematical construct or a cleverly selected observation, presents a larger and more fundamental problem: it is not enough.

## Only enough is enough

In a standard data set, with a fairly Gaussian distribution and without outliers, the sufficiency principle says that we need *two* summary figures: the average and the standard deviation. If you report only one, you will be withholding essential information about the data set. One average, two averages, one hundred averages, pieced together into an average person is still not enough. In order to properly describe a population, the numerical variation in the data proves to be just as important as the geometrical centre.

It is no wonder people do not see themselves reflected in the average person, or that no one feels that they themselves are 'the man in the street'. This is because Jane and John Doe do not represent them. By focusing on the average person, we lose essential information about the group. At the same time, our understanding of what is typical, common and normal becomes narrower. For normally distributed data, approximately 95 per cent of the observations will be in the interval of the average  $\pm$  two standard deviations. This is an interval that often provides considerable leeway for what is considered normal.

## Popular

Quetelet's average person is still popular. At the time of writing, a Google search for 'average

person' gives close to 15 million hits. That is more than Whitney Houston, Marlon Brando and Sir Isaac Newton. Science killed the average man over 100 years ago because he did not make sense, but the society he was supposed to describe has kept him on life support ever since.

---

REFERENCES:

1. Faerstein E, Winkelstein W. Adolphe quetelet: statistician and more. *Epidemiology* 2012; 23: 762-3. [PubMed][CrossRef]
2. «Mannen i gata er 200 år» [The man in the street celebrates 200 years]  
<https://forskning.no/2016/12/har-funnet-mannen-i-gata/produisert-og-finansiert-av/universitetet-i-agder> (visited 8 November 2018)
3. Birnbaum A. On the foundations of statistical inference. *J Am Stat Assoc* 1962; 57: 269-306. [CrossRef]

---

Published: 14 January 2019. Tidsskr Nor Legeforen. DOI: 10.4045/tidsskr.18.0824

© The Journal of the Norwegian Medical Association 2020. Downloaded from tidsskriftet.no