

# Science, numerology and critical sense

How can numerology become fact? Neither editorial boards, nor readers, should neglect their own expertise and critical sense, even when reference is made to sophisticated statistical distributions with appurtenant eponyms and the material apparently approved by statisticians.

Liv Ariane Augestad  
*l.a.augestad@medisin.uio.no*  
 Mathias Barra  
 Samantha Salvesen Adams

The following is taken from the column *From other journals* in issue no 19/2014 of the Journal of the Norwegian Medical Association (1): «Benford's Law, also referred to as the First-Digit Law, says that the first digit 1–9 for numbers in natural growth processes occurs most frequently for 1 and least frequently for 9. It has long been assumed that births are distributed evenly over the dates of the month, but in an article recently published in the *Tilfeldig gang* journal, issued by the Norwegian Statistical Association, [T. Dønvoid] shows that Benford's Law applies to births as well [(2, 3) ...] «As a rule, more births occur, the lower the digit sum of the date number, i.e. more births occur when the digit sum equals 1 and fewer when this sum equals 9. In turn, this means that more births occur on the 1st 10th, 19th and 28th of each month, and fewer on the 9th, 18th and 27th. [...] The study may help improve roster planning in maternity wards.»

Shortly after the publication of the article, the authors of this piece were gathered at a dinner party, and the conversation quickly turned to this finding. Samantha Salvesen Adams and Liv Ariane Augestad were both confused. They were unfamiliar with Benford's Law, but thought it strange that births should occur more frequently on dates with a low digit sum. Nor did the original article in *Tilfeldig gang* (2) provide any convincing explanation. What kind of mechanism could account for a cyclical distribution of births per day over the course of a month, apparently governed by the decimal system? To the possible chagrin of the other dinner guests, we spent a great deal of time clarifying the nature of Benford's Law, as well as pondering its possible association with births.

## What do we know from before?

Births rates exhibit seasonal variations that are caused by seasonal variations in con-

ception, related to holiday seasons and other public holidays, and most likely also to deadlines for admission to day-care centres (3, 4). There are also reports of systematic variations by days of the week, with fewer births on weekends and on Sundays in particular (5,6). This may be related to social organisation – fewer planned elective Caesarean sections, less intensive perinatal care, such as inducement of births, and possibly also related to the lives of pregnant women during weekends. Could something similar be associated with dates with a low digit sum, i.e. a social phenomenon that may affect the lifestyle of pregnant women or the intensity of perinatal care, linked to the digit sum of the date numbers?

## Benford

Mathias Barra had not read the article, but could contribute his mathematical skills and a certain familiarity with Benford. Could an explanation of Benford's Law provide theoretical support to the claim that births follow the Benford distribution?

In brief, we say that Benford's Law applies to a sequence of numbers if the distribution of first digits follows the so-called Benford distribution. Typically, Benford's Law applies to sequences of measurements, meaning numbers that quantify some kind of magnitude. An example of a sequence of measurements that comply – most likely in an approximate fashion – with Benford's Law is provided by the lengths of all the world's rivers. Measured in kilometres, there will be a majority of short rivers and streams with a length of 1.00–1.99 kilometres. All of these have 1 as their first digit. Somewhat fewer are 2.00–2.99 kilometres in length, and even fewer are 9.00–9.99 kilometres long, with 9 as the first digit in their measurement number. In general, at each interval from and including  $10^n$  to  $10^{n+1}$  km, there is a tendency for rivers of a length within this interval to be more numerous in the first part, where 1 is the first digit.

When using the decimal system, we will see a specific Benford distribution of the first digits – approximately 30 per cent of 1's, falling to approximately 5 per cent of 9's. A single Benford distribution exists for each base,  $b \geq 2$ , and to be able to say

that Benford's Law applies, the first digits must follow the Benford distribution irrespective of the base used to represent the numeraire of the measurement and irrespective of the unit of measurement used. If we instead measure the rivers in feet or inches, choose binary numbers or decimal – this phenomenon will recur because of the underlying phenomenon: that these numbers are measurements of physical objects of which there are more small ones than large ones. Interested readers may turn to Wikipedia for more information (7).

## Is it a good hypothesis?

Even after an intense brainstorming session we were unable to come up with any social phenomenon organised in such a way as to result in higher birth frequencies on dates with higher digit sums. Moreover, Benford's Law refers to the first digits in *measurements*. Date numbers are *not* measurements, they are ordinal numbers and say nothing about the size of the date in question.

Even if the empirical data set should be found to corroborate the hypothesis that births occur more frequently on dates with a low digit sum, Benford's Law would still not apply, unless we also found the same by dividing the week into weeks or hours, or decided to use a base-5 numeral system for the dates of the month. Thus we are not in any case operating within a domain where Benford's Law might be valid. Having established this basic fact, we were inclined to investigate the matter in more detail. The first step was to subject the original article to a closer scrutiny.

## Source criticism

When we telephoned the student journal *Tilfeldig gang*, acting editor Turid Follestad could immediately confirm that theirs was not a peer-reviewed journal, and that none of the editors, nor the Norwegian Statistical Association, could vouch for the content of the article in question (2). The web pages of the student journal gave a link to a longer version of the original article that included the data (8). The link from the web page to the full version disappeared a few days after our contact with the editors of *Tilfeldig gang*.

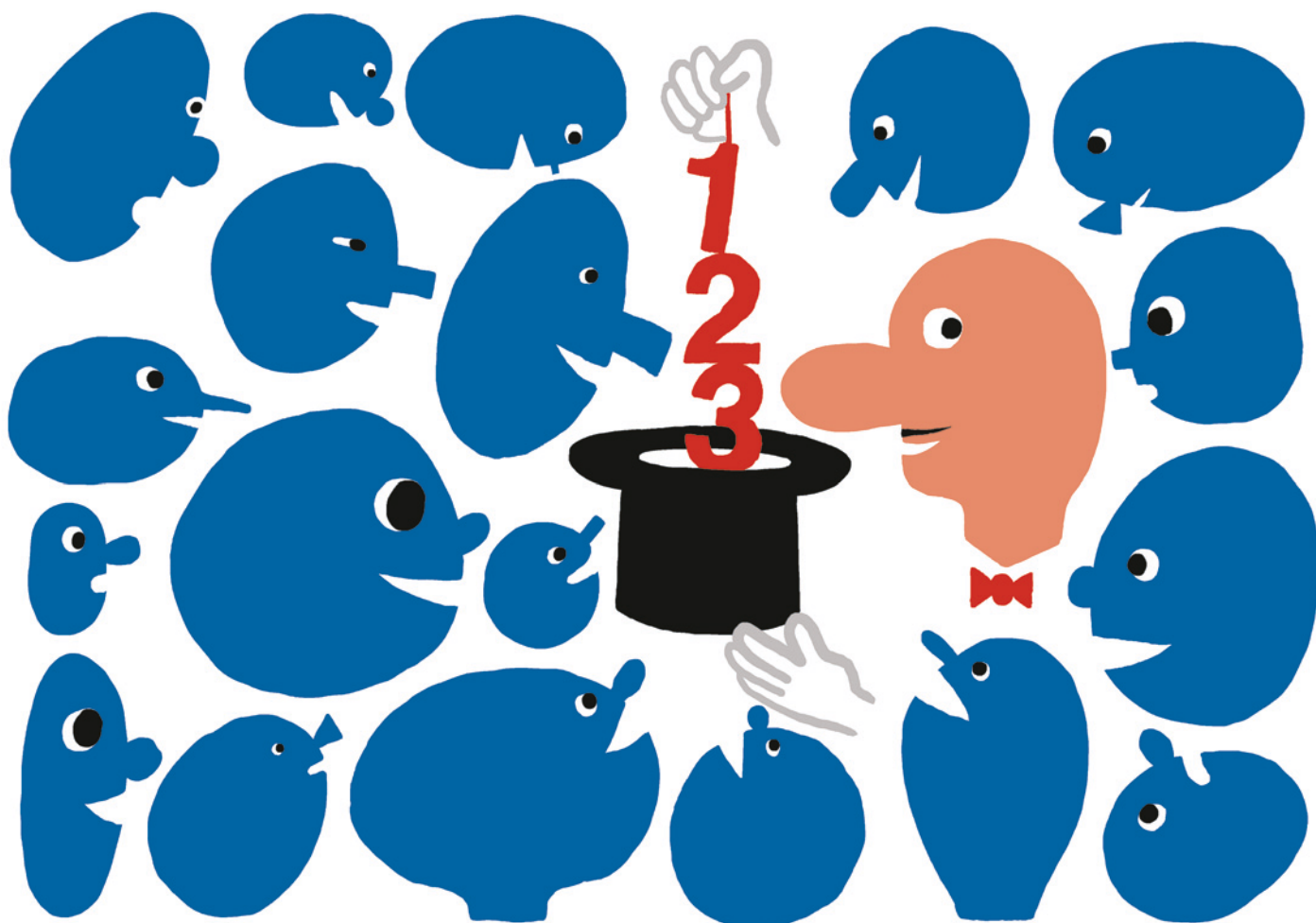


Illustration: Espen Friberg

### Can the data support the conclusion?

The hypothesis in the articles (1, 2, 8) can nevertheless be rejected even without access to such innovative tools for source criticism as the telephone, since *no* valid mathematical argument is provided. The data material in the full version (8) consists of public data on the registered dates of birth of American insurance buyers, and the analyses in the original article in fact provide *no* support to the conclusion that births follow the Benford distribution.

On the contrary, the data presented in the articles do *not* follow the Benford distribution. Simple calculations show that each first digit from 1 to 9 appears with a frequency ranging from 11.3 per cent to 10.9 per cent. This is not exactly a convincing representation of a falling tendency – unless, as in the articles in question, it is graphically represented with a shortened *y*-axis. Any data that fail to fit the conclusion are explained *ad hoc*. The final conclusion that births follow the Benford distribution is therefore directly contrary to the data actually presented.

### Something to learn?

The fact that this «article» is highlighted in the *From other journals* column and thus appears to have been subjected to peer review bears witness to a failure on the part of the editors to comply with routines. Such things happen to the best of us, but such a failure is particularly grave in a matter like this. It is inconceivable to us how any known theory could lead to the hypothesis that births follow the Benford distribution, since the discovery of such an association would have revolutionised our views of what a date number represents.

Science cannot provide answers to everything, and empirical material occasionally lends support to associations that may be hard to explain initially. History provides several examples of this. To us, a hypothesis stating that births will tend to cluster on dates with a low digit sum nevertheless appears to be based more on numerology than on knowledge about births or Benford's Law.

### Peer reviews

In other words, the reported distribution of births was so surprising that the editors ought to have smelled a rat. Simple source criticism would have revealed that this article was not worthy of attention from the Journal of the Norwegian Medical Association. A world of scientific publication that includes a growing number of dubious «peer-reviewed» journals increasingly requires readers to use their critical sense and ability to assess the content of articles for themselves. Even if an article has been published in a prestigious journal, this is no guarantee that it has been subject to peer review: a number of well-established publishing houses have recently withdrawn more than 100 articles because of peer-review fraud (9). Peer review is largely unpaid work, and the quality varies considerably. Even when no direct fraud is involved, the assessment may be beneath contempt and undertaken by incompetent peer reviewers.

Nor are peer reviewers immune to the

dazzling effect of numbers. Including a mathematical formula increases the likelihood that the research will be deemed to be of high quality, even when the formula is unrelated to the research presented (10). There are also certain peer-reviewed journals in which we may have reason to question the very academic paradigm. For example, large publishing houses such as Elsevier are also home to homoeopathy journals, with peer review and an «impact factor» indexed in Medline, and ranked as level 1 journals in this country.

### How important is the hypothesis?

Hypotheses that are tested empirically in the natural sciences in general and particularly in medicine have often been generated as a component of a theory that seeks to explain observations. For example, clinical specialists may establish a hypothesis on the basis of their experience and knowledge of pathophysiology, and this may constitute an appropriate candidate hypothesis given the existing knowledge base. Historically, there have been major practical limitations to the types and numbers of hypotheses that could be experimentally investigated, and there is reason to assume that the selection of hypotheses for testing has concentrated on those that were held to have a high likelihood of being true. With increasing amounts of data, powerful computers and pressure to publish, we may assume that an increasing amount of more or less random testing of correlations is undertaken in the quest for significant p-values. This makes greater demands on the readers' ability to assess how well the researchers justify their testing of a given hypothesis.

### A thought experiment

Imagine that someone tells you that in a large data set it has been found that 30-day survival after gastric surgery increases (the odds ratio or another statistical indicator) if the third letter in the surgeon's first name is «d», such as in Lydia or Andrew. Moreover, the p-value of this effect is reported to be  $p = 0.034$ . Under the null hypothesis that the third letter of the surgeon's first name does not lead to a higher likelihood of survival, «there is (only) a 3.4 % probability of observing a similarly biased or more biased distribution of 30-day survival». Does this mean that there is a 96.6 % probability that the alleged «d-effect» is true? No! This finding says nothing *directly* about the probability that a hypothesis is true or untrue. It needs to be interpreted in light of the confidence in the hypotheses being tested, here exemplified by the null hypothesis, stating that the third letter in the surgeon's first name does *not* have an effect

on 30-day survival – which ought to be a reasonable assumption. This confidence should not be significantly undermined, even by a so-called significant p-value.

When Italian researchers in 2011 «discovered» particles that moved faster than the speed of light, they asked others to help them find the error. Under the null hypothesis «the theory of relativity is true», the experiments had a p-value of 0.000002. A major international collaboration was initiated, and several months later a systematic measurement error was discovered, meaning that Einstein's theory of relativity remains valid (11).

What if analyses of an empirical data set in fact showed a clustering of births on days with a low digit sum, and that the null hypothesis of an even distribution over date numbers had a p-value lower than 0.05? If we have a well-justified (null) hypothesis saying that date numbers are unrelated to birth rates, we should not reject the null hypothesis. Regarding the empirical data material as reflecting a random observation of an underlying even distribution is more likely than assuming an association between date numbers and birth rates. The more theory there is to indicate that the null hypothesis is correct, the stronger the statistical evidence needed to disprove it.

### When the scepticism alarm goes off

For an inquisitive human being, few things are more intriguing than new and revolutionary knowledge. Research requires inquisitive researchers and frequently a pinch of creativity. However, good research also requires familiarity with established knowledge and what can be termed common sense. As a general rule, we may assert that paradigm-shifting conclusions in scientific work ought to be backed by very convincing evidence. Even the editors and readers of journals have the opportunity to consult their own critical sense and expertise when assessing new results.

If the scepticism alarm goes off, we recommend journal editors in particular to undertake further investigations. «Keep an open mind, but not so open that your brain falls out» (12).

### Liv Ariane Augestad (born 1980)

PhD and post-doctoral scholar at the Department of Health Management and Health Economics, University of Oslo, and South-Eastern Norway Health Authority Competence Centre for Health Research at Akershus University Hospital.

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

### Mathias Barra (born 1977)

PhD in mathematical logic and researcher. He works on mathematical modelling of patient pathways, health economics, statistics and life-quality research at the South-Eastern Norway Health Authority Competence Centre for Health Research at Akershus University Hospital. The author has completed the ICMJE form and declares no conflicts of interest.

### Samantha Salvesen Adams (born 1981)

PhD and locum GP at Frysja Health Centre. The author has completed the ICMJE form and declares no conflicts of interest.

### References

1. Dønvold T. Sesongjusterte fødsler er ikke jevnt fordelt på datoer. Tidsskr Nor Legeforen 2014; 134: 1834.
2. Dønvold T. Benfords lov i fødselsdata. 2014. Tilfeldig gang. <https://sites.google.com/site/statistiskforening/tilfeldig-gang> [31.8. 2014].
3. MacFarlane A. Variations in number of births and perinatal mortality by day of week in England and Wales. *BMJ* 1978; 2: 1670–3.
4. Brunborg H. Fødsler i jule- og nyttårshelgen: Få fødsler, mange unnfanger. SSB 2001. <http://ssb.no/befolkning/artikler-og-publikasjoner/fa-fodsler-mange-unnfanger> [20.11. 2014].
5. Gam CM, Tanniou J, Keiding N et al. A model for the distribution of daily number of births in obstetric clinics based on a descriptive retrospective study. *BMJ Open* 2013; 3: e002920.
6. Barra M, Lindstrøm J, Adams S et al. Sesongjusterte fødselsfrekvenser er poissonfordelte. *Tidsskr Nor Legeforen* 2015; 135.
7. Benford's law. [http://en.wikipedia.org/wiki/Benford's\\_law](http://en.wikipedia.org/wiki/Benford's_law) [1.6. 2015].
8. Dønvold T. Sesongjusterte fødsler er ikke uniformt fordelt. *Tilfeldig gang* 2015.
9. Ferguson C, Marcus A, Oransky I. Publishing: The peer-review scam. *Nature* 2014; 515: 480–2.
10. Eriksson K. The nonsense math effect. *Judgm Decis Mak* 2012; 7: 746–9.
11. Pressemeldinger fra CERN. OPERA experiment reports anomaly in flight time of neutrinos from CERN to Gran Sasso. <http://press.web.cern.ch/press-releases/2011/09/opera-experiment-reports-anomaly-flight-time-neutrinos-cern-gran-sasso> [30.5. 2015].
12. Quote investigator. <http://quoteinvestigator.com/2014/04/13/open-mind/> [30.5. 2015].

Received 1 July 2015, accepted 7 October 2015.

Editor: Hanne Støre Valeur.