Editorial

Kidney Res Clin Pract 2024;43(2):131-132
pISSN: 2211-9132 • eISSN: 2211-9140
https://doi.org/10.23876/j.krcp.23.228

KIDNEY RESEARCH AND CLINICAL PRACTICE

On operational definitions of mortality

Hakmook Kang

Department of Biostatistics, Vanderbilt University Medical Center, Nashville, TN, USA

See Article on Pages 156–164

Mortality or mortality rate has been collected since 1750's [1] and are utilized across various research fields such as epidemiology, biostatistics, and biomedical and biopharmaceutical research. Excess mortality, denoting mortality above the normal rate, typically demands deeper investigation as it can signal fatal diseases or infections. In biomedical research, unraveling the causes of excess mortality often sparks new lines of inquiry, such as research into developing vaccines for coronavirus disease 2019. Consequently, instances of increased mortality or excess mortality consistently attract significant attention in the biomedical research realm.

In the era of big data, encompassing electronic health records (EHRs), multi-modal magnetic resonance imaging data, and multiomics data, comprehending data structures and integrating diverse information sources to address critical scientific questions is paramount. In large-scale datasets, encountering missing data is inevitable due to various factors, such as randomness or systematic issues [2]. Particularly, EHR-type data tend to exhibit missing observations [3,4].

One strategy to tackle the missing data challenge involves imputing missing values using non-missing information within the dataset, provided the non-missing information predicts the missing values. For example, missing body weight values could be imputed using height, sex, and age data via a linear regression model [5]. Herein, understanding variable associations within a dataset is pivotal for successful imputation.

Nevertheless, when missing observations arise from systematic missingness, such as the absence of death information in the Health Insurance Review and Assessment Service database, imputation becomes more limited and complex. Integrating a dataset containing systematic missingness with another containing the missing information can help obtain the necessary data without constructing an imputation model. In cases where obtaining another dataset with the missing information isn't feasible, imputation methods such as single or multiple imputation may need to be considered [5].

Mortality stands as a cornerstone parameter in biomedical research [6,7]. Sometimes, addressing critical scientific inquiries becomes infeasible without knowledge of mortality rates. In such instances, operational definitions of mortality, as outlined in [8,9], can be established. A recent article by Lee et al. [10], titled "*Validation of operational definitions of mortality in a nationwide hemodialysis population using the Health Insurance Review and Assessment Service databases of Korea*," delves into the validation of several operational definitions of mortality. These defi-

Received: August 26, 2023; **Accepted:** October 24, 2023
Correspondence: Hakmook Kang
Department of Biostatistics, Vanderbilt University Medical Center, 2525 West End Ave, Suite 1100, Nashville, TN 37203, USA.
E-mail: hakmook.kang@vanderbilt.edu
ORCID: https://orcid.org/0000-0001-6876-4021

nitions encompass intervals of 30, 60, 90, 120, 150, and 180 days of no health insurance claims. The study reveals that an operational definition requiring 150 days free from health insurance claims yielded the most accurate results.

From a statistical standpoint, the uniformity of a single definition across all age and sex groups is intriguing. Nonetheless, leaning towards more conservative or lenient mortality estimates based on a single operational definition within a particular age group could introduce bias to study outcomes. Employing statistical models to account for potential confounding factors like age or sex might be essential. If one definition consistently overestimates mortality within a specific age and sex combination, proposing an alternative definition for that combination could prove beneficial. Similarly, suggesting multiple operational definitions for various strata, such as specific age and sex combinations, could help mitigate bias. However, the trade-off between bias and variance needs careful consideration, as too many definitions may lead to overfitting and reduced generalizability.

The assessment of estimation can be approached through external and internal validations. While external validation using external datasets is the gold standard, internal validation can be achieved through k-fold cross-validation combined with bootstrapping when external data are unavailable.

In this context, exploring the distribution of deviations from true values for each age stratum via bootstrapping could be illuminating. Repeatedly resampling data within each age group and calculating deviations from true values generates a bootstrap distribution of deviations, essentially reflecting the variation of mortality within each age group. By proposing multiple operational definitions and generating corresponding bootstrap distributions of deviations, it becomes possible to identify approaches that exhibit the smallest variance, thus being more generalizable to similar datasets. Beyond proposing definitions, assessing the variance associated with each definition contributes to finding an optimal balance between bias (measured by mean deviation) and variance (bootstrap variance).

Conflicts of interest

The author has no conflicts of interest to declare.

Data sharing statement

The data presented in this study are available upon reasonable request from the corresponding author.

ORCID

Hakmook Kang, https://orcid.org/0000-0001-6876-4021

References

1. Högberg U, Wall S. Secular trends in maternal mortality in Sweden from 1750 to 1980. *Bull World Health Organ* 1986;64:79–84.
2. Atkinson HH, Rosano C, Simonsick EM, et al. Cognitive function, gait speed decline, and comorbidities: the health, aging and body composition study. *J Gerontol A Biol Sci Med Sci* 2007;62:844–850.
3. Jazayeri A, Liang OS, Yang CC. Imputation of missing data in electronic health records based on patients' similarities. *J Healthc Inform Res* 2020;4:295–307.
4. Hu Z, Melton GB, Arsoniadis EG, Wang Y, Kwaan MR, Simon GJ. Strategies for handling missing clinical data for automated surgical site infection detection from the electronic health record. *J Biomed Inform* 2017;68:112–120.
5. Sterne JA, White IR, Carlin JB, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ* 2009;338:b2393.
6. Son HE, Moon JJ, Park JM, et al. Additive harmful effects of acute kidney injury and acute heart failure on mortality in hospitalized patients. *Kidney Res Clin Pract* 2022;41:188–199.
7. Neyra JA, Ortiz-Soriano V, Liu LJ, et al. Prediction of mortality and major adverse kidney events in critically ill patients with acute kidney injury. *Am J Kidney Dis* 2023;81:36–47.
8. Lee HS, Song YR, Kim JK, et al. Outcomes of vascular access in hemodialysis patients: analysis based on the Korean National Health Insurance Database from 2008 to 2016. *Kidney Res Clin Pract* 2019;38:391–398.
9. Choi H, Kim M, Kim H, et al. Excess mortality among patients on dialysis: comparison with the general population in Korea. *Kidney Res Clin Prac* 2014;33:89–94.
10. Lee DH, Kim YJ, Kim H, Lee HS. Validation of operational definitions of mortality in a nationwide hemodialysis population using the Health Insurance Review and Assessment Service databases of Korea. *Kidney Res Clin Pract* 2024;43:156–164.