

# BMJ Open Pinocchio testing in the forensic analysis of waiting lists: using public waiting list data from Finland and Spain for testing Newcomb-Benford's Law

Jaime Pinilla,<sup>1</sup> Beatriz G López-Valcárcel,<sup>1</sup> Christian González-Martel,<sup>1</sup> Salvador Peiro<sup>2</sup>

**To cite:** Pinilla J, López-Valcárcel BG, González-Martel C, *et al.* Pinocchio testing in the forensic analysis of waiting lists: using public waiting list data from Finland and Spain for testing Newcomb-Benford's Law. *BMJ Open* 2018;8:e022079. doi:10.1136/bmjopen-2018-022079

► Prepublication history and additional material for this paper are available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2018-022079>).

Received 1 February 2018  
Revised 10 April 2018  
Accepted 11 April 2018



<sup>1</sup>Departamento de Métodos Cuantitativos en Economía y Gestión, Universidad de Las Palmas de Gran Canaria - Campus de Tafira, Las Palmas de Gran Canaria, Spain

<sup>2</sup>Fundación para el Fomento de la Investigación Sanitaria y Biomédica de la Comunidad Valenciana (FISABIO), Red de Investigación en Servicios de Salud en Enfermedades Crónicas (REDISSEC), València, Spain

## Correspondence to

Dr Jaime Pinilla;  
[jaime.pinilla@ulpgc.es](mailto:jaime.pinilla@ulpgc.es)

## ABSTRACT

**Objective** Newcomb-Benford's Law (NBL) proposes a regular distribution for first digits, second digits and digit combinations applicable to many different naturally occurring sources of data. Testing deviations from NBL is used in many datasets as a screening tool for identifying data trustworthiness problems. This study aims to compare public available waiting lists (WL) data from Finland and Spain for testing NBL as an instrument to flag up potential manipulation in WLs.

**Design** Analysis of the frequency of Finnish and Spanish WLs first digits to determine if their distribution is similar to the pattern documented by NBL. Deviations from the expected first digit frequency were analysed using Pearson's  $\chi^2$ , mean absolute deviation and Kuiper tests.

**Setting/participants** Publicly available WL data from Finland and Spain, two countries with universal health insurance and National Health Systems but characterised by different levels of transparency and good governance standards.

**Main outcome measures** Adjustment of the observed distribution of the numbers reported in Finnish and Spanish WL data to the expected distribution according to NBL.

**Results** WL data reported by the Finnish health system fits first digit NBL according to all statistical tests used ( $p=0.6519$  in  $\chi^2$  test). For Spanish data, this hypothesis was rejected in all tests ( $p<0.0001$  in  $\chi^2$  test).

**Conclusions** Testing deviations from NBL distribution can be a useful tool to identify problems with WL data trustworthiness and signalling the need for further testing.

## INTRODUCTION

Waiting lists (WLs) have been recognised as an inescapable side effect of public National Health Services<sup>1</sup> and recently the performance of health systems has been evaluated by comparing WL indicators.<sup>2</sup> Healthcare systems face extreme difficulties when managing WLs due to inherent issues related to their origin and maintenance. Reaching conclusions on the importance of WLs from their volume is not an easy task because reliable data are often lacking or hidden,

## Strengths and limitations of this study

- Frequent contradictions and conflicts of interests occur between different actors actively involved in waiting list management.
- Statistical tools can be used to verify the compliance of waiting list data.
- The Newcomb-Benford's Law seems to be a valid tool for screening the trustworthiness of public waiting list data and for signalling the need of further analysis.
- The method proposed in this article can be applied to other healthcare data.

operational definitions are not standard and change frequently, and information on the severity of a patient's condition is barely included.<sup>2</sup> Administrative updates and the standardisation of WLs, although necessary, are often associated with a 'successful' decrease in WLs size, which is not necessarily associated with productivity improvements or better management, especially when this decrease is achieved without ensuring that the patient no longer requires the intervention he/she was waiting for.<sup>3</sup>

In countries with publicly funded National Health Systems, large WLs erode the confidence of citizens in the health system, its leaders and its professionals, and WLs are commonly used as an element of political confrontation. Because WLs have become a relevant issue, both from a social and political point of view, and the numbers describing a diminishing WL are taken as a noticeable signal of 'success' in healthcare policy, WL management may be more focused on offering appealing short-term 'numbers' than developing coherent long-term solutions consistent with the complex nature of the WL problem. So there is a potential temptation to play with these numbers, as documented in the past.<sup>4,5</sup>

Unlike Pinocchio, most liars do not provide telltale signs that they are being dishonest, so there is a need for methods to detect manipulation in WL data and to distinguish between accurate data and data that are false or which are omitting information. Originally described by Newcomb,<sup>6</sup> in 1938 a physicist called Benford rediscovered a remarkable empirical phenomenon<sup>7</sup>: for an extensive collection of heterogeneous numerical data expressed in decimal form, the frequency of numbers which have  $d$  as the first significant digit, with  $d=1, 2, \dots, 9$ , is not  $1/9$  (11.1%), as one would expect, but is approximately equal to  $\log_{10} \left[ \left( \frac{d+1}{d} \right) \right]$ . According to what is today referred to as Newcomb-Benford's Law (NBL; or Benford's Law or First digit law), the probabilities that the leading digits will be 1, 2 or 3 account for >60% of the total probability distribution. Although this law does not apply to datasets of truly random numbers (eg, lottery), sequential numbers, assigned numbers (eg, zip codes) and numerical series with some restrictions, many real-world datasets do conform to an NBL first digit distribution. Knowing that the frequencies of the first significant digits should fall off in a particular way, suspicious data may be statistically tested against this empirical NBL distribution to evaluate their reliability,<sup>7</sup> providing a solid basis for screening the trustworthiness of large amount of data, which might have been manipulated.

Over many years, the analysis of the digital frequency in datasets has emerged as a powerful tool for detecting data irregularities in tax audit—so-called forensic auditing—to detect financial accounting manipulation,<sup>8,9</sup> and large auditing and consulting firms increasingly use this analysis in their fight against financial fraud.<sup>10,11</sup> NBL even had a moment of fame with the finding that the Greek Government had been presenting 'invented' economic data to the European Union (EU).<sup>12</sup> In the medical sector, NBL has been applied to clinical questionnaire data,<sup>13</sup> to test falsification of interview data,<sup>14</sup> to evaluate the performance of public health surveillance systems during epidemics,<sup>15</sup> the accuracy of cancer incidence in cancer registries<sup>16</sup> and to recognise fraud in scientific medical publications<sup>17,18</sup> but, to the best of our knowledge and although administrative data on WLs are similar in nature to financial accounting, so far no one has used NBL as a 'stress' test for the reliability of WL data in publicly funded National Health Systems.

The aim of this study is to compare public WL data corresponding to Finland and Spain, both countries with National Health Systems but characterised by different levels of transparency and good governance standards, for testing NBL as an instrument to screen for irregularities in WL data.

## METHODS

### Study design and conceptual framework

According to NBL, for an extensive collection of heterogeneous numerical data expressed in decimal form, the frequency of numerical data with the first significant

digit equal to 1 appeared to be about 30%, and equal to 1, 2 or 3, about 60%. For the second and further digits, NBL predicts a more uniform distribution.<sup>19</sup> Our analysis examines Spanish and Finnish WL data, specifically the frequency of numbers appearing in first position, and compares it with the expected pattern following NBL distribution.

## SETTING

WL data for publicly funded elective treatments is not equally accessible and transparent in all countries. Finland first published data on WLs in 1993. In 1996, the Finnish government recognised by law patients' rights to know WL times and the possibility of complaining in the event of not being satisfied with them, and establishing specific maximum times for accessing healthcare interventions in primary and secondary care. In 2005, the government introduced a new National Healthcare Guarantee system into Finnish law. Currently in Finland, the national monitoring of queues, waiting times and hospital productivity has been intensified and quality has improved during the past decade. This has given the supervisory and other bodies information to act on in order to live up to the Healthcare Guarantee.<sup>20</sup>

In December 2003, after a formal requirement by the Ombudsman,<sup>21</sup> Spain started publishing homogeneous data collected nationwide focusing on the number of patients and the duration of WLs for surgical procedures. These data included information from only 14 different health services belonging to the 17 Spanish Autonomous Communities (ACs) with Regional Healthcare Services. Data published by the Ministry of Health did not include information on all 17 ACs until June 2012. Each AC has responsibilities in health planning, information systems and service delivery, but they do not always provide information on WLs to the Spanish Ministry of Health central authority. As a consequence, the available information can be considered as a proxy to the challenge posed by WLs, but not state of the art or definitive.<sup>22</sup> Although most ACs do provide the necessary information to the Ministry, there is no agreement to make that disaggregated information publicly available. Some regional data are available on the websites of some ACs, but these data are not generally homogeneous.<sup>23</sup>

## Data sources

We used publicly available data on WLs collected from the Ministry of Health, Social Policy and Equality in Spain<sup>24</sup> and the National Institute for Health and Welfare in Finland,<sup>25</sup> respectively. Both databases provide official information on the periodic evolution of WLs for surgery and for outpatient visits to medical and surgical specialties. 'Patients waiting' refers to the number of patients waiting for their first visit to an outpatient specialist or to an elective surgical intervention on the date of reference. In Spain, the frequency of reporting data is biannual (June and December) while for Finnish data it is

triannually (April, August and December). The dataset consists of a total of 764 total valid observations for Spain (35 lists, 19–23 time periods), and 594 valid observations (14 registers with 0 patients waiting were eliminated) for Finland (38 lists, 16 time periods), from December 2003 to December 2015 and December 2007 to December 2012, respectively. Data are accessible as an online supplementary appendix file.

### Statistical analysis

The data were automatically screened for leading digits and numbers. Extracted numbers were transferred to a comma-separated values spreadsheet file and the occurrence of each number was determined using the R Package *benford.analysis*.<sup>26</sup>

We used three different statistical tests to determine whether the distribution of the first and higher order digits conformed to NBL: the Pearson's  $\chi^2$  test, the mean absolute deviation (MAD) test and the Kuiper test, with the null hypothesis that data would follow the Benford distribution. Pearson's  $\chi^2$  test is a natural candidate for testing whether an observed sample satisfies NBL. The statistical test for first digit is defined as  $\chi^2 = \sum_{i=1}^9 \frac{(O_i - E_i)^2}{E_i}$ , where  $O_i$  and  $E_i$  are the observed and expected absolute frequencies for digit  $i$ , respectively. Under  $H_0$ , the statistic follows a  $\chi^2$  distribution with 8 df. For specific digits, the standard normal statistic  $Z_i = \frac{\sqrt{N} |O_i - E_i| - \left(\frac{1}{2N}\right)}{\sqrt{(E_i)(1 - E_i)}}$  can be used to check whether the observed frequency deviates significantly from its theoretical value.

Because the  $\chi^2$  test is very sensitive to sample size, having enormous power with large  $N$  and low power for moderately small sample sizes, we also used the MAD test that ignores sample size. The MAD statistic is calculated as  $\frac{\sum_{i=1}^9 |O_i - E_i|}{9}$ . Finally, a third alternative is the Kuiper test, a modification of the Kolmogorov-Smirnov test. The Kuiper's test is calculated as  $T_k = (D_N^+ + D_N^-) [\sqrt{N} + 0.155 + 0.24/\sqrt{N}]$ , where  $D_N^+ = \sup [F(O_i) - F(E_i)]$  and  $D_N^- = \sup [F(E_i) - F(O_i)]$ , and  $F(\cdot)$  stands for cumulated relative frequencies.<sup>27</sup>

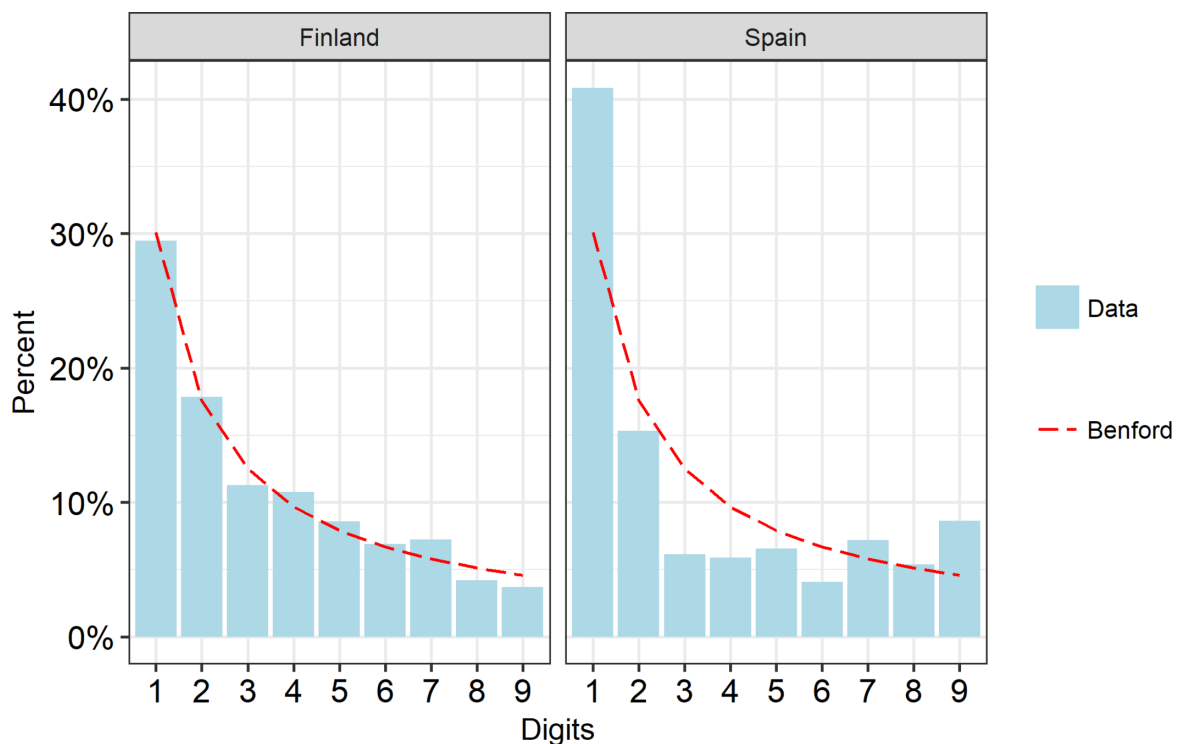
Finally, we tested if NBL applies to our datasets using criteria suggested by Miller<sup>28</sup> and Wallace,<sup>29</sup> and we further repeated the fitness NBL analysis but using only data for the same point in time (December) for both Finnish and Spanish datasets (see online supplementary annex 1).

### Patient involvement

This study uses publicly available data sources and did not include patients as study participants. No patients were involved in setting the research question, the study design or the overall conduct of the study. There are no plans to involve patients in the dissemination of study findings.

### RESULTS

Figure 1 compares the overall frequency distributions of the first significant digit for WL data in Finland and Spain against the expected NBL frequency distribution. On visual inspection, Finnish data seem to be satisfactorily adjusted to an NBL distribution, while Spanish data do not seem to satisfy the NBL distribution at all.



**Figure 1** Theoretical (line) and observed distributions (columns) of first digit for Finnish and Spanish waiting list data.

**Table 1** Test statistics for the first digits of Finnish data

Value	Count	Frequency observed	Frequency expected (Benford's Law)	Diff. (MAD)	P values of Z-test for each digit
1	175	0.29461	0.30103	-0.00642	0.7544
2	106	0.17845	0.17609	0.00236	0.8717
3	67	0.11279	0.12494	-0.01214	0.4196
4	64	0.10774	0.09691	0.01083	0.3671
5	51	0.08586	0.07918	0.00660	0.5429
6	41	0.06902	0.06695	0.00208	0.8055
7	43	0.07239	0.05799	0.01440	0.1352
8	25	0.04209	0.05115	-0.00906	0.3521
9	22	0.03704	0.04576	-0.00872	0.3757
Total	594				

Pearson's  $\chi^2$  test: 5.9584 ( $p=0.6519$ ); mean test (absolute value): 0.8077; Kuiper test: 0.8338. All p values are non-significant at the 1% level. The respective critical test values for the 5% and 1% significance levels are: Pearson's  $\chi^2$  test (8 df): 15.51 and 20.09; mean test: 1.96 and 2.58; Kuiper test: 1.75 and 2.00. MAD, mean absolute deviation.

Tables 1 and 2 show the results of the statistical fitness of the Finnish and Spanish datasets to Benford's Law. These results are consistent with the previous graph, showing that for Finnish data it is not possible to reject the null hypothesis that they follow the Benford distribution. Nevertheless, for Spanish data, the null hypothesis was rejected in all tests.

## DISCUSSION

Our results basically show that the WL data from Finland follow the NBL, while the data from Spain do not, raising suspicion about their trustworthiness. This example illustrates how statistical tools can be used to verify the compliance of WL data with a regularity law applicable to many different naturally occurring sources of data.

Frequent contradictions and conflicts of interests occur between different actors actively involved in WL management. The public and the media consider regularly published data on WLs as the quintessence of healthcare policy success or failure. Therefore, to respond to these expectations, data manipulation is a temptation for both policymakers and managers. The two countries chosen for the present illustration are perceived as remarkably different in their behaviour. Finland has been consistently classified at the top of the international ranking of transparency and good governance, social control of the political class and more likely to enforce penalties in the case of irregularities than other countries. Finland was ranked fourth in the world in the World Justice Project (WJP) Rule of Law Index 2015,<sup>30</sup> and was evaluated as

**Table 2** Test statistics for the first digits of Spanish data

Value	Count	Frequency observed	Frequency expected (Benford's Law)	Diff. (MAD)	P values of Z-test for each digit
1	312	0.40838	0.30103	0.10735	0.0000**
2	117	0.15314	0.17609	-0.02295	0.0966
3	47	0.06152	0.12494	-0.06342	0.0000**
4	45	0.05890	0.09691	-0.03801	0.0002**
5	50	0.06545	0.07918	-0.01374	0.1798
6	31	0.04058	0.06695	-0.02637	0.0023**
7	55	0.07199	0.05799	0.01400	0.1035
8	41	0.05366	0.05115	0.00251	0.7422
9	66	0.08639	0.04576	0.04063	0.0000**
Total	764				

Pearson's  $\chi^2$  test: 107.511\*\* ( $p>0.0001$ ); mean test (absolute value): 3.6553\*\*; Kuiper test (absolute value): 4.5732\*\*. \*\*Significant test value on the 1% level.

The respective critical test values for the 5% and 1% significance levels are: Pearson's  $\chi^2$  test (8 df): 15.51 and 20.09; mean test: 1.96 and 2.58; Kuiper test: 1.75 and 2.00. MAD, mean absolute deviation.

the most efficient country in producing public services of high quality at moderate cost.<sup>31</sup> Moreover, it is well known internationally that the publicly organised Finnish healthcare service system has been a success story.<sup>32</sup> Spain came in 17th position (out of 24 neighbouring countries) in the latest WJP Rule of Law Index. In 2015, it was distinguished for its high level of corruption,<sup>33</sup> which seems to be worsening, surpassed only by Italy and Greece in the 15 EU countries before the Eastern enlargement. In Spain, the publication of WL data was introduced in 2002 by an Ombudsman mandate, which reported that data were sparse, broken and sometimes not very truthful.<sup>22</sup> Fifteen years have gone by since then, but even now the Ministry of Health recognises the limitations and lack of rigour of such data.

Neither in the case of Spain nor in any others, does failure to obey NBL necessarily provide evidence that WL data are inaccurate or have been manipulated. The NBL universal empirical distribution provides a tool to check data quality in the sense of data accuracy, which denotes the closeness of computations or estimates to the exact or true values. If real values following the NBL are replaced with fabricated numbers, the result is typically a deviation from NBL. The fabrication of numbers may not necessarily be an act of deliberate manipulation; even rounding up can cause a deviation from Benford's Law. Thus, a deviation from the Benford distribution does not provide conclusive proof of manipulation, just as conformity does not prove the cleanliness of the data. Rather, non-conformity should be seen as a signal flagging up data that need closer inspection and further testing. Benford's Law could thus be used in addition to existing control mechanisms as a first step in checking the possible manipulation of data.

Among the limitations of this study it should be first noted that we only analyse the first WL digit. Adding second-order digit and first-two digit combination tests in a forensic analysis is an essential part of a thorough forensic examination. This should be done as a separate test apart from the first-digit test. An exception is made when the dataset under consideration contain too few values, in which case only a first-digit test is performed. As suggested, an empirical threshold in this context is established to avoid second-order digit test and first-two digit combination tests for any dataset having fewer than 1000 records,<sup>34</sup> as is our case. There is no formal statistical theory capable of giving significant threshold points for applicable sizes; rather the above suggestions are subjective judgements derived from experiences in dealing with datasets and forensic digital analysis.

Second, the published WLs could, as with some restricted series, not adjust to the NBL distribution, but in the preliminary analyses carried out (see online supplementary annex 1) the two lists analysed seem to meet the requirements needed<sup>28 29</sup> to follow this distribution. Third, datasets from Finland and Spain have some differences (size, time of collection) that could influence their distribution, although the analysis carried out using

only the lists collected in December was consistent with the overall results (see online supplementary annex 1). Finally, it should be noted that both countries use (legitimate) administrative mechanisms to 'clean' WLs of deceased persons, people already operated or that no longer want to have surgery, people who could not be reached for appointments or people out of coverage. Also the entries into, and the exits from WLs could have some seasonal variability. These factors clearly influence the number of people waiting, but should not influence the distribution of the first digits, nor the adjustment to the NBL distribution. On the contrary, fabricated data hardly will conform NBL distribution.<sup>35</sup>

The method proposed in this article can be applied to other healthcare data, as long as control mechanisms or alarm signals for intensifying efforts to monitor and control the clinical and economic information of health centres are in place. Other areas where NBL testing would be interesting are reporting systems for adverse events, files on professional activity, operating theatre times, length of stay statistics and research data-files (clinical trials, observational studies and analogous data).

**Contributors** JPD, BGL-V and SP conceptualised and designed the study. JPD and CG-M oversaw data collection and performed the statistical analysis. JPD drafted the manuscript and all coauthors participated equally in the revision and final approval of the manuscript. JPD is the guarantor for the study. All authors had full access to all the data in the study (including statistical analysis, tables and figures) and can take responsibility for the integrity of the data and the accuracy of the data analysis.

**Funding** This paper forms part of the research funded by Grant ECO2017-83771-C3-2 under the 'National Programme for Research, Development and Innovation to Address the Challenges of Society: National Plan for Scientific Research and Technical Innovation 2017–2020' funded by the Ministry of Economy and Competitiveness of Spain.

**Disclaimer** The funder had no influence on the conduct of this study or on the drafting of this manuscript.

**Competing interests** None declared.

**Patient consent** Not required.

**Ethics approval** The study, with secondary data available from public sources and no patients as participants, does not require ethics approval according Spanish law and international regulations.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data sharing statement** The full dataset (Dataset.csv) is available without restrictions in the Appendix.

**Open Access** This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

© Article author(s) (or their employer(s) unless otherwise stated in the text of the article) 2018. All rights reserved. No commercial use is permitted unless otherwise expressly granted.

## REFERENCES

1. Martin RM, Sterne JA, Gunnell D, *et al.* NHS waiting lists and evidence of national or local failure: analysis of health service data. *BMJ* 2003;326:188–92.

2. Siciliani L, Borowitz M, Moran V, eds. *Waiting time policies in the health sector: what works? OECD Health Policy Studies*. Paris: OECD Publishing, 2013.
3. Radical Statistics Health Group. NHS "indicators of success": what do they tell us? *BMJ* 1995;310:1045–50.
4. Yates J, Lies YJ. Lies, damned lies, and waiting lists. *BMJ* 1991;303:802.
5. Ferriman A. Trusts fail to discipline those who manipulate waiting lists. *BMJ* 2002;325:614b–614.
6. Newcomb S. Note on the frequency of use of the different digits in natural numbers. *American Journal of Mathematics* 1881;4:39–40.
7. Benford F. The Law of Anomalous Numbers. *Proc Amer Phil Soc* 1938;78:551–72.
8. Nigrini MJ. A taxpayer compliance application of Benford's Law: tests and statistics for auditors. *J Am Taxation Assoc* 1996;18:72–91.
9. Nigrini MJ. An assessment of the change in the incidence of earnings management around the Enron-Andersen episode. *Review of Accounting and Finance* 2005;4:92–110.
10. Kumar K, Bhattacharya S. Benford's law and its application in financial fraud detection. *Adv Financial Plann Forecast* 2003;11:57–70.
11. Bhattacharya S, Kumar K, Smarandache F. Conditional probability of actually detecting a financial fraud – a neutrosophic extension to Benford's law. *Int J Appl Math* 2005;17:7–14.
12. Rauch B, Götsche M, Brähler G, et al. Fact and fiction in EU-Governmental economic data. *German Economic Review* 2011;12:243–55.
13. Taylor RN, McEntegart DJ, Stillman EC. Statistical techniques to detect fraud and other data irregularities in clinical questionnaire data. *Drug Inf J* 2002;36:115–25.
14. Bredl S, Winker P, Kötschau K. A statistical approach to detect interviewer falsification of survey data. *Surv Methodol* 2012;38:1–10.
15. Idrovo AJ, Fernández-Niño JA, Bojórquez-Chapela I, et al. Performance of public health surveillance systems during the influenza A(H1N1) pandemic in the Americas: testing a new method based on Benford's Law. *Epidemiol Infect* 2011;139:1827–34.
16. Crocetti E, Randi G. Using the Benford's Law as a first step to assess the quality of the cancer registry data. *Front Public Health* 2016;4:225.
17. Hein J, Zobrist R, Konrad C, et al. Scientific fraud in 20 falsified anesthesia papers: detection using financial auditing methods. *Anaesthesist* 2012;61:543–9.
18. Hüllemann S, Schüpfer G, Mauch J. Application of Benford's law: a valuable tool for detecting scientific papers with fabricated data? : A case study using proven falsified articles against a comparison group. *Anaesthesist* 2017;66:795–802.
19. Hill TP. A statistical derivation of the significant-digit Law. *Statistical Science* 1995;10:354–63.
20. Jonsson PM, Hakkinen P, Jarvelin J, et al. In: Siciliani L, Borowitz M, Moran V, eds. *Waiting time policies in the health sector: what works?* Paris: OECD Publishing, 2013:133–46. (accessed on 12 Jan 2018).
21. Múgica E. Oficina del Defensor del Pueblo. Listas de espera en el Sistema Nacional de Salud. Diario de sesiones de las Cortes Generales. 2002;119 [http://www.congreso.es/public\\_oficiales/L7/CORT/DS/CM/CM\\_119.PDF](http://www.congreso.es/public_oficiales/L7/CORT/DS/CM/CM_119.PDF)
22. García-Goñi M, Spain C-FJ. In: Siciliani L, Borowitz M, Moran V, eds. *Waiting time policies in the health sector: what works?* Paris: OECD Publishing, 2013:263–74. (accessed 12 Jan 2018).
23. Peiró S, Ridaio M. *Experiencias autonómicas en la gestión de las listas de espera. Report to the Agencia de Calidad del Sistema Nacional de Salud, Ministerio de Sanidad y Consumo*. Madrid: Fundación Instituto de Investigación en Servicios de Salud, 2004. (accessed 12 Jan 2018).
24. Ministerio de Sanidad. *Sistema de información sobre listas de espera en el Sistema Nacional de Salud. Agencia de Calidad del Sistema Nacional de Salud*. Madrid: Instituto de Información Sanitaria, 2015. (accessed 28 Jun 2016).
25. National Institute for Health and Welfare, Finland. Trends in access to specialised health care 2007–2012, Statistical report 25/2013. 2013. <https://www.thl.fi/en/web/thlfi-en/statistics/statistics-by-topic/specialised-health-care-services/trends-in-access-to-specialised-health-care> (accessed 28 Jun 2016).
26. Cinelli C. Benford analysis for data validation and Forensic analytics. R package benford.analysis. <https://cran.rproject.org/web/packages/benford.analysis/index.html> (accessed 12 Jan 2017).
27. Stephens MA. Use of the Kolmogorov-Smirnov, Cramer-Von Mises and related statistics without extensive tables. *J R Statist Soc B* 1970;32:115–22.
28. Miller SJ. *Benford's Law: theory and applications*. Princeton University Press: Princeton, NJ, 2015.
29. Wallace WA. Assessing the quality of data used for benchmarking and decision making. *J Gov Financ Manag* 2002;51:16–22.
30. World Justice Project. Rule of Law Index. 2015. <http://data.worldjusticeproject.org>
31. Boyle R. *Measuring public sector productivity: lessons from international experience (CPMR Discussion Paper 35)*. Dublin: Institute of Public Administration, 2006. (accessed 12 Jan 2018).
32. Teperi J, Porter ME, Vuorenkoski L, et al. *The Finnish health care system: a value-based perspective (Sitra reports 82)*. Helsinki: Sitra, 2009. (accessed 12 Jan 2018).
33. Transparency International. Corruption perceptions index. 2015 <https://www.transparency.org/cpi2015#results-table>.
34. Kossovsky AE. *Bedford's Law: theory, the general law of relative quantities, and forensic fraud detection applications*. Singapore: World Sci Pub Co Inc, 2014.
35. Gauvrit NG, Houillon JC, Delahaye JP. Generalized Benford's Law as a lie detector. *Adv Cogn Psychol* 2017;13:121–7.