Terminal digit preference and the accuracy of breast cancer diameter reporting based on Benford’s law

Shahzaib Ahmad a,*, Amber Latif b, Mehwish Mehmoond c, Ramisha Aslam a, Zain Ul Abiddin a, Hassan Mumtaz d,e, Khadija Ahmed f, Waqas Mehdi a,g, Waheeda Begum h

a King Edward Medical University Lahore/Mayo Hospital Lahore, Pakistan
b KRL Hospital, Pakistan
c Fauji Foundation Hospital, Pakistan
d Forensic & Toxicology, Riphah International University, Rawalpindi, Pakistan
e Master’s in Public Health: Health Services Academy, Islamabad, Pakistan
f Doctor of Physical Therapy (DPT), Foundation University Islamabad (Fui), Pakistan
g King Edward Medical University Lahore, Pakistan
h Army Medical College, Pakistan

ABSTRACT

Background: Breast cancer is the most frequent cancer in women all over the world, and it is one of the leading causes of cancer-related deaths in women. A pathologist’s partiality for the last digit of a patient’s name can lead to errors in the measurement of malignancies. This means that, rather than recording the exact measurement of a tumor, a pathologist might round it off to his preferred terminal digit.

Methods: It is a retrospective cross-sectional study in which data on primary tumor resection for 1000 breast cancer patients was obtained from KRL Hospital’s patient directory from November 2016 to December 2020. The tumors were measured in cm to one decimal point along their longest dimension. Ki-67 markers were used to categorize the tumors into nine categories. Terminal digit preference was evaluated using Benford’s law.

Results: The recording of the Ki-67 index revealed evidence of pentameric preference. The numbers three, five, and six appeared more frequently in the histogram of the Ki-67 index distribution measured in percentage. The frequency of nine has dropped dramatically. However, the influence of tumor size terminal digits on Ki-67 staining scores (low proliferative vs high proliferative) assessed using the Mann–Whitney U Test demonstrated that tumor size terminal digits had no significant effect on Ki-67 staining scores (p = 0.114).

Conclusion: The Ki-67 index shows evidence of pentameric preference for digits three, five, and six. The frequency of nine has dropped dramatically. The influence of tumor size on terminal digits on staining scores (low proliferative vs. high proliferative) was assessed using the Mann–Whitney U Test.

1. Introduction

Breast cancer represents the most prevalent form of cancer in women all over the world, making it one of the leading causes of cancer-related deaths in women [1]. However, considering recent advances in the field of research, the prognosis and treatment plan for breast cancer have changed dramatically. Hence, physicians must choose from a wide variety of treatment options by the condition of the patient. TNM staging is often used to classify a breast tumor, which in turn determines the management of the patient [3]. The TNM staging system is based on the size of the tumor and its extent of dissemination along with other factors [4].

Hence, the size of a tumor as well as its cytology have a major impact on its staging and subsequently guide the physician in its management. However, errors can be made during the measurement of these tumors. One of these errors is the terminal digit preference of the pathologist. This means that instead of recording the exact measurement of a tumor, a pathologist might round off the measurement to his preferred terminal digit, usually 0 or 5. This error has been documented in previous studies for different kinds of carcinomas [5]. In the case of breast carcinoma, however, more research is required. The analysis of breast cancer 0a4re0ents established terminal digital preference utilizing histograms in a previous investigation [6]. This terminal digital preference can also occur during a sample’s Ki-67 indexing. The Ki-67 index

* Corresponding author.
E-mail addresses: shahzaib.ahmad@kemu.edu.pk (S. Ahmad), amberlatif033@gmail.com (A. Latif), mehwishmehmood62@gmail.com (M. Mehmoond), Ramishaaslam17@gmail.com (R. Aslam), Zawan031@gmail.com (Z.U. Abiddin), Hassanmumtaz.dr@gmail.com (H. Mumtaz), khadijaahmed907@gmail.com (K. Ahmed), Dr.waqasmehdi@gmail.com (W. Mehdi), Waheedaamin1996@gmail.com (W. Begum).

https://doi.org/10.1016/j.amsu.2022.103993
Received 21 April 2022; Received in revised form 8 June 2022; Accepted 12 June 2022
Available online 16 June 2022
2049-0801/© 2022 The Authors. Published by Elsevier Ltd on behalf of IJS Publishing Group Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
assesses the expression of the Ki-67 marker on cancer cells, which indicates their proliferative ability [7].

This study aims to evaluate terminal digital preference among pathologists using Benford’s Law. Benford’s Law has previously been used to evaluate COVID-19 data [8,9] and other fabrications in the data [10]. It states that in a large dataset containing naturally recorded measurements, the incidence of digits 1 to 9 is not distributed equally. Instead, the distribution of these digits follows a counter-intuitive trend in which each digit occurs at a frequency of 11.11%, the digit “1” occurs at 30.1%, “2” at 17.6%, “3” at 12.5%, and so on [12].

By comparing the frequencies of different digits (especially 0 and 5) with the expected frequencies according to Benford’s Law, we can establish if terminal digit preference occurs in breast cancer size measurements and the Ki-67 index.

2. Materials and methods

It is a retrospective cross-sectional study. All the recorded data was obtained from the patient directory of KRL Hospital. At KRL Hospital, primary tumor resection for breast cancer patients was done in 1000 breast cancer patients from November 2016 to December 2020. Data related to these specimens were recorded and later interpreted by a histopathologist.

The calculated sample size for the study was 1000 breast cancer specimens by the WHO sample size calculator. Hence, the sizes and Ki-67 indexes of 1000 breast carcinoma tumors were included in the study. Fine needle aspiration tests and benign breast lesions were not included in the study. The expected frequencies of different digits in the data set were calculated according to Benford’s Law via the mathematical equation:

\[
F(d) = \log (1 + 1/d)
\]

where F is the frequency and d are the digits in question [1–9].

CAP protocols were followed in all the specimens included in the dataset. The recorded specimens were obtained during breast conservative procedures and modified radical mastectomies. The tumors were measured along their greatest dimension and recorded in cm to one decimal point. The tumor diameter was calculated from microscopic slides by measuring the outermost borders of the invasive lesion and measuring to the nearest millimeter using a transparent ruler. If no slides were available, the measurement was based on a macroscopic inspection (formalin-fixed material), either of a single tissue slice or as an average of all tissue slices containing microscopically validated invasive tumor tissue. The study proposal was approved by the ethical review committee of KRL Hospital. Our study is fully compliant with the STROCSS 2021 guidelines [13]. A complete STROCSS 2021 checklist has been provided as a supplementary file. Our study has been registered on Research Registry with the following UIN: researchregistry7969 [14]. Our study is by the Declaration of Helsinki.

The tumors were also stained for Ki-67 markers and were divided into nine groups depending upon the percentage of staining, i.e., (11–29%, 20–29%, 30–39%, 40–49%, 50–59%, 60–69%, 70–79%, 80–89%, 90–99%).

A Pearson’s chi-square goodness-of-fit test was performed. The null hypothesis is that the observed distribution of digits in the data set for breast tumor size and Ki-67 index follows Benford’s law; hence a distribution having a P-value > 0.05 is considered to adhere to Benford’s distribution. Mann–Whitney The U test was used to compare the calculated frequencies of digits 1 to 9 with the expected frequencies calculated according to Benford’s Law.

3. Results

Data from 1000 patients with primary malignant breast tumors was retrieved from KRL Hospital and analyzed. We found no pentameric preference for histopathological measurement of tumor size. On computing the last digit frequency for histopathological measurements of tumor size recorded in centimeters to one decimal place, the histogram displayed a higher frequency of digits three and nine. There was a decrease in the frequency around the peak of digit nine which correlates to digits seven, eight, and one.

The evidence of pentameric preference was found in the recording of the Ki-67 index. The histogram of the distribution of the Ki-67 index recorded in percentage revealed that the digits three, five, and six were in higher frequency. There was a marked drop in the frequency of digit nine. The rest of the terminal digits, although differing in frequency from that predicted by Benford’s law, somewhat followed Benford’s distribution curve, as shown in Fig. I.

The effect of terminal digits of tumor size on Ki-67 staining scores (low proliferative vs high proliferative) was evaluated using Mann–Whitney U Test. The test revealed the insignificant effect of terminal digits of tumor size on the Ki-67 staining scores (p = 0.114), as shown in Fig. II.

4. Discussion

It has not been thoroughly researched. Our study using 1000 specimens of primary malignant breast tumors collected from November 2016 to December 2020 at KRL Hospital aimed to assess terminal digit preference in breast tumor size and Ki–67 indices using Benford’s law.

Studies have shown that terminal digit preference occurs in pathological reporting of malignant tumors, including colorectal adenocarcinomas and breast tumors [5,6]. This type of preference can affect T size categorization and thus have serious implications in patient management and prognosis, particularly in tumor types where macroscopic tumor size affects staging, such as breast and thyroid tumors.

The histograms showing the distribution of the first digits of the largest diameter of tumor size and that of Ki-67% showed that some digits were reported at a higher frequency than others. The preferred terminal digits were 3 and 9 in the reporting of histopathological breast tumor size and 3, 5, and 6 in Ki-67 staining scores. The result of digit preference in tumor size is inconsistent with the previous nationwide study done in Norway, in which there was clear evidence of terminal digit preference for 0 and 5 (pentameric preference) in the measurements of mammographic and histopathologic tumor diameter [6].

The increased frequency of digits 3 and 9 in tumor size measurement found in our study does not affect T category classification according to the TNM system as 30 mm and 90 mm values do not lie at the border values. For example, a tumor diameter of 28 mm rounded off to 30 mm due to a digit preference of 3 would fall into the T2 category either way. Similarly, tumor diameters falling in the range of 80 mm–89 mm rounded off to 90 mm owing to digit preference would be classified as T3. We can say that histopathologists adopted a practical approach by avoiding T category border values, a finding consistent with a recent Dutch study [15].

The Ki-67% score is defined as the percentage of positively stained cancer cells among the total number of malignant cells evaluated. It is used as a prognostic marker in breast cancer. A higher Ki-67% score shows that the tumor is aggressive, more likely to spread, and is associated with a worse prognosis. In our study, there was a preference in the reporting of Ki-67 staining scores for digits 3, 5, and 6. As there is no common consensus on a cut-off value of Ki-67, the effects of terminal digit preference in the reporting of these scores remain unpredictable and need to be investigated further.

The Mann–Whitney U test was used to assess the effect of tumor size.
on Ki-67 staining scores, with 10% set as the cut-off value for low and high Ki-67 scores. It has been reported that a higher Ki-67 index is correlated with increasing tumor size [16]. However, in our study, the relationship between tumor size and Ki-67 values was found to be insignificant.

Using immunohistochemistry, researchers from Anyang Tumor Hospital have found evidence that distinguishing between TNBC subtypes with varying degrees of aggressiveness and prognosis can be done using p53 and ki-67 proteins. The prognosis is significantly worse for patients with high p53 or Ki-67 indices or a history of cancer in the family. Prognostic markers in TNBC include p53, Ki-67, and family history [17].

5. Limitations

The limitation of our study is that it included data from a single hospital with small sample size. The study did not include information about individual histopathologists who performed the measurements, and it was not possible to study intra-observer or inter-observer trends. With no knowledge of the specific cases affected by terminal digit preference, it is not possible to postulate the clinical implications of the measurement error. The current TNM staging guidelines do not consider terminal digit preference. More explicit guidelines are required to minimize errors in the pathological reporting of malignant tumors. To make our findings generally applicable, there should be uniformity in using the cut-off value for Ki-67. Further studies are required to probe the consequences resulting from this source of measurement error.

6. Strengths

It is essential to measure the impact of terminal digit preference especially when treatment guidelines can change the course of management. This study quantifies the impact of this covert tendency to alter the exact dimensions of the tumors. Hence, this study is an approach to unveiling the impact of terminal digit preference.

7. Conclusion

The recording of the Ki-67 index revealed evidence of pentameric preference. The numbers three, five, and six appeared more frequently in the histogram of the Ki-67 index distribution measured in percentage. The frequency of nine has dropped dramatically. The influence of tumor size terminal digits on Ki-67 staining scores (low proliferative vs. high proliferative) was assessed using the Mann–Whitney U test, which demonstrated that tumor size terminal digits did not affect Ki-67 staining scores.

Ethical approval

The ethical approval was obtained from ethical review committee of Shifa International Hospital Islamabad.

Reference number: 09-ERC/17/10/2015.

Sources of funding

No funding received.

Author contribution

All the authors contributed to the outlines of the research proposal, data collection, Analysis, writing, editing, proofing, and final approval of the paper.

Availability of data and material

Yes.
Registration of research studies

Name of the registry: Not Applicable.
Unique Identifying number or registration ID: Not Applicable.
Hyperlink to your specific registration (must be publicly accessible and will be checked):

Consent

Not Applicable.

Guarantor

Hassan Mumtaz.

Provenance and peer review

Not commissioned, externally peer-reviewed.

Declaration of competing interest

None to disclose.

Acknowledgements

Nill.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.amsu.2022.103993.

References

[1] S. Libson, M. Lippman, A review of clinical aspects of breast cancer, Int. Rev. Psychiatry. 26 (1) (2014 Feb) 4–15, https://doi.org/10.3109/09540621.2013.852071.PMID:24716497.
[2] M. Fahad Ullah, Breast cancer: current perspectives on the disease status, Adv. Exp. Med. Biol. 1152 (2019) 51–64, https://doi.org/10.1007/978-3-030-20301-6_4. PMID: 31456179.
[3] K.L. Maughan, M.A. Lutterbie, P.S. Ham, Treatment of breast cancer, Am. Fam. Physician 81 (11) (2010 Jun 1) 1339–1346. PMID: 20521754.
[4] G. Cserni, E. Chmielik, B. Cserni, T. Tot, The new TNM-based staging of breast cancer, Virchows Arch. 472 (5) (2018 May) 697–703, https://doi.org/10.1007/s00428-018-2301-9. Epub 2018 Jan 27. PMID: 29380126.
[5] Hayes SJ Terminal digit preference occurs in pathology reporting irrespective of patient management implications, J. Clin. Pathol. 61 (2008) 1071–1072.
[6] Kaitlyn M. Tsaruda, Solveig Hofvind, Lars A. Aksten, Solveig R. Hoff, Marit B. Veierød, Terminal digit preference: a source of measurement error in breast cancer diameter reporting, Acta Oncol. 59 (3) (2020) 260–267, https://doi.org/10.1080/0284186X.2019.1669917.
[7] Maggie C.U. Cheang, SMaggie C.U. Cheang, Stephen K. Chia, et al., Ki67 index, HER2 status, and prognosis of patients with luminal B breast cancer, JNCC. J. Natl. Cancer Inst. 101 (10) (20 May 2009) 736–750, https://doi.org/10.1093/jnci/djp082.
[8] M. Campolieti, COVID-19 deaths in the USA: Benford’s law and under-reporting, J. Public Health 24 (2021 May) fdab161, https://doi.org/10.1093/PubMed/fdab161. Epub ahead of print. PMID: 34027553; PMCID: PMC8194618.
[9] A.P. Kennedy, S.C.P. Yam, On the authenticity of COVID-19 case figures, PLoS One 15 (12) (2020), e0243123, https://doi.org/10.1371/journal.pone.0243123.
[10] Feifei Li, Shuqing Han, Hongyv Zhang, Jiaojiao Ding, Jianhua Zhang, Jianzhai Wu, Application of Benford’s law in data analysis, J. Phys. Conf. 1168 (2019), 032133, https://doi.org/10.1088/1742-6596/1168/3/032133.
[11] Beer TW Terminal digit preference: beware of Benford’s law, J. Clin. Pathol. 62 (2009) 192.
[12] Frank Benford, The law of anomalous numbers, in: Proceedings of the American Philosophical Society, vol. 78, American Philosophical Society, 1938, pp. 551–572, http://www.jstor.org/stable/984802.
[13] G. Mathew, R. Agha, for the STROCSS Group, Stroccs 2021: strengthening the Reporting of cohort, cross-sectional and case-control studies in Surgery, Int. J. Surg. 96 (2021), 106165.
[14] https://www.researchregistry.com/browse-the-registry#home/registrationdetails/62998956facf11001fac6ee4/.
[15] M.A. den Bakker, R.A. Damhuis, Pentameric last-digit preference and stage border avoidance in pathology measurement, Histopathology 73 (3) (2018 Sep) 510–513.
[16] J.S. Nigam, T. Kumar, S. Bharti, R. Sinha, P.P. Bhadani, Association of ki-67 with clinicopathological factors in breast cancer, Cureus 13 (6) (2021 Jun 13).
[17] Y. Pan, Y. Yuan, G. Liu, Y. Wei, P53 and Ki-67 as prognostic markers in triple-negative breast cancer patients, PLoS One 12 (2) (2017), e0172524, https://doi.org/10.1371/journal.pone.0172324.