Survival Time in Hematologic Malignancies Obeys Newcomb-Benford Law

To The Editor: Newcomb-Benford law (NBL) states that the distribution of the first digit in most empirical numerical series is not uniform, and there is inherent predilection toward lower digit values. Specif-
ically, the probability of the digit 1 being first is approximately 30.1%, whereas for the digit 9, the probability decreases to approximately 4.6%. Not all numerical sets abide by this law; however, for those that do, diversions indicate data modifications including potentially malicious behaviors such as using fraudulent data. It was previously shown that incidence of cancer follows NBL, but it is unknown whether survival data in cancer registries abide by this law; therefore, we sought to investigate that using a large national cancer database in the United States.

Basic demographic and survival data for all patients with the following hematologic malignancies—acute myeloid leukemia (AML), chronic myeloid leukemia (CML), multiple myeloma (MM), Hodgkin lymphoma (HL), diffuse large B-cell lymphoma (DLBCL), and follicular lymphoma (FL)—diagnosed between 1975 and 2016 were extracted from the Surveillance, Epidemiology, and End Results (SEER) program database. Survival time is reported in months. Patients with missing survival data or follow-up less than 1 month were excluded. Patients with each disease were divided into 9 groups based on the first digit of their overall survival numerical value. Pearson’s correlation coefficient (Pearson’s $r^2$) was used to quantify the concordance between the distribution of first digits of the survival time (in months), and Newcomb-Benford law predicted percentages. Statistical analyses were done using Statistical Software for Social Sciences, version 24. Institutional Review Board approval was not required, as SEER provides deidentified patient records.

A total of 455,229 patients were included in the final analysis. At a median follow-up of 119 months, 278,629 patients had died. The distribution of the first digits of the survival time in all of our selected diseases followed NBL (Figure [A]). Concordance by Pearson’s $r^2$ between the observed digits distribution in our data set and that predicted by NBL for any numerical series that abide was high (AML $= 0.99$, CML $= 0.97$, DLBCL $= 0.95$, FL $= 0.93$, HL $= 0.97$, MM $= 1.00$; scale of 0 to 1, in which 1 indicates higher concordance). In the 228,975 patients with lymphoma in whom staging data (using the Ann Arbor staging system) were available, a similar concordant
pattern was seen, as shown in the Figure (B).

To our knowledge, this is the first report to show that registry-based survival data in the most common hematologic malignancies do follow NBL. As mentioned previously, in numerical sets that abide by NBL, deviations indicate nonrandom data modifications. These modifications can be fraudulent or could be resulting from human error in data maintenance or effects of an external agent not related to the data itself (perhaps the effects of a certain therapy on survival). Nevertheless, NBL is an easy and simple way to evaluate quality of survival data. Our results warrant validation in other data sets, especially those with different therapeutical interventions.

Taha Al-Juhaishi, MD
Stephenson Cancer Center
University of Oklahoma Health Sciences Center
Oklahoma City OK

Sadeer G. Al-Kindi, MD, FAC
Harrington Heart and Vascular Institute
University Hospitals School of Medicine
Case Western Reserve University
Cleveland, OH

Grant Support: No grant support was received for this article.

Potential Competing Interests: The authors report no competing interests.


https://doi.org/10.1016/j.mayocp.2021.11.029

Association Between Weekly Exercise Time and Mortality

To the Editor: We read with great interest the work of Schnohr et al1 about the association between weekly exercise time and cardiovascular disease (CVD) mortality. During ±25 years of follow-up, 4779 deaths (54.9% of study population) were recorded, of which 2054 (23.6%) were CVD related. The authors reported a U-shaped association between weekly exercise time and CVD mortality and all-cause mortality, with the lowest risk for individuals exercising 2.6 to 4.5 h/wk. These findings remained consistent after correction for competing risks, stratified analyses, and analyses to account for reverse causation. The outcomes of this study are contradictory to the current dogma that exercise is a potent medicine at any dose.2 We are intrigued by these findings and want to share some thoughts with the authors.

The categorization of individuals into 5 distinct groups is an important methodologic consideration as the middle (third) group is used as the reference group to explore the presence of a U shape. The authors describe that participants were allocated to an exercise group of 0 h/wk, 0.1 to 2.5 h/wk, 2.6 to 4.5 h/wk, 4.6 to 10 h/wk, or more than 10 h/wk on the basis of the frequency distribution of the data. It appeared, however, that the reference group had a substantially smaller sample size (n=494, 6% of total cohort) compared with the other groups (n=1602 to n=2679, 18% to 31% of total cohort). It is unclear why this approach was chosen, whether the exercise time data had a Gaussian distribution (or not), and whether different group distributions (eg, quintiles) would have led to different outcomes. To improve our understanding of the dose-response association between exercise volumes and mortality, it would also be interesting to use nonlinear models (eg, restricted cubic spline analysis) with weekly exercise time as the continuous variable.

Another concern is the assessment of the exposure variable. Exercise habits are known to be highly variable across the human life span,3,4 and a single measurement of weekly exercise time could induce a nondifferential measurement error, especially after a long follow-up time, such as in this study. This error could subsequently result in an underestimation of the true effect of exercise on health. Indeed, a recent study reported a U-shaped association between exercise volumes and CVD mortality and all-cause mortality when a single measure of physical activity was used.5 Adding repeated measures of physical activity to the analysis changed the outcome dramatically as the U-shape association was replaced by a curvilinear association, indicating that the largest exercise volumes yielded the greatest health benefits. Outcomes of this study should therefore be interpreted with care. We also encourage the authors to analyze their data in a similar fashion as repeated measurements appear to be available in the Copenhagen City Heart Study.

This work adds novel data to the ongoing debate about the dose-response association of exercise and