Death Certificate Misclassification and Risk Models

Stacey L. McComish, *Associate in Research*

United States Transuranium and Uranium Registries

1845 Terminal Drive, Suite 201, Richland, WA 99354

ustur.wsu.edu | s.mccomish@wsu.edu

"Learning from Plutonium and Uranium Workers"
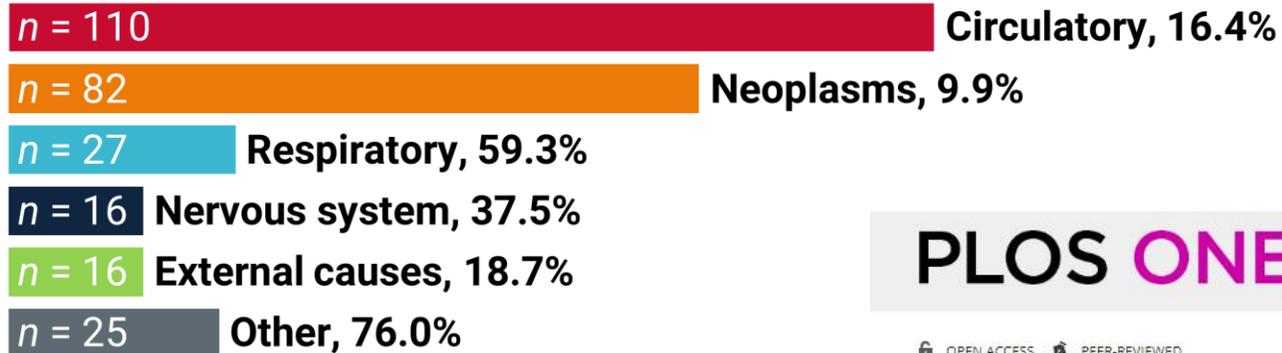
Where we began

- Gold and Kathren (1998) studied causes of death among 260 USTUR Registrants
  - ✓ Incidental finding: 89% had a good or complete match between death certificate and autopsy report
  - ✓ Higher than the published range of 50% to 70%

- As of ~2020, 100 additional Registrants had passed away

- Time to look more closely at the level of agreement between death certificates and autopsy reports among USTUR Registrants

How well did death certificates match autopsy reports?

- 25.5% of death certificates and autopsy reports did not match at the category level

Mismatch rate: Top 5 disease categories

n = 110	Circulatory, 16.4%
n = 82	Neoplasms, 9.9%
n = 27	Respiratory, 59.3%
n = 16	Nervous system, 37.5%
n = 16	External causes, 18.7%
n = 25	Other, 76.0%

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 [co] [✉], Xirui Liu [co], Florencio T. Martinez [co], Joey Y. Zhou [co], Sergey Y. Tolmachev [co]

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

Article	Authors	Metrics	Comments	Media Coverage

How might this misclassification affect epidemiological findings?

- Conventional understanding:

    If the misclassification rate is independent of dose, it will bias findings away from significance (i.e. toward the null hypothesis)

- Which implies that:

    If an observed dose-response relationship is barely significant (or even almost significant), correction of misclassification would just make it more significant
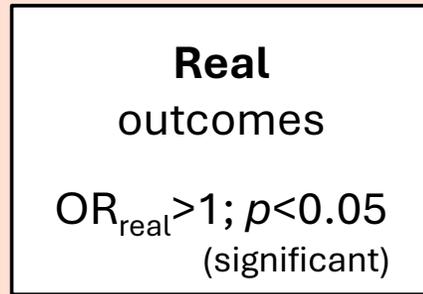
Paper in preparation asks two questions

- Can misclassification of disease bias dose-response relationships away from the null?

- Can it bias a relationship far enough away from the null that that the conclusion of the study is erroneously changed from non-significant to significant?
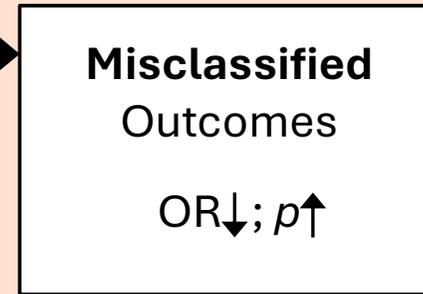
Basic approach to investigating these questions

Conventional heuristic

Paper findings

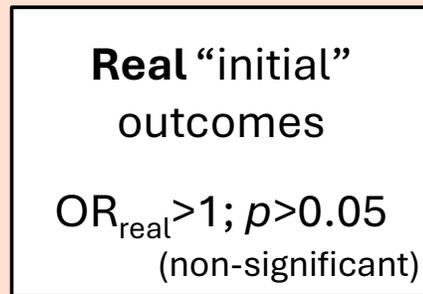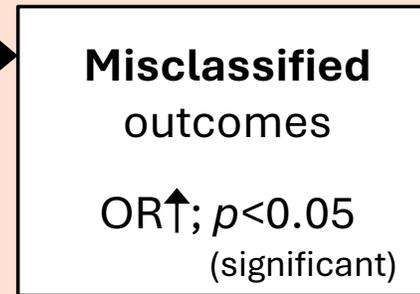| Real outcomes	Misclassified Death Certificates	Misclassified Outcomes
$OR_{real} > 1$; $p < 0.05$ (significant)	→	$OR\downarrow$; $p\uparrow$

Implies that correction of misclassification **would make a significant OR more significant**

Do scenarios exist where the opposite happens?

Real or Generated

Simulated findings

Real "initial" outcomes	Simulated misclassification	Misclassified outcomes
$OR_{real} > 1$; $p > 0.05$ (non-significant)	→	$OR\uparrow$; $p < 0.05$ (significant)

Implies that correction of misclassification **may not make a significant OR more significant**

A couple of definitions

- ## False positive rate

$$Over - classification\ rate = \frac{\#\ False\ Positives}{False\ Positives + True\ Negatives}$$

- ## False negative rate
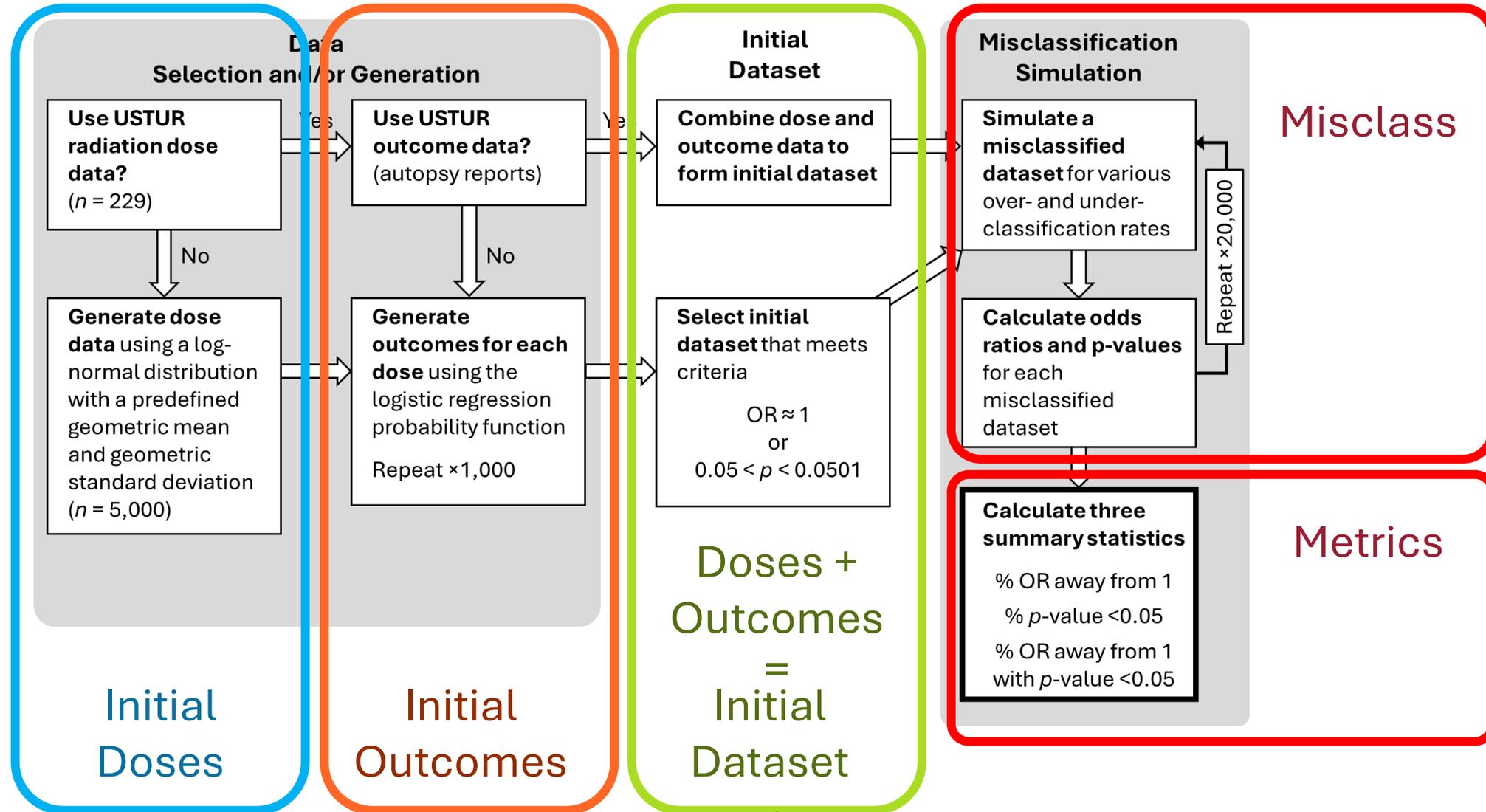
$$Under - classification\ rate = \frac{\#\ False\ Negatives}{False\ Negatives + True\ Positives}$$

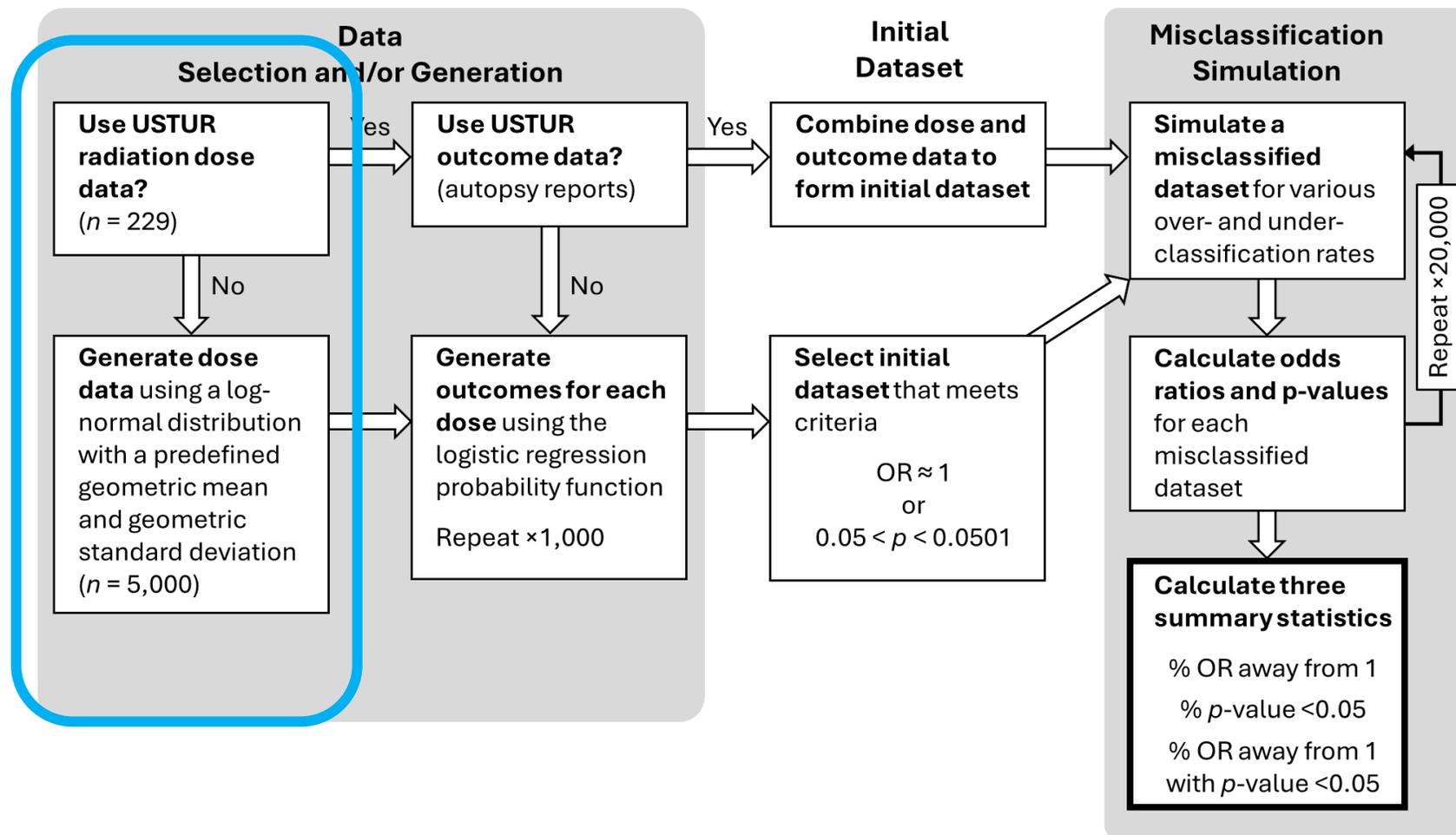- ## Outcome of interest is cancer for this study, though the concept can be generalized to other diseases
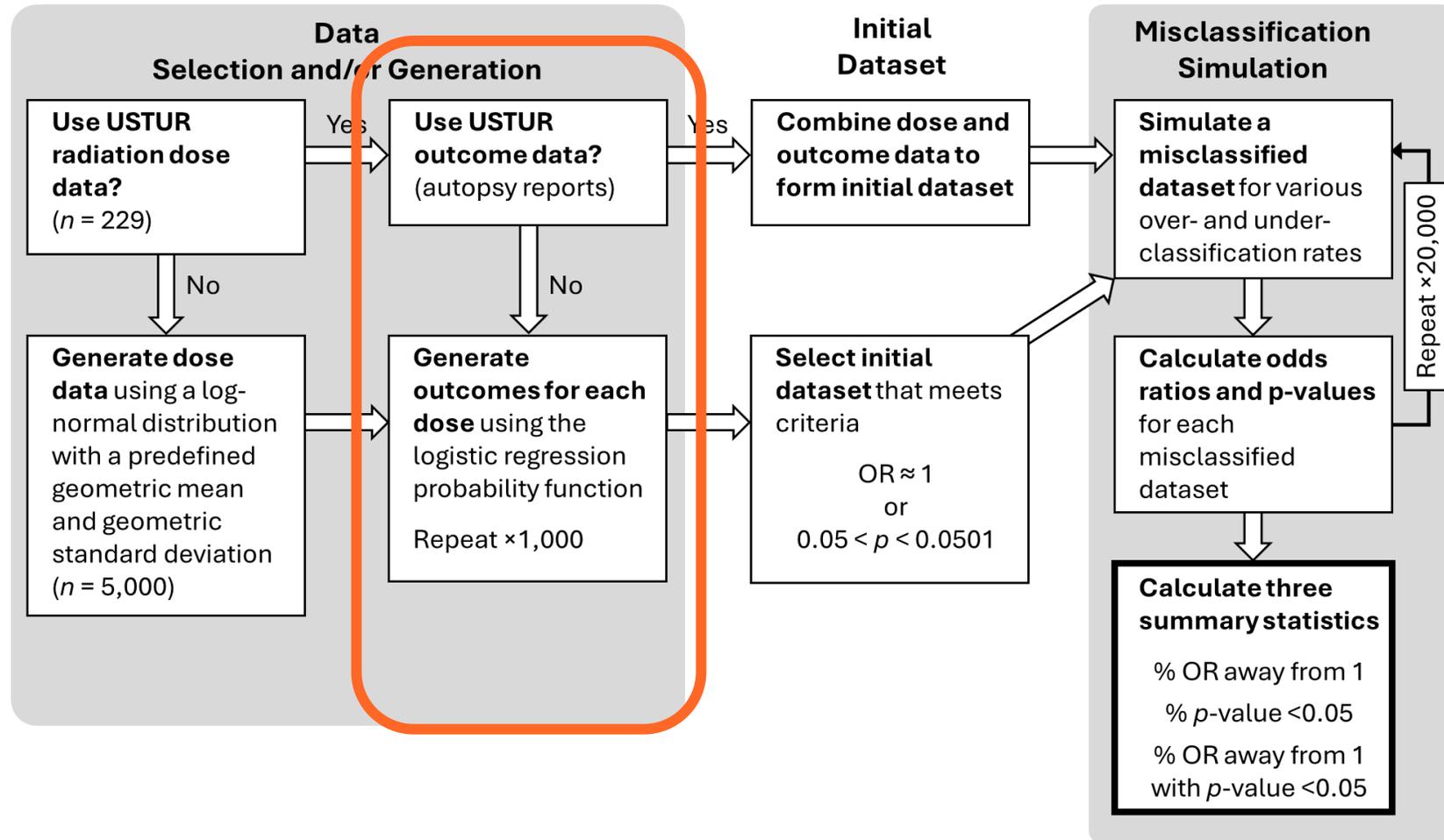
Basic Methods

Data Selection and/or Generation

Use USTUR radiation dose data?
($n = 229$)

Yes →

Generate dose data using a log-normal distribution with a predefined geometric mean and geometric standard deviation
($n = 5{,}000$)

No ↓

Use USTUR outcome data?
(autopsy reports)

Yes →

Generate outcomes for each dose using the logistic regression probability function

Repeat ×1,000

No ↓

Initial Dataset

Combine dose and outcome data to form initial dataset

Select initial dataset that meets criteria

$OR \approx 1$
or
$0.05 < p < 0.0501$

Misclassification Simulation

Simulate a misclassified dataset for various over- and under-classification rates

Calculate odds ratios and p-values for each misclassified dataset

Repeat ×20,000

Calculate three summary statistics

% OR away from 1

% *p*-value <0.05

% OR away from 1 with *p*-value <0.05

Misclass

Metrics

Initial Doses

Initial Outcomes

Doses + Outcomes = Initial Dataset

Represents actual distribution of disease in a population

WASHINGTON STATE UNIVERSITY

Multiple initial scenarios explored

Data Selection and/or Generation

Use USTUR radiation dose data? ($n = 229$)

— Yes → **Use USTUR outcome data?** (autopsy reports)

— No → **Generate dose data** using a log-normal distribution with a predefined geometric mean and geometric standard deviation ($n = 5,000$)

Use USTUR outcome data? (autopsy reports)

— Yes → **Combine dose and outcome data to form initial dataset**

— No → **Generate outcomes for each dose** using the logistic regression probability function

Repeat ×1,000

Initial Dataset

Combine dose and outcome data to form initial dataset

Select initial dataset that meets criteria

OR ≈ 1
or
$0.05 < p < 0.0501$

Misclassification Simulation

Simulate a misclassified dataset for various over- and under-classification rates

Calculate odds ratios and p-values for each misclassified dataset

Repeat ×20,000

Calculate three summary statistics

% OR away from 1

% p-value <0.05

% OR away from 1 with p-value <0.05

USTUR external doses vs. generated doses

Multiple initial scenarios explored

USTUR underlying causes of death vs. generated outcomes
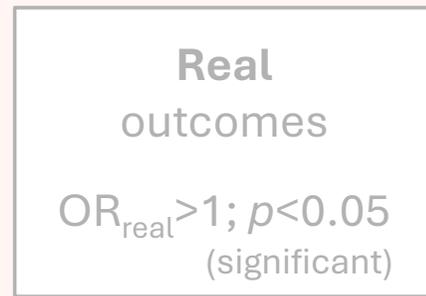
Multiple initial scenarios explored

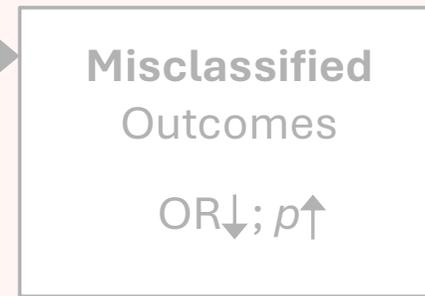Borderline scenarios: OR≈1 or $p_{initial}$ barely non-significant

Basic approach to investigating these questions

Conventional heuristic

| **Real** outcomes $OR_{real}>1$; $p<0.05$ (significant) | → Misclassified Death Certificates → | Paper findings **Misclassified** Outcomes $OR\downarrow$; $p\uparrow$ |

Implies that correction of misclassification **would make a significant OR more significant**

Do scenarios exist where the opposite happens?
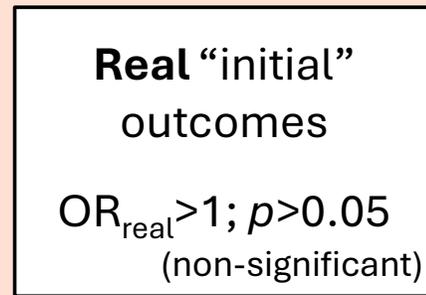
| Real or Generated **Real** "initial" outcomes $OR_{real}>1$; $p>0.05$ (non-significant) | → Simulated misclassification → | Simulated findings **Misclassified** outcomes $OR\uparrow$; $p<0.05$ (significant) |

Implies that correction of misclassification **may not make a significant OR more significant**

WASHINGTON STATE UNIVERSITY
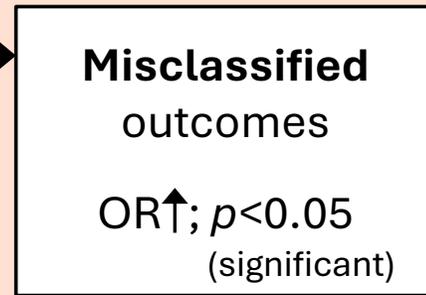
# Results: Actual USTUR doses and outcomes

- USTUR external doses, autopsy reports (initial outcomes), death certificates (misclassified outcomes)

- *p*-values are so large that odds ratios for both scenarios are the same (statistically speaking)

| Scenario	Calculation method	True (AR[a])		Misclassified (DC[b])	
		OR	*p*-value	OR	*p*-value
1	2 x 2 table[c]	0.708	0.260	0.700	0.253
2	logistic regression	0.360	0.326	0.530	0.540

Results: Actual USTUR doses and outcomes

- USTUR external doses, USTUR autopsy reports, simulated misclassification on death certificates

- $OR_{init}=0.708$; $p_{init}=0.260$ (actual USTUR)
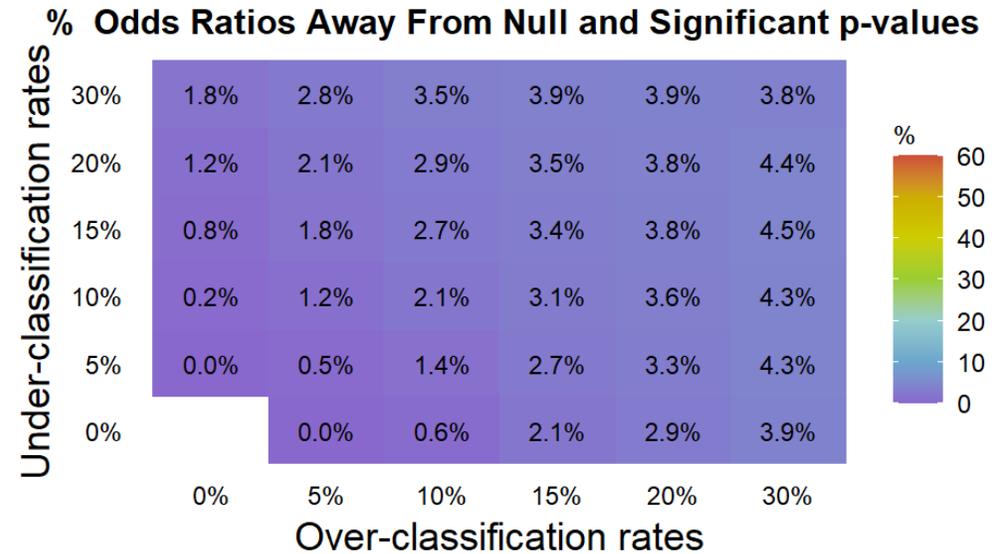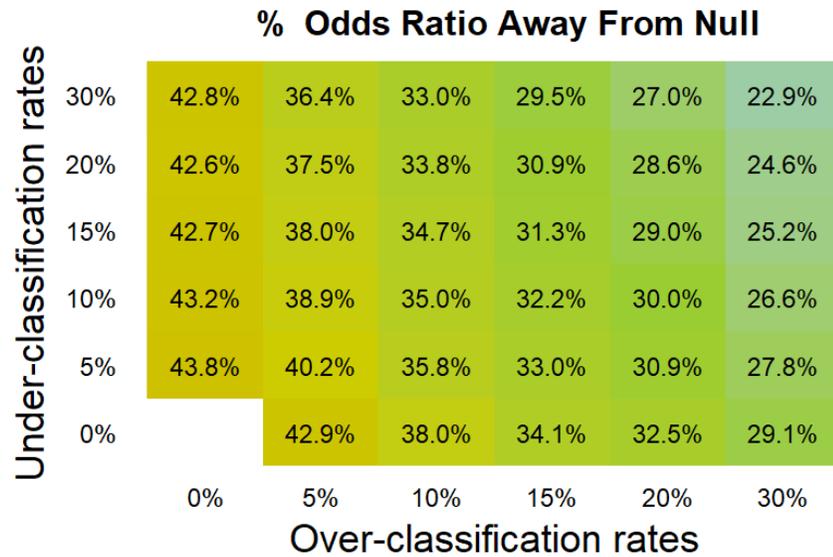
Results: Generated doses and outcomes

- Generated external doses, generated initial "real" outcomes, simulated death certificate misclassification

- $OR_{init}=1.0003$; $p_{init}=0.999$ (OR≈1)
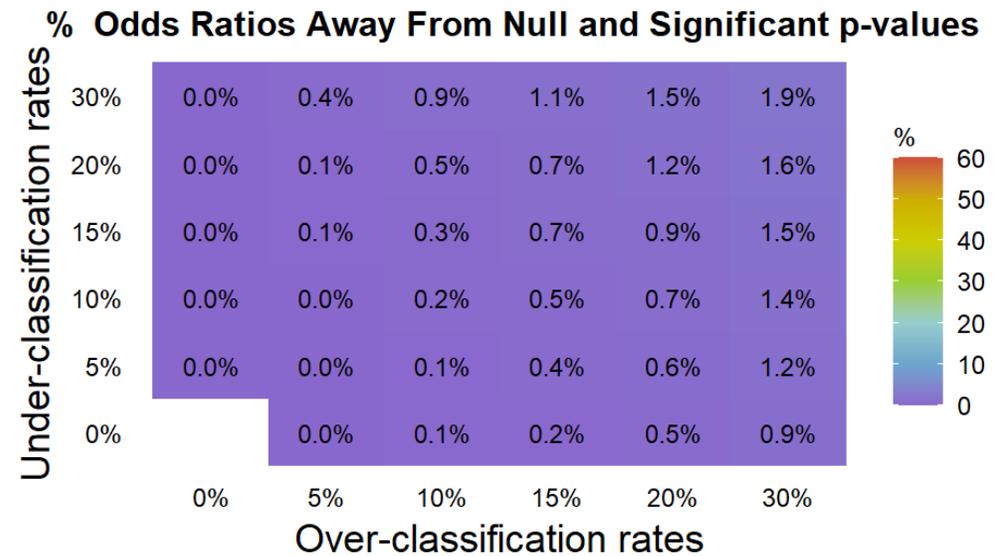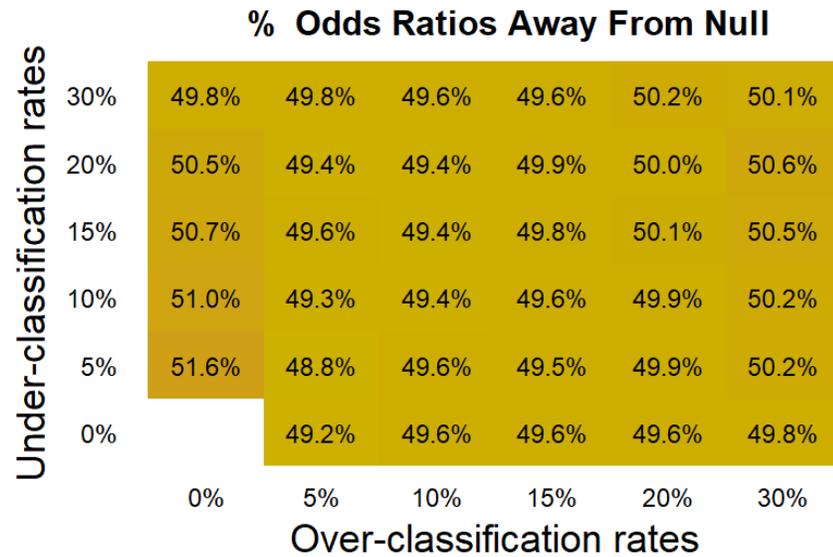
Results: Generated doses and outcomes

- Generated external doses, generated initial "real" outcomes, simulated death certificate misclassification

- $OR_{init}=1.47$; $p_{init}=0.05006$ (p slightly non-significant)
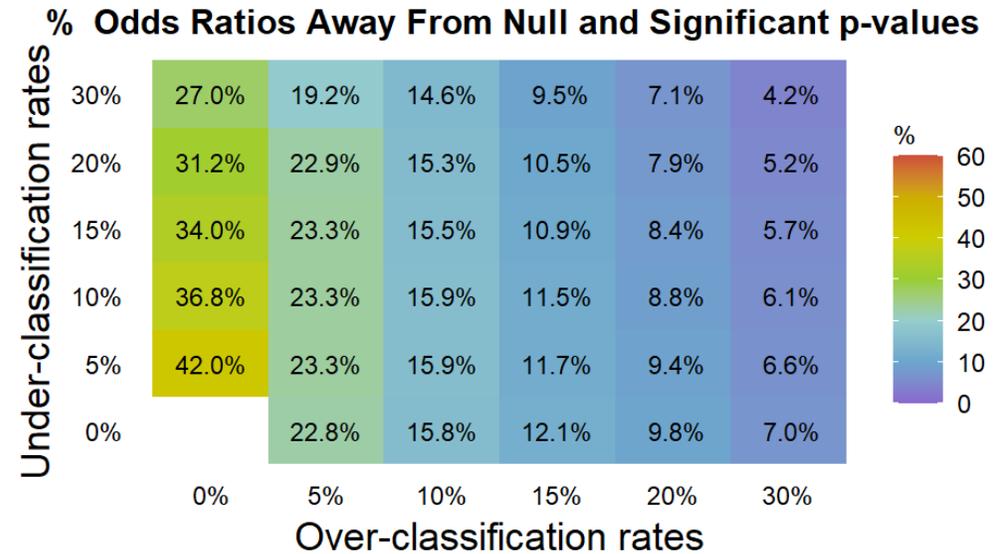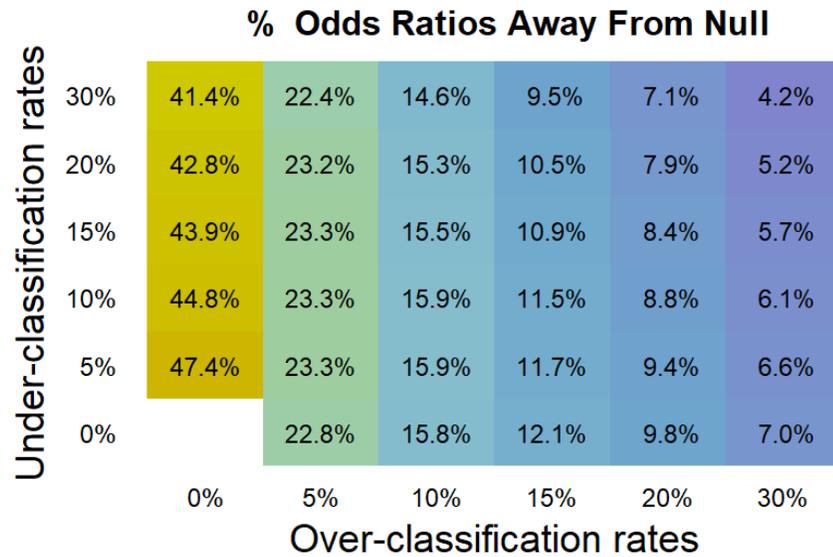
One explanation: the nature of random chance

Coin Toss

Heads = 50%
Tails = 50%

If I toss 10 quarters, I will not always get 5 heads

Misclassified Outcomes

Low dose = 5%
High dose = 5%

OR_{init}=0.70
p=0.25

For 5% overall misclassification, it won't always work out to exactly 5% in high dose cases and 5% in low dose cases

Real dataset (under-misclassification):
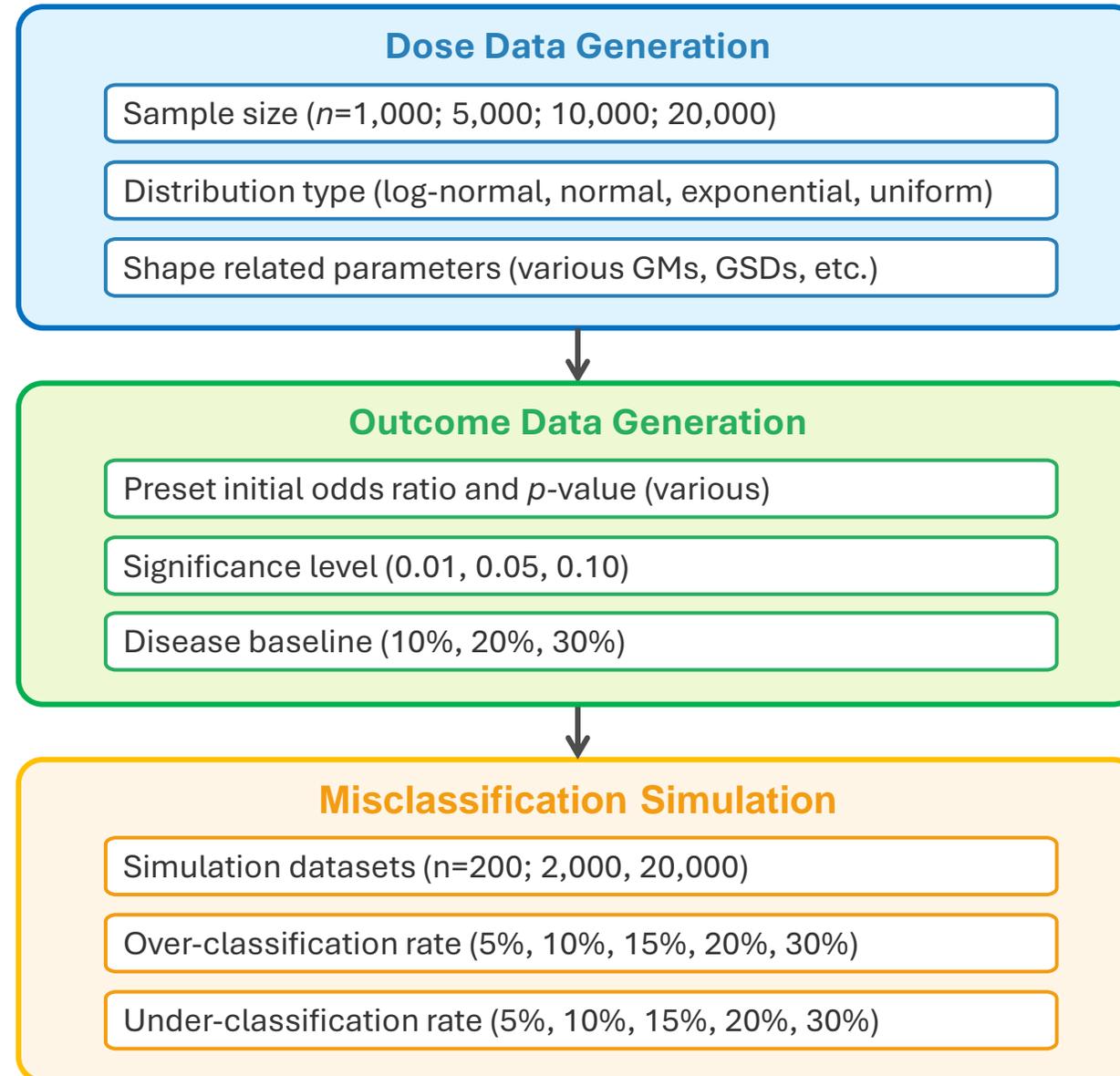   Total: 5.06%
    Low dose: 8.00%
    High dose: 2.41%

OR_{mis}=0.55↓
p=0.049

*Non-differential misclassification is sometimes functionally differential

Other factors come into play: a work in progress

Dose Data Generation

Sample size (n=1,000; 5,000; 10,000; 20,000)

Distribution type (log-normal, normal, exponential, uniform)

Shape related parameters (various GMs, GSDs, etc.)

Outcome Data Generation

Preset initial odds ratio and p-value (various)

Significance level (0.01, 0.05, 0.10)

Disease baseline (10%, 20%, 30%)

Misclassification Simulation

Simulation datasets (n=200; 2,000, 20,000)

Over-classification rate (5%, 10%, 15%, 20%, 30%)

Under-classification rate (5%, 10%, 15%, 20%, 30%)

WASHINGTON STATE UNIVERSITY

Questions? Thoughts?