Impact of death certificate misclassifications on radiation cancer risk estimates

Xirui Liu[1], Stacey L. McComish[1], Joey Y. Zhou[2], Sergei Y. Tolmachev[1]

[1]United States Transuranium and Uranium Registries, College of Pharmacy and Pharmaceutical Sciences, Washington State University, Richland, WA, USA

[2]Office of Domestic and International Health Studies, United States Department of Energy, Washington, D.C., USA

Introduction

According to previous research[1] observations, misclassification of cause of death can occur on death certificates compared to the diagnoses from autopsy reports, which are considered the gold standard. In our previous study, the overall misclassification rate was 25.5%, with varying rates for different disease categories. The over-classification rate was 4.2% for cancer cause of death and 11.8% for cardiovascular diseases, while the under-classification rate for cancer was 14.1% and for cardiovascular diseases was 25.8%. These misclassifications can influence the results of studies when investigating the association between radiation dose and cause of death. Therefore, this analysis focuses on cancer cause of death and aims to evaluate whether these misclassifications have a statistically significant impact on the conclusions. We generated a dataset containing 1,000 cases with simulated dose and cancer outcome data by setting odds ratio equaling 1.0 and a baseline cancer incidence of 30%. Odds ratios and corresponding *p*-values were computed and compared after 10,000 simulations conducted with predetermined over- and under-classification rates.

Dose Data Generation

- A normal distribution is used to simulate dose data where the geometric mean (*GM*) is calculated based on real USTUR external dose data and a geometric standard deviation (*GSD*) value of 1.1 is set to ensure that the generated dose data follows a realistic normal distribution.

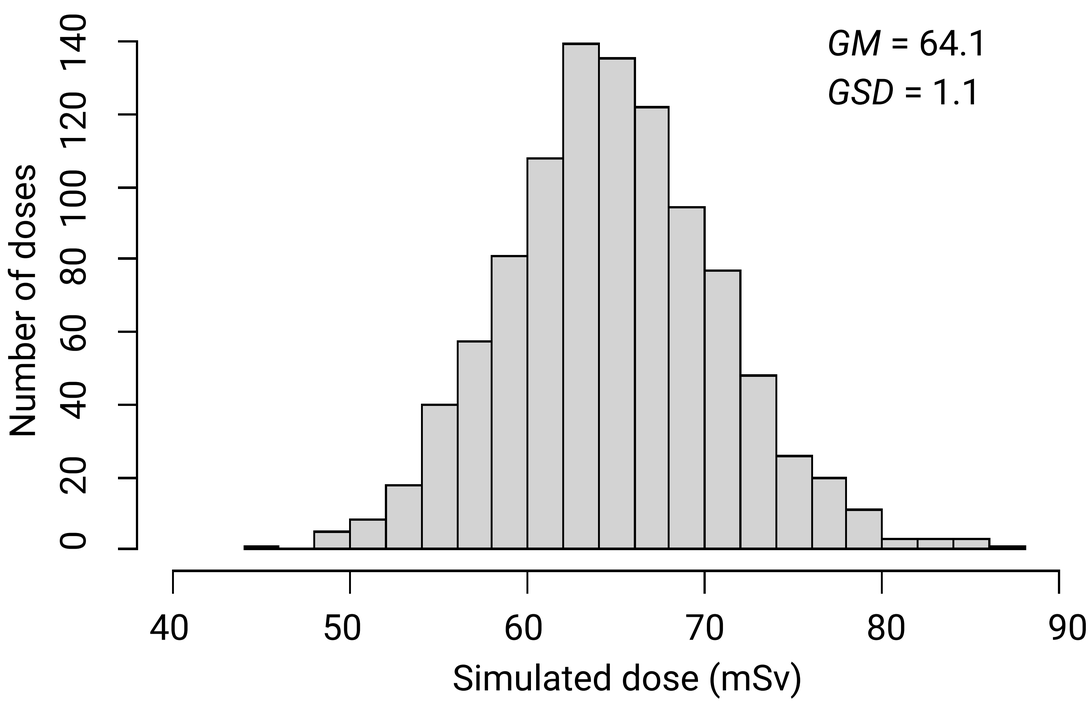


Figure 1: Dose data distribution generated based on selected *GM* and *GSD*.

Cancer Outcome Data Generation

- Logistic function is used to calculate the probability of cancer outcome $p(x)$ for each case:

$$p(x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}}$$

  x: The dose value (mSv)

  β_0: The constant derived from a baseline cancer incidence of 30% when x equals zero

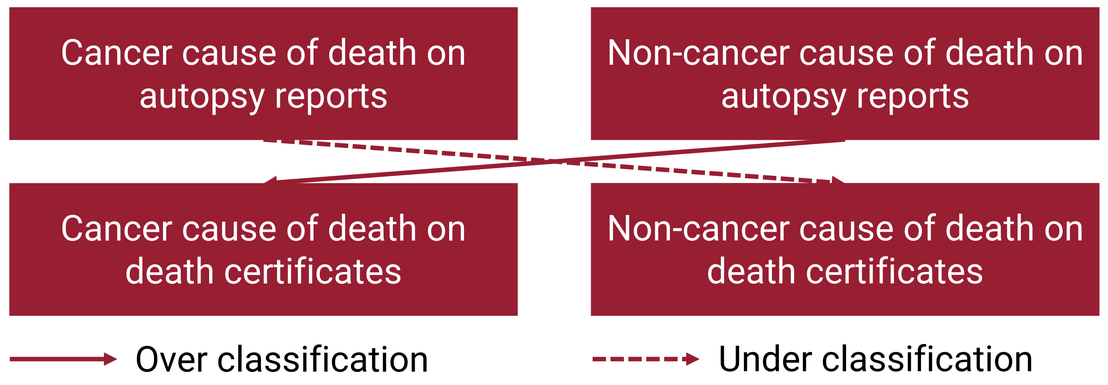
  β_1: The logarithm of the odds ratio, where the odds ratio was set to 1.0

- This process is repeated 1,000 times to obtain a dataset that includes both dose values and outcomes.

Misclassification Simulation

$$\text{Over classification rate} = \frac{Number\ of\ False\ Positives}{Total\ Number\ of\ Noncancer\ on\ ARs}$$

$$\text{Under classification rate} = \frac{Number\ of\ False\ Negatives}{Total\ Number\ of\ Cancer\ on\ ARs}$$



Misclassification Rates Impact on Significance of Results

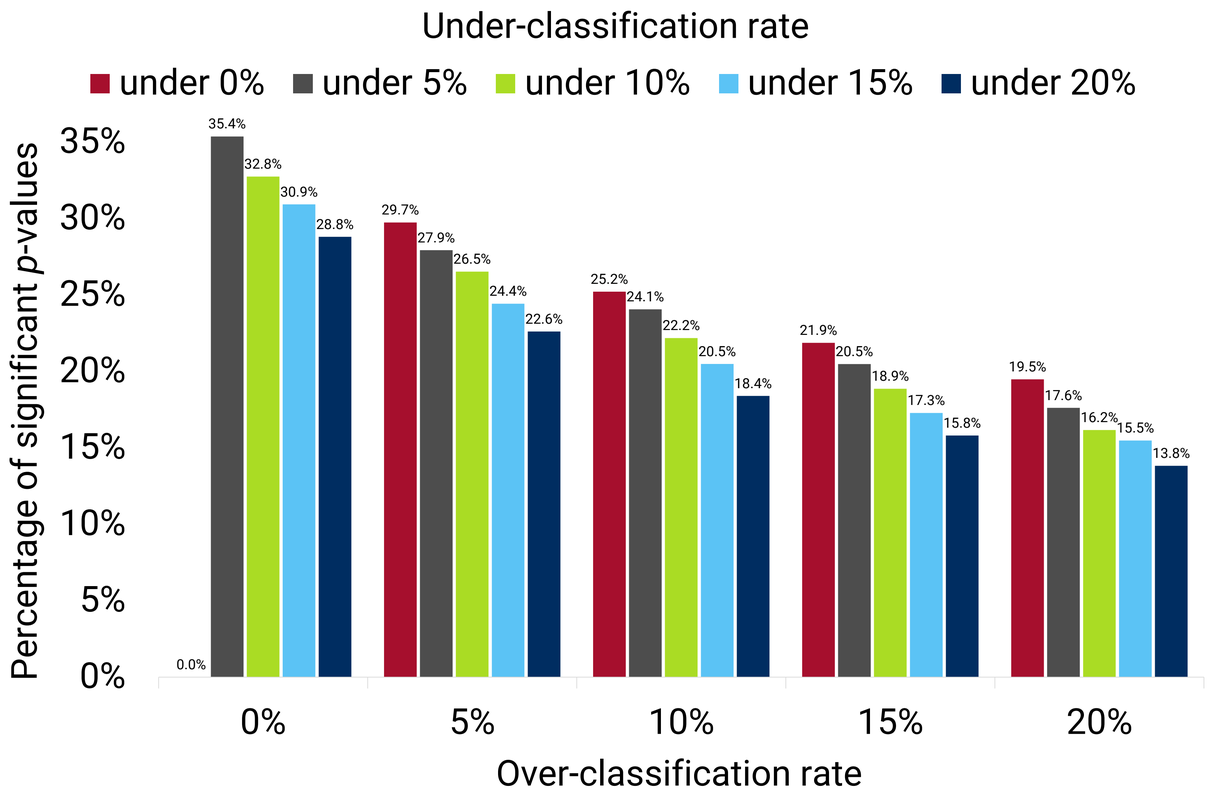


Figure 2: How different over- and under-classification rates impact on percentage of significant *p*-values.

Discussion

- This experiment selected one possible initial dataset for simulations. Modification of the initial dataset can result in different scenarios with different findings.
- The objective of this experiment is to reveal the trends of how the misclassification impacts conclusions under nonsignificant initial conditions. The possible reasons for these effects will be further investigated in subsequent studies, and more representative data will be simulated by altering the parameters.

Conclusions

- In a population with a non-significant odds ratio of 1.0, misclassification errors on death certificates can result in odds ratios that have significant *p*-values 10 - 35% of the time.
- The probabilities of obtaining significant *p*-values decrease as the rates of the over or under classification increase, which indicates that smaller misclassification rates are more likely to result in incorrect conclusions.

Reference

1. Gold B, Kathren RL. Causes of death in a cohort of 260 plutonium workers. Health Phys. 1998;75(3):236-40. doi: 10.1097/00004032-199809000-00001