WASHINGTON STATE UNIVERSITY
College of Pharmacy and
Pharmaceutical Sciences

Impact of Death Certificate Misclassifications on Health Risk Models

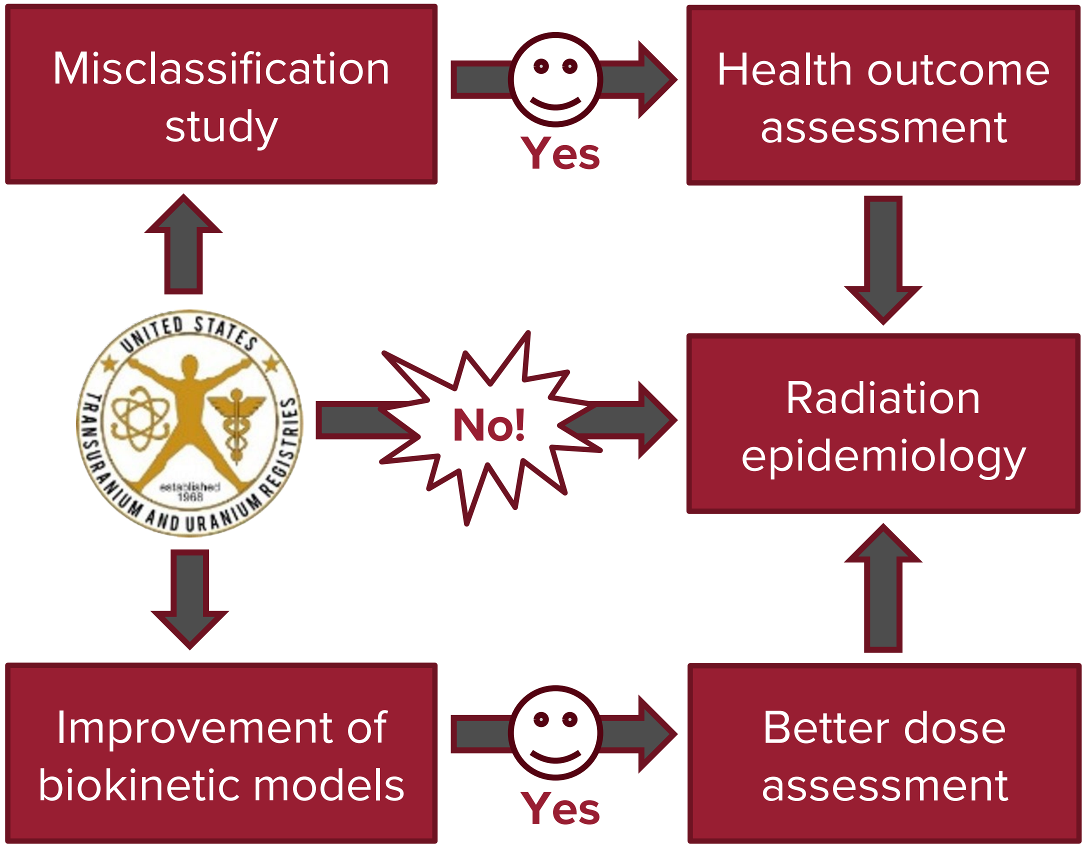
Xirui Liu[1], Stacey L. McComish[1], Joey Y. Zhou[2], Sergey Y. Tolmachev[1]

[1]United States Transuranium and Uranium Registries, College of Pharmacy and Pharmaceutical Sciences, Washington State University
[2]Office of Domestic and International Health Studies, U.S. Department of Energy

WASHINGTON STATE UNIVERSITY

US Transuranium and Uranium Registries

- Established in 1968, a research program to follow up with occupationally exposed workers by studying the biokinetics and tissue dosimetry of the actinides [*in support of radiation epidemiology studies*]
- USTUR Registrants: individuals with documented history of exposure to the actinides (mainly plutonium)
- Since 1992, it has been a US Department of Energy grant to Washington State University



WASHINGTON STATE UNIVERSITY

Misclassification Study

PLOS ONE

OPEN ACCESS PEER-REVIEWED

RESEARCH ARTICLE

Misclassification of causes of death among a small all-autopsied group of former nuclear workers: Death certificates vs. autopsy reports

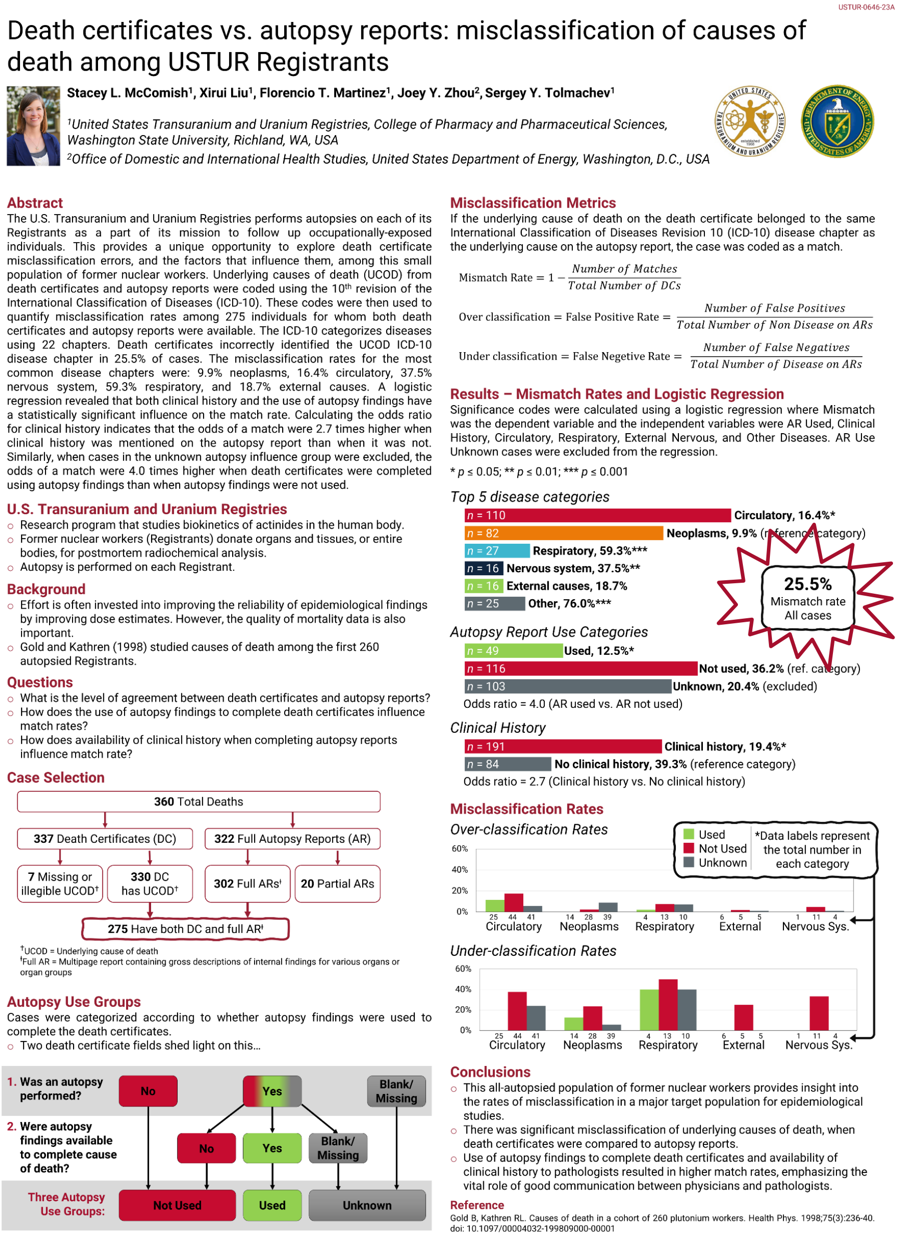
Stacey L. McComish, Xirui Liu, Florencio T. Martinez, Joey Y. Zhou, Sergey Y. Tolmachev

Published: May 3, 2024 • https://doi.org/10.1371/journal.pone.0302069

Article | Authors | Metrics | Comments | Media Coverage

$$\text{Over-classification rate} = \frac{\text{Number of false positives}}{\text{Total number of non disease on ARs}}$$

$$\text{Under-classification rate} = \frac{\text{Number of false negatives}}{\text{Total number of disease on ARs}}$$



USTUR-0646-23A

Death certificates vs. autopsy reports: misclassification of causes of death among USTUR Registrants

Stacey L. McComish[1], Xirui Liu[1], Florencio T. Martinez[1], Joey Y. Zhou[2], Sergey 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

Abstract

The U.S. Transuranium and Uranium Registries performs autopsies on each of its Registrants as a part of its mission to follow up occupationally-exposed individuals. This provides a unique opportunity to explore death certificate misclassification errors, and the factors that influence them, among this small population of former nuclear workers. Underlying causes of death (UCOD) from death certificates and autopsy reports were coded using the 10th revision of the International Classification of Diseases (ICD-10). These codes were then used to quantify misclassification rates among 275 individuals for whom both death certificates and autopsy reports were available. The ICD-10 categorizes diseases using 22 chapters. Death certificates incorrectly identified the UCOD ICD-10 disease chapter in 25.5% of cases. The misclassification rates for the most common disease chapters were: 9.9% neoplasms, 16.4% circulatory, 37.5% nervous system, 59.3% respiratory, and 18.7% external causes. A logistic regression revealed that both clinical history and the use of autopsy findings have a statistically significant influence on the match rate. Calculating the odds ratio for clinical history indicates that the odds of a match were 2.7 times higher when clinical history was mentioned on the autopsy report than when it was not. Similarly, when cases in the unknown autopsy influence group were excluded, the odds of a match were 4.0 times higher when death certificates were completed using autopsy findings than when autopsy findings were not used.

U.S. Transuranium and Uranium Registries

- Research program that studies biokinetics of actinides in the human body.
- Former nuclear workers (Registrants) donate organs and tissues, or entire bodies, for postmortem radiochemical analysis.
- Autopsy is performed on each Registrant.

Background

- Effort is often invested into improving the reliability of epidemiological findings by improving dose estimates. However, the quality of mortality data is also important.
- Gold and Kathren (1998) studied causes of death among the first 260 autopsied Registrants.

Questions

- What is the level of agreement between death certificates and autopsy reports?
- How does the use of autopsy findings to complete death certificates influence match rates?
- How does availability of clinical history when completing autopsy reports influence match rate?

Case Selection

360 Total Deaths
337 Death Certificates (DC)
322 Full Autopsy Reports (AR)
7 Missing or illegible UCOD†
330 DC has UCOD†
302 Full ARs‡
20 Partial ARs
275 Have both DC and full AR‡

†UCOD = Underlying cause of death
‡Full AR = Multipage report containing gross descriptions of internal findings for various organs or organ groups

Autopsy Use Groups

Cases were categorized according to whether autopsy findings were used to complete the death certificates.

- Two death certificate fields shed light on this...

1. Was an autopsy performed? No / Yes / Blank/Missing
2. Were autopsy findings available to complete cause of death? No / Yes / Blank/Missing

Three Autopsy Use Groups: Not Used / Used / Unknown

Misclassification Metrics

If the underlying cause of death on the death certificate belonged to the same International Classification of Diseases Revision 10 (ICD-10) disease chapter as the underlying cause on the autopsy report, the case was coded as a match.

$$\text{Mismatch Rate} = 1 - \frac{\textit{Number of Matches}}{\textit{Total Number of DCs}}$$

$$\text{Over classification} = \text{False Positive Rate} = \frac{\textit{Number of False Positives}}{\textit{Total Number of Non Disease on ARs}}$$

$$\text{Under classification} = \text{False Negetive Rate} = \frac{\textit{Number of False Negatives}}{\textit{Total Number of Disease on ARs}}$$

Results – Mismatch Rates and Logistic Regression

Significance codes were calculated using a logistic regression where Mismatch was the dependent variable and the independent variables were AR Used, Clinical History, Circulatory, Respiratory, External Nervous, and Other Diseases. AR Use Unknown cases were excluded from the regression.

* $p \leq 0.05$; ** $p \leq 0.01$; *** $p \leq 0.001$

Top 5 disease categories

- n = 110 **Circulatory, 16.4%***
- n = 82 **Neoplasms, 9.9%** (reference category)
- n = 27 **Respiratory, 59.3%*****
- n = 16 **Nervous system, 37.5%****
- n = 16 **External causes, 18.7%**
- n = 25 **Other, 76.0%*****

25.5% Mismatch rate All cases

Autopsy Report Use Categories

- n = 49 **Used, 12.5%***
- n = 116 **Not used, 36.2%** (ref. category)
- n = 103 **Unknown, 20.4%** (excluded)

Odds ratio = 4.0 (AR used vs. AR not used)

Clinical History

- n = 191 **Clinical history, 19.4%***
- n = 84 **No clinical history, 39.3%** (reference category)

Odds ratio = 2.7 (Clinical history vs. No clinical history)

Misclassification Rates

Over-classification Rates

Under-classification Rates

Conclusions

- This all-autopsied population of former nuclear workers provides insight into the rates of misclassification in a major target population for epidemiological studies.
- There was significant misclassification of underlying causes of death, when death certificates were compared to autopsy reports.
- Use of autopsy findings to complete death certificates and availability of clinical history to pathologists resulted in higher match rates, emphasizing the vital role of good communication between physicians and pathologists.

Reference

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

Objectives

In general, it is believed that if outcome misclassification were incorporated into epidemiological studies that have significant dose-response associations, the findings would be more significant.

- Can misclassification of disease increase the risk estimate of dose-response associations?
- Can misclassification of disease change the conclusion of significant dose-response associations?

Methods: Initial Dataset Generation I (a)

Generate dose data using a log-normal distribution with predefined values for the geometric mean and the geometric standard deviation

→

Generate outcome data randomly using logistic probability function based on dose data, pre-set odds ratio and outcome baseline

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

×1,000 times

→

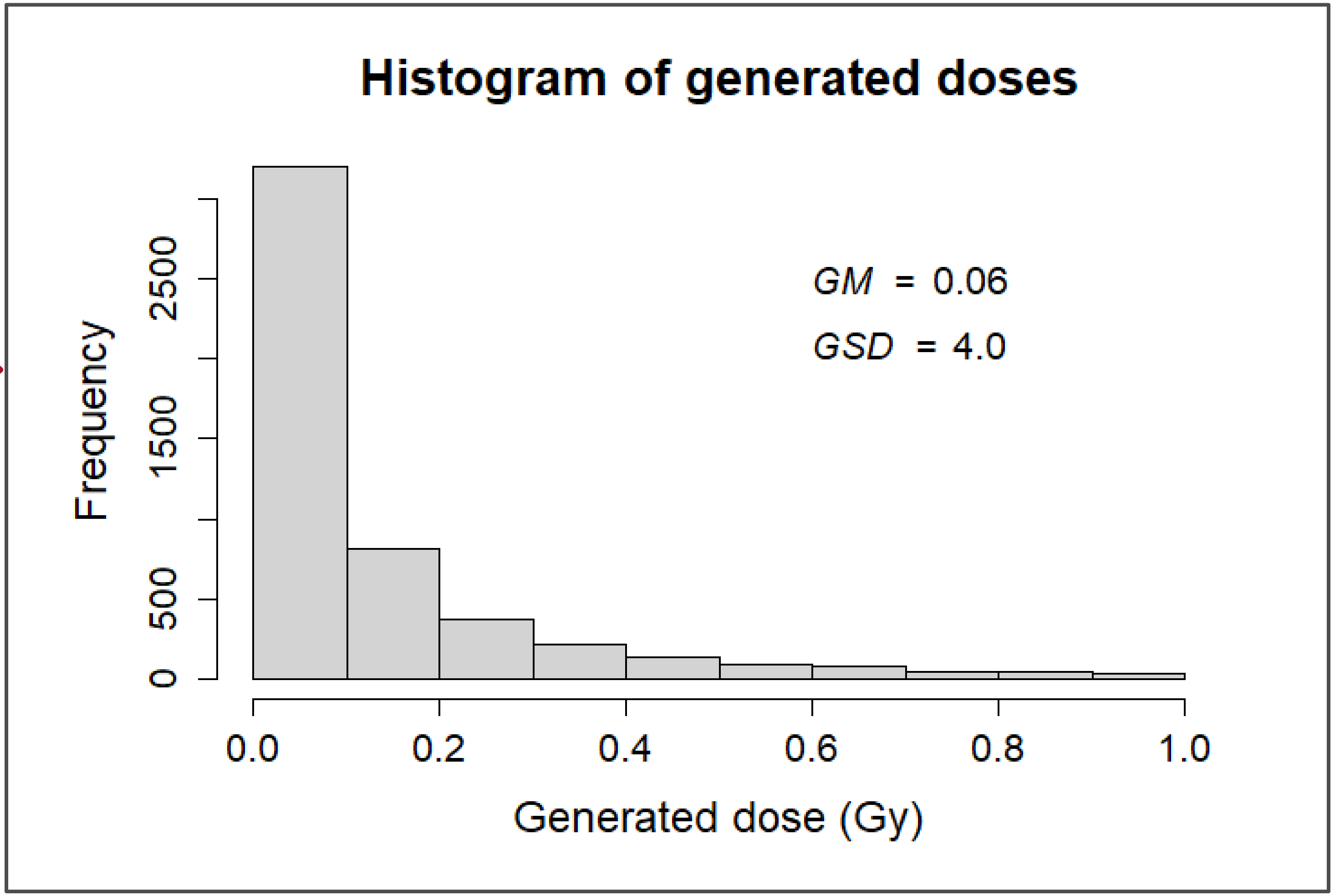
Select initial dataset of odds ratio close to 1.0 (statistically non-significant association)

odds ratio = 1.001

p-value = 0.998

Methods: Initial Dataset Generation I (b)

Generate dose data using a log-normal distribution with predefined values for the geometric mean and the geometric standard deviation



WASHINGTON STATE UNIVERSITY

Methods: Initial Dataset Generation I (c)

Generate dose data using a log-normal distribution with predefined values for the geometric mean and the geometric standard deviation

Generate outcome data randomly using logistic probability function based on dose data, pre-set odds ratio and outcome baseline

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

×1,000 times

Select initial dataset of odds ratio close to 1.0 (statistically non-significant association)

odds ratio = 1.001

p-value = 0.998

x: The **dose** value

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

β_1: The logarithm of the **odds ratio**

Methods: Initial Dataset Generation I (d)

Generate dose data using a log-normal distribution with predefined values for the geometric mean and the geometric standard deviation

→

Generate outcome data randomly using logistic probability function based on dose data, pre-set odds ratio and outcome baseline

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

×1,000 times

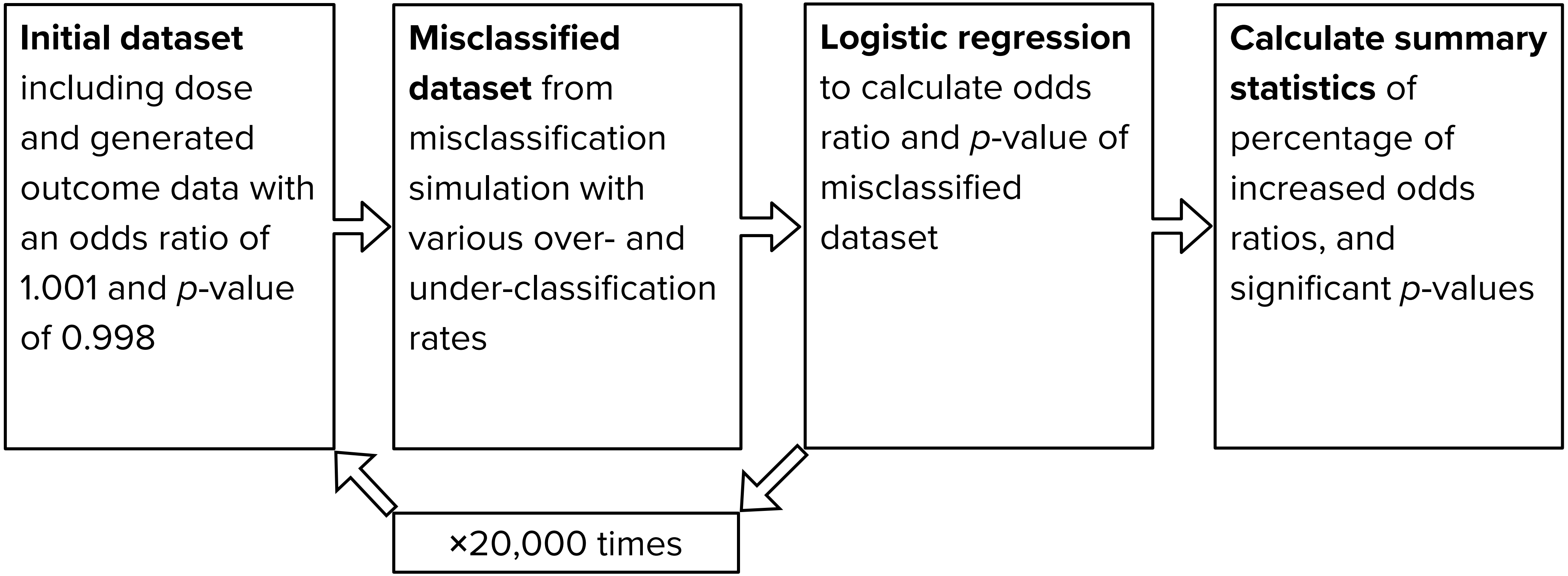
→

Select initial dataset of odds ratio close to 1.0 (statistically non-significant association)

odds ratio = 1.001

p-value = 0.998

Methods: Misclassification Simulation I

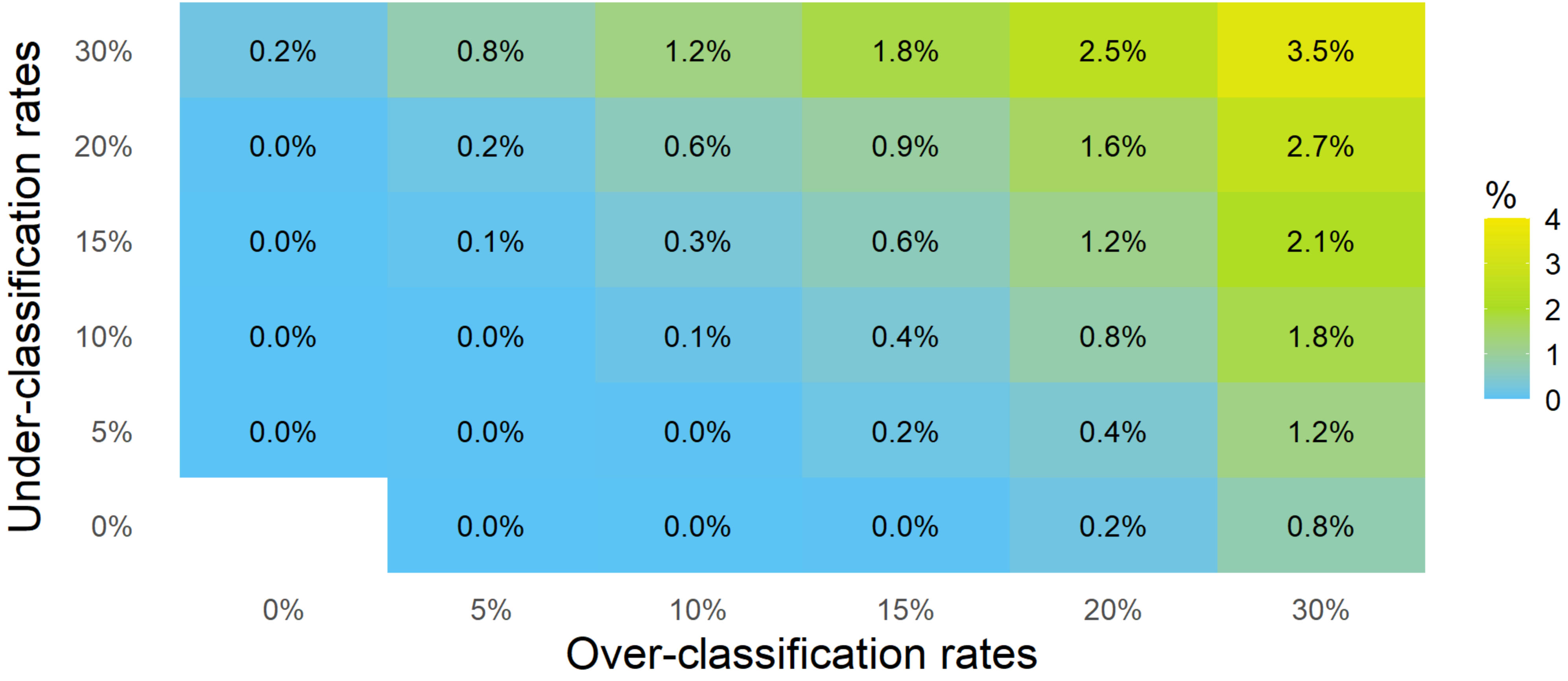


Normally it is expected the impact of misclassification would decrease odds ratio and increase *p*-value, which would tend to reduce the significance of conclusion.

WASHINGTON STATE UNIVERSITY

Results I: Percentage of Increased Odds Ratios after Misclassification

Results I: Percentage of Significant *p*-values after Misclassification



WASHINGTON STATE UNIVERSITY

Conclusions I

- In a population with a non-significant odds ratio close to 1.0 (statistically non-significant association) and preset over- and under-classification rates from 0% to 30%, misclassification errors on death certificates can result in
  - increased odds ratio between 48.3% and 51.9% of the time
  - significant *p*-values between 0% and 3.5% of the time
- While misclassification frequently impacted the odds ratio, it rarely led to incorrectly concluding that a relationship was statistically significant when it wasn't.

Methods: Initial Dataset Generation II

Generate dose data using a log-normal distribution with predefined values for the geometric mean and the geometric standard deviation

→

Generate outcome data randomly using logistic probability function based on dose data, pre-set odds ratio and outcome baseline

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

×1,000 times

→

Select initial dataset with *p*-value barely larger than 0.05 (statistically non-significant association)

odds ratio = 1.39

p-value = 0.05001

WASHINGTON STATE UNIVERSITY

Methods: Misclassification Simulation II

Initial dataset including dose and generated outcome data with an odds ratio of 1.39 and *p*-value of 0.05001

→ **Misclassified dataset** from misclassification simulation with various over- and under-classification rates

→ **Logistic regression** to calculate odds ratio and *p*-value of misclassified dataset

→ **Calculate summary statistics** of percentage of increased odds ratios, and significant *p*-values

×20,000 times

Results II: Percentage of Increased Odds Ratios after Misclassification

Under-classification rates \ Over-classification rates	0%	5%	10%	15%	20%	30%
30%	34.7%	24.4%	18.7%	14.5%	11.8%	8.4%
20%	37.0%	26.1%	20.9%	16.3%	13.9%	10.5%
15%	38.3%	27.4%	21.7%	17.7%	15.2%	12.0%
10%	41.2%	29.0%	22.9%	18.9%	16.3%	13.6%
5%	44.7%	30.0%	24.0%	20.4%	18.2%	15.4%
0%		30.9%	25.9%	22.0%	20.0%	16.9%

% (colour scale): 45, 35, 25, 15, 5

WASHINGTON STATE UNIVERSITY

Results II: Percentage of Significant *p*-values after Misclassification

Conclusions II

- In a population with a non-significant *p*-value barely larger than 0.05 (statistically non-significant association) and preset over- and under-classification rates from 0% to 30%, misclassification errors on death certificates can result in
  - increased odds ratio between 8.4% and 44.7% of the time
  - significant *p*-values between 9.8% and 41.5% of the time
- As the rates of the over- or under-classification increased, the probabilities of both increased odds ratio and significant *p*-values decreased

Future Work

- Cancelling out effect between over- and under-classification
- Impact of various factors such as radiation dose distributions, outcome baselines, risk levels, significance levels, sample sizes, and confounding variables
- Test of different models, such as Poisson model with person-year tables
- Simulation starting with a misclassified initial dataset to understand how it could reflect the true underlying cause of death, providing a more applicable approach for real-world situations

WASHINGTON STATE UNIVERSITY

Thank you!