WASHINGTON STATE UNIVERSITY
College of Pharmacy and Pharmaceutical Sciences

2026 USTUR Scientific Advisory Committee Meeting
Hampton Inn, Richland, Washington; April 16–17, 2026

Quantification of Outcome Misclassification

Xirui Liu, *Associate in Research*

United States Transuranium and Uranium Registries

1845 Terminal Drive, Suite 201, Richland, WA 99354

ustur.wsu.edu | xirui.liu@wsu.edu

Study Overview

Phase 3

Correction

Correct outcome misclassification to improve study accuracy

Phase 2

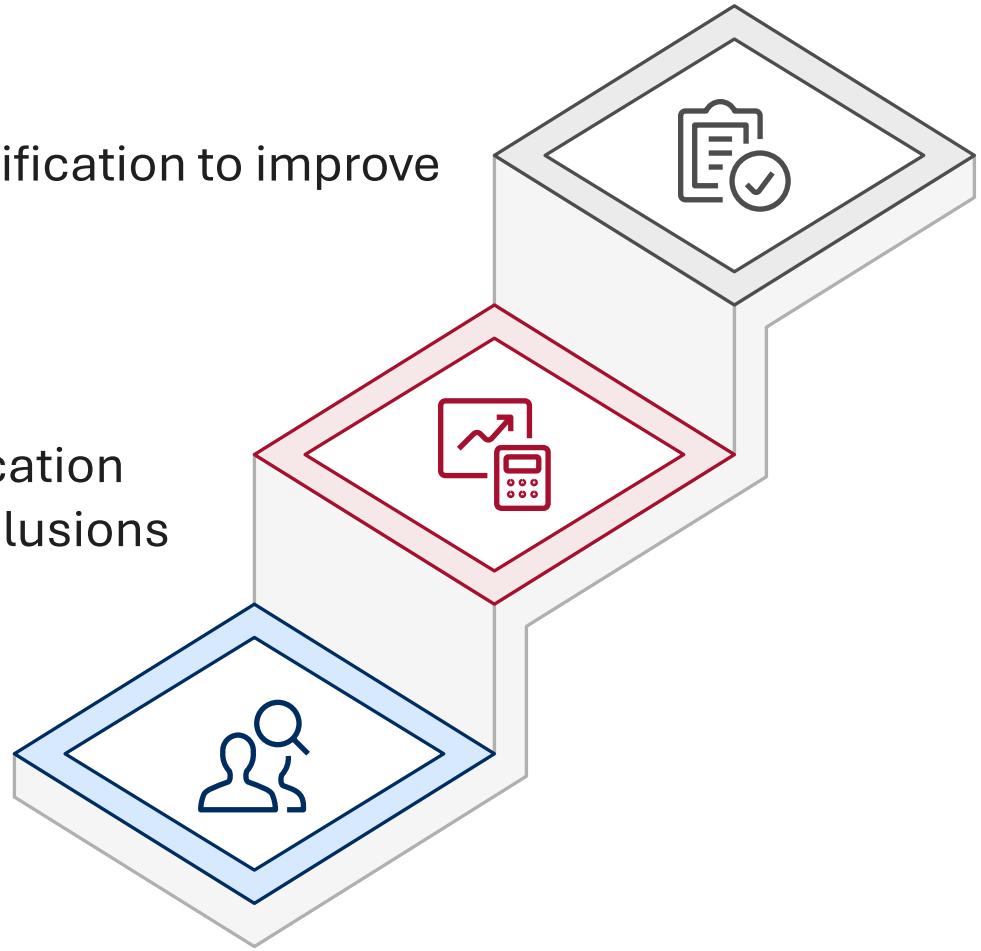
Impact

Quantify how outcome misclassification impact epidemiological study conclusions

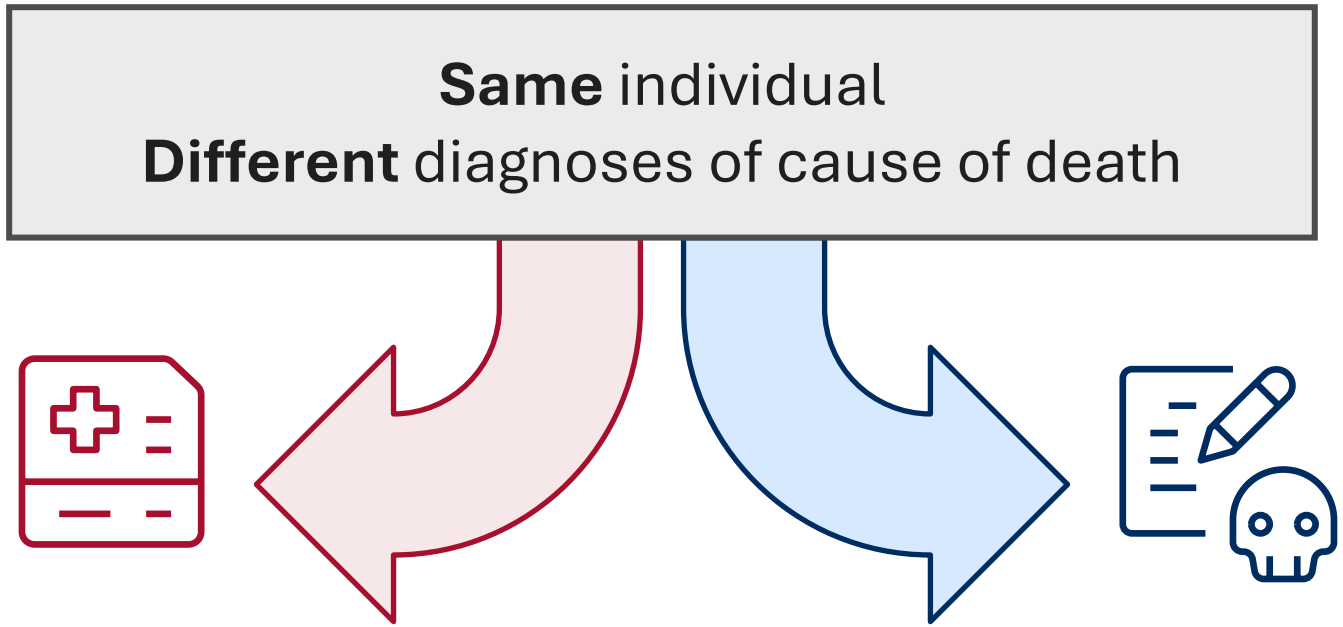
Phase 1

Identification

Identify outcome misclassification among USTUR registrants



Phase 1: Identification



Death Certificate

- Based on clinical symptoms
- Commonly used in epidemiological studies

Autopsy Report

- Based on postmortem findings
- Considered the gold standard for underlying cause of death

Phase 1: Published Findings


PLOS One

Publish About Browse Search advanced search

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

0 Save | 1 Citation | 1,150 View | 0 Share

ICD-10 Chapters		n^b	Matches	Over-	Under-
2	Neoplasms	80	72 (90.0%)	8 (4.3%)	11 (13.3%)
6	Nervous system	15	10 (66.7%)	5 (2.0%)	3 (23.1%)
9	Circulatory	108	90 (83.3%)	18 (12.2%)	31 (25.6%)
10	Respiratory	26	10 (38.5%)	16 (6.4%)	9 (47.4%)
20	External causes	15	12 (80.0%)	3 (1.2%)	1 (7.7%)
	Other	24	6 (25.0%)	n/a	n/a
	All Causes	268	200 (74.6%)	n/a	n/a

[a]Number of false positives and false negatives, with the corresponding over- and under-classification rates in parentheses.
[b]Number of death certificates with UCODs in each chapter.

Phase 2: Research Question

Quantify outcome misclassification impact on epidemiology study conclusion

Phase 2: Study Design

Quantify **outcome misclassification** **impact** on **epidemiology study** **conclusions**

Quantify

- Maximum possible impact
  - start from boundary scenario
  - measure percentage

Outcome misclassification

- Over misclassification
- Under misclassification
- Combined effects

Impact

- Odds ratio: increase
- *p*-value: non-significant → significant

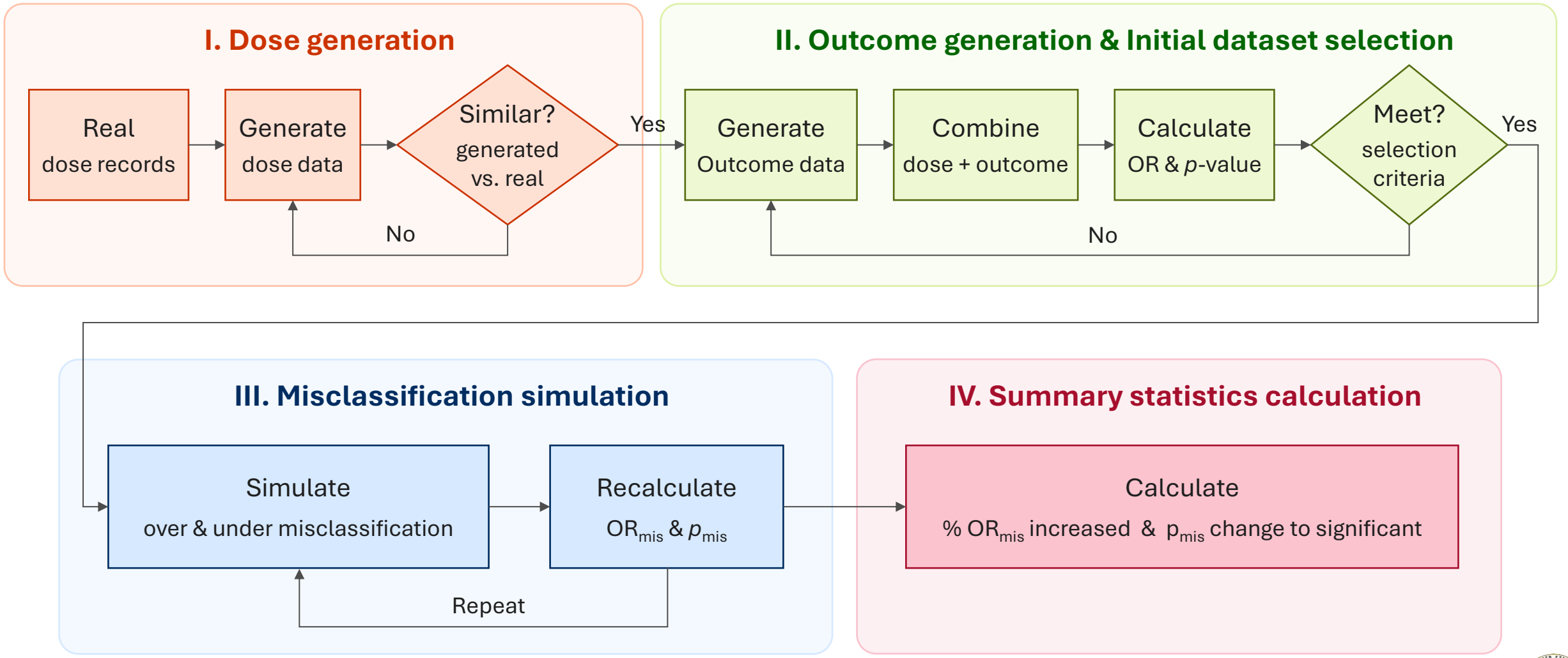
Epidemiology study

- Independent variable: radiation dose exposure
- Dependent variable: cancer cause of death

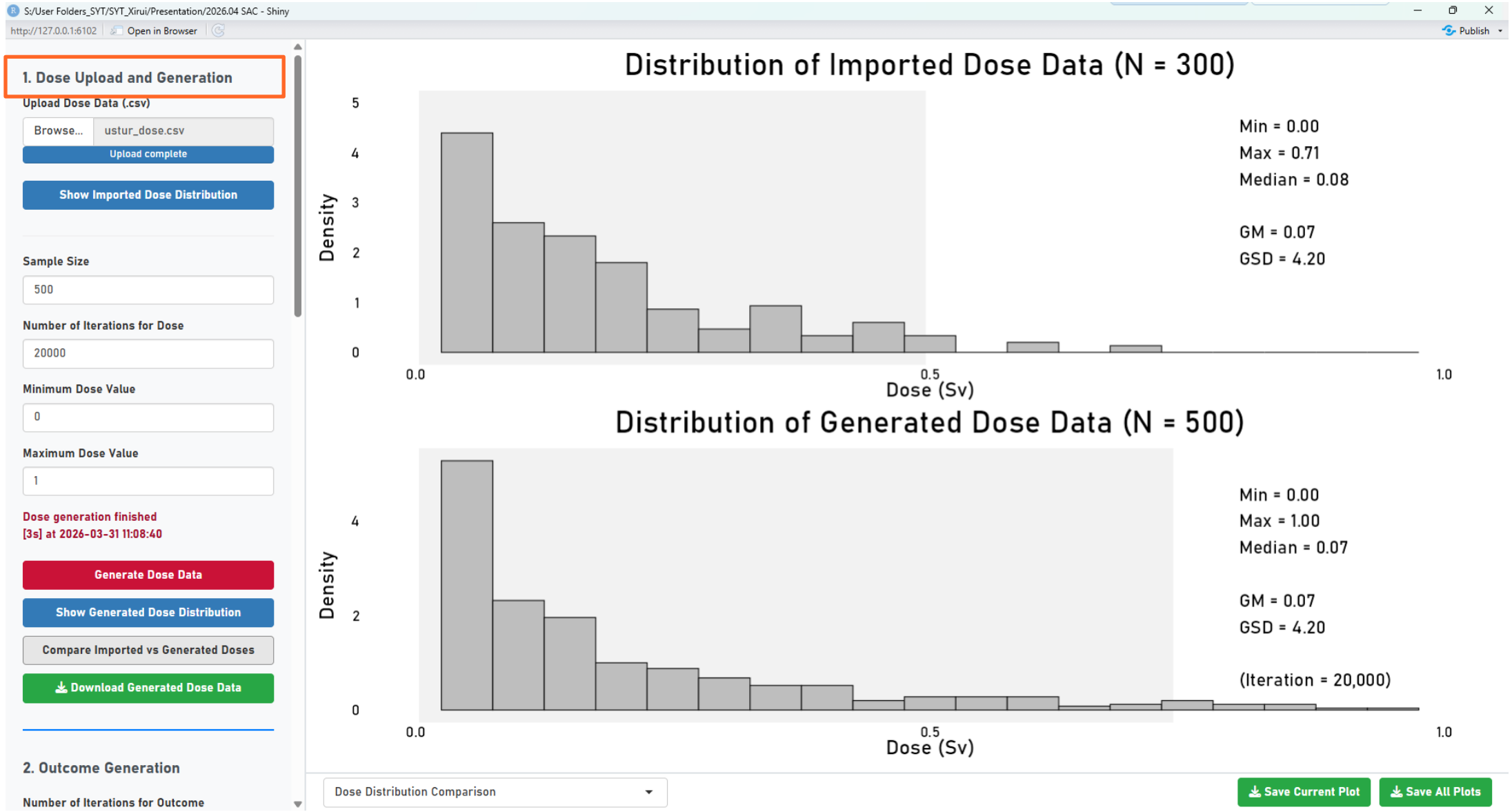
Conclusion

- Odds ratio
- *p*-value

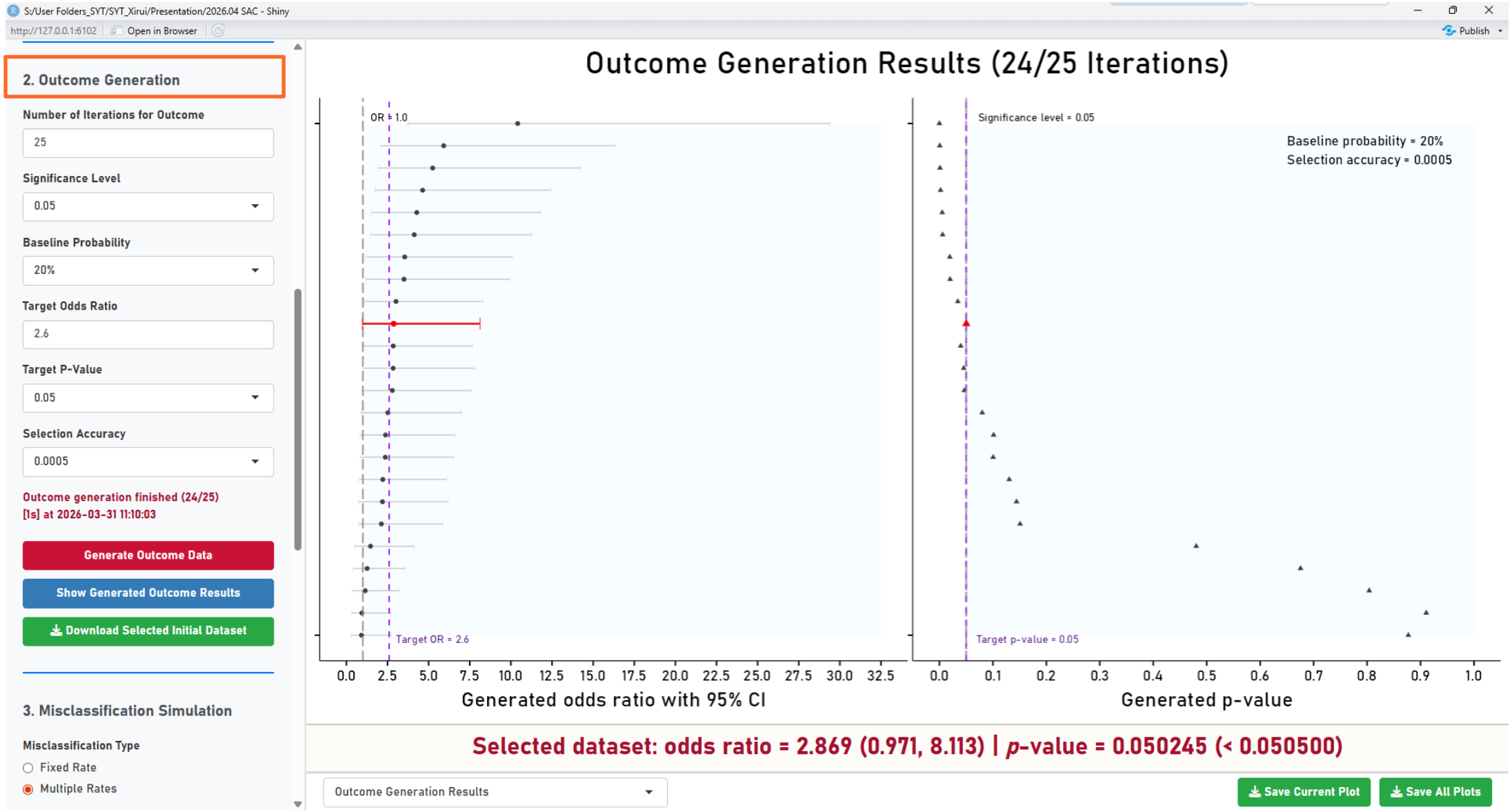
Phase 2: Methodology



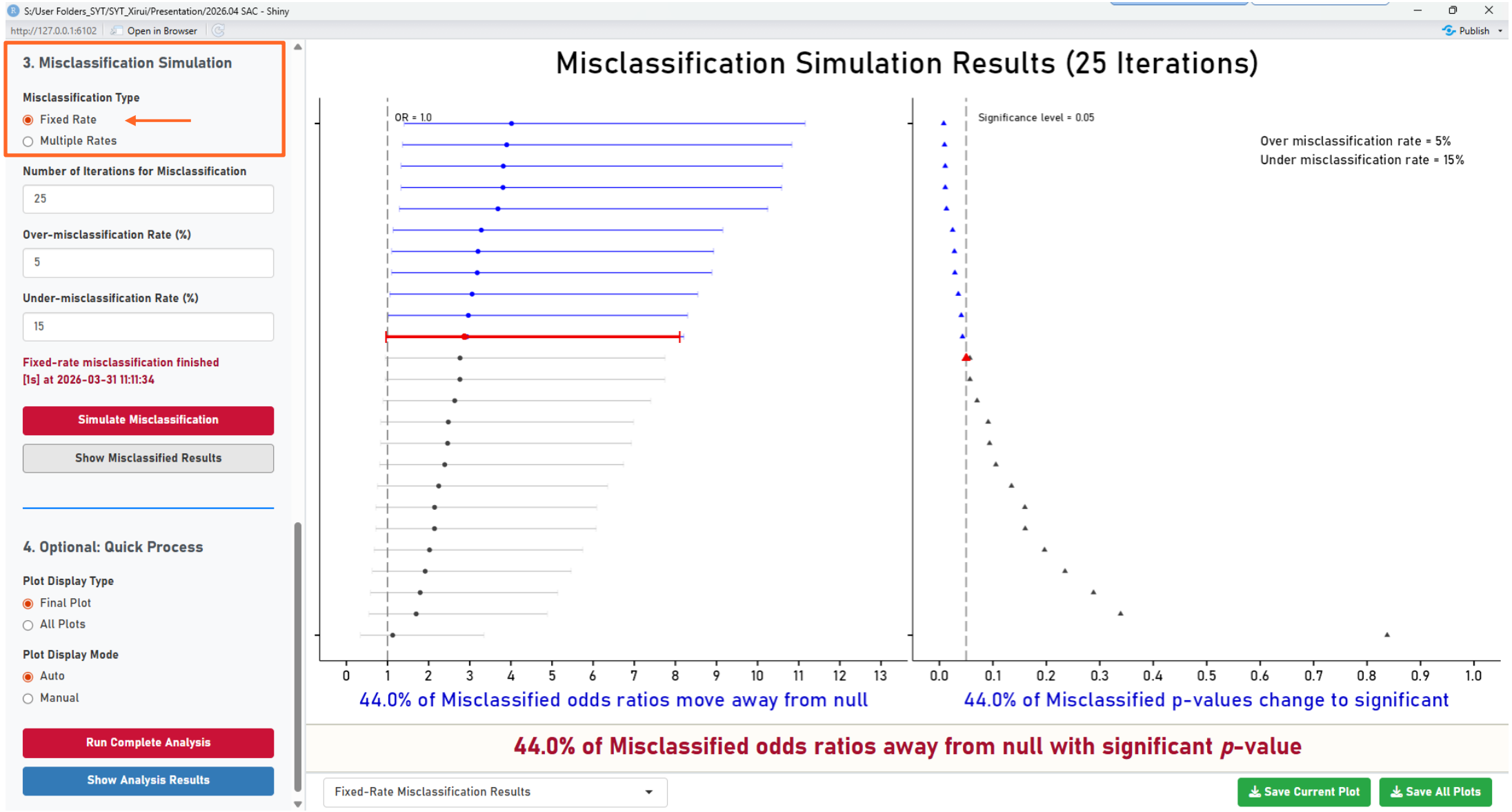
R Shiny Tool: Dose Upload and Generation



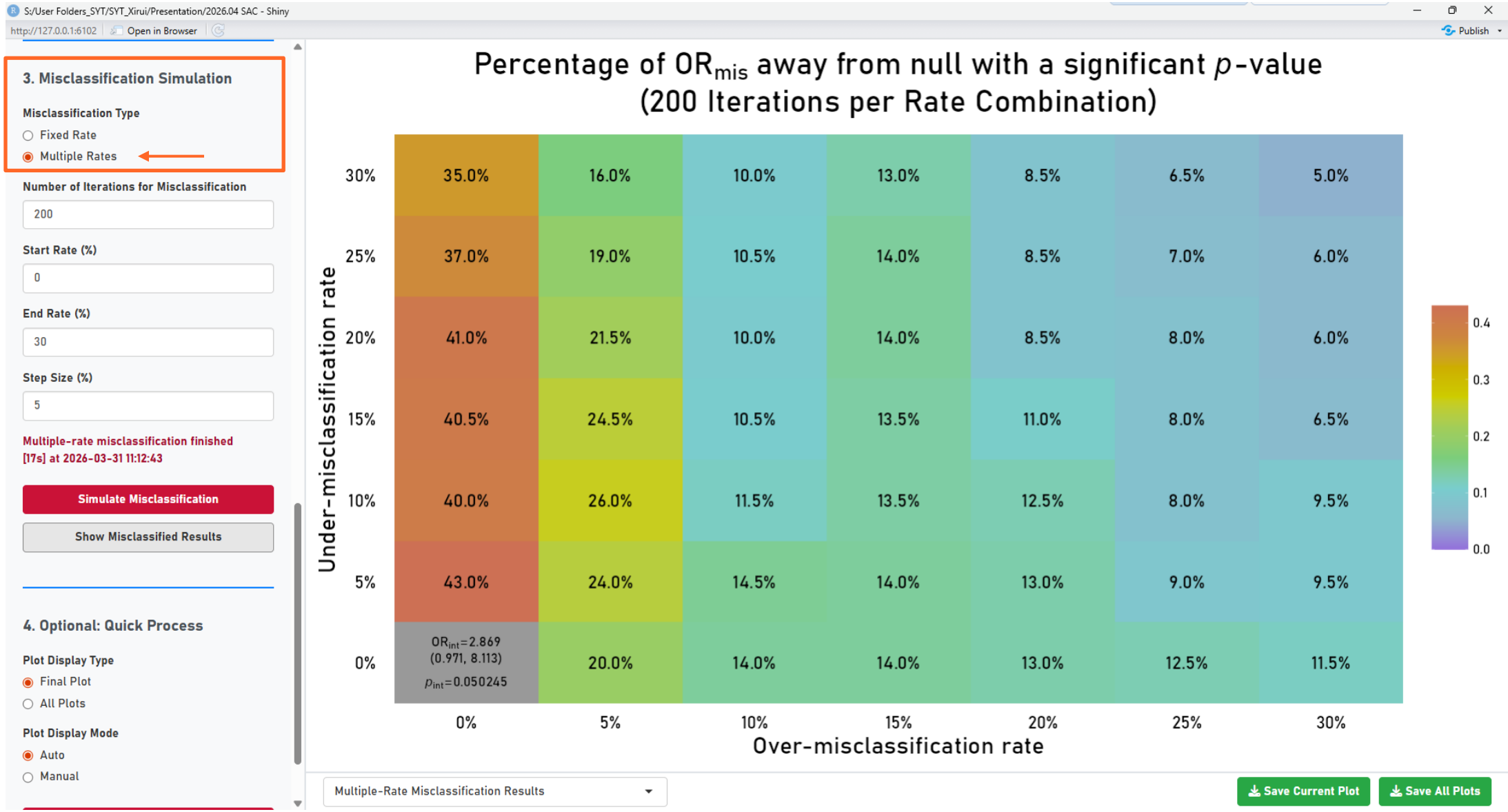
R Shiny Tool: Outcome Generation and Selection



R Shiny Tool: Misclassification Simulation



R Shiny Tool: Statistic Calculation



Phase 2: Published Findings

General understanding

- Non-differential misclassification biases dose-response association **toward the null**

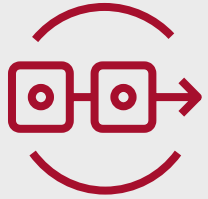
Our finding

- **Non-trivial probability** that outcome misclassification moves an individual dose-response association **away from the null**
  - even changing conclusions **from non-significant to significant**
- In **borderline scenarios**, up to **42.8%** of conclusions could potentially be impacted by the outcome misclassification
- Impact was **more sensitive** to **low misclassification rates**

Phase 2: Factor Analysis

Sample Size

200 → 5,000



Number of simulations

50 → 5,000

Disease baseline probability

10% → 20% → 30%

Significance threshold

0.01 → 0.05 → 0.10

Phase 2 Factor: Sample Size

Parameter Setting

I. Generate dose

Sample size: 200

II. Generate outcome

Disease baseline: 20%

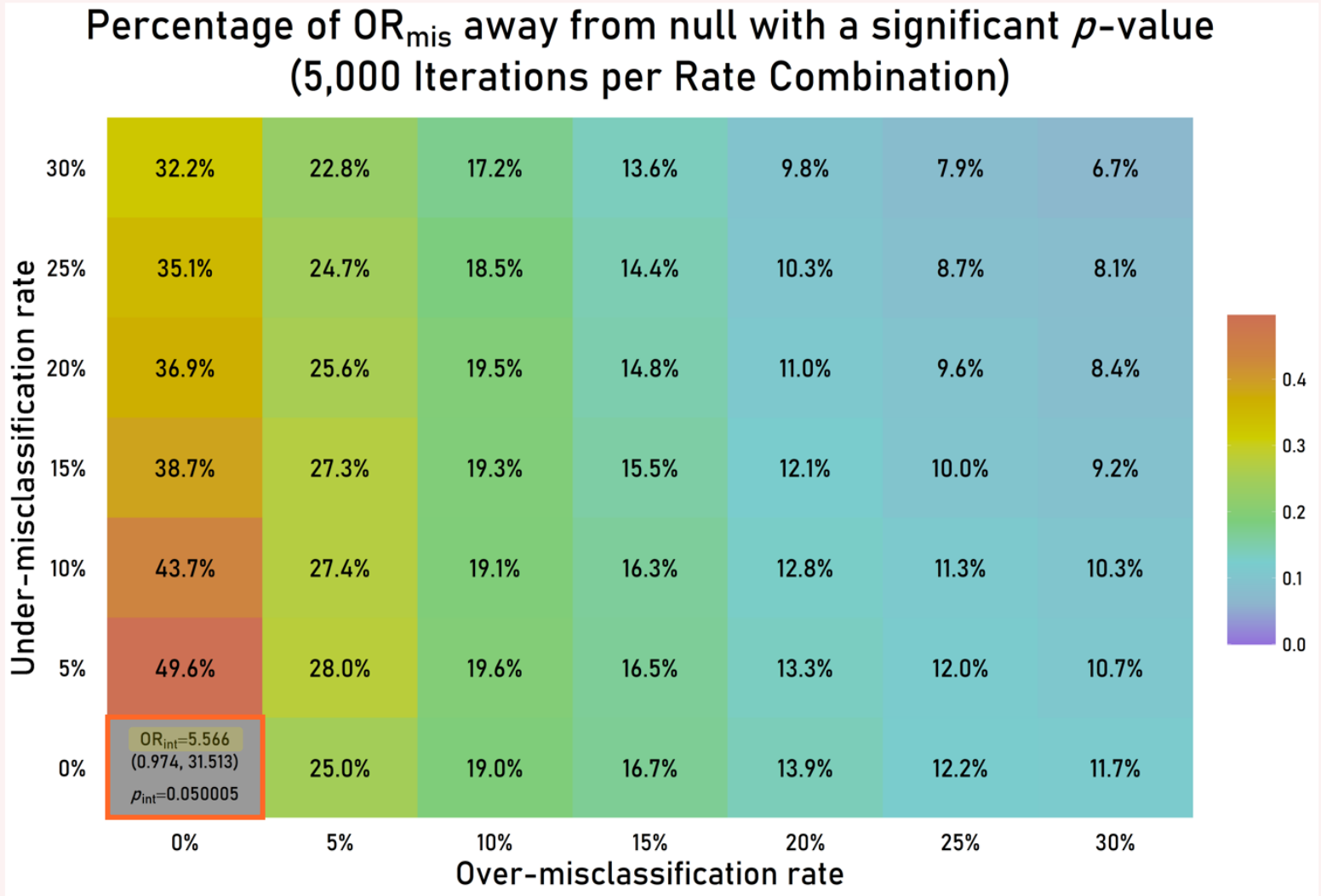
Sig threshold: 0.05

Accuracy: 0.000005

III. Simulate misclassification

Iteration times: 5,000

Heatmap Results



Phase 2 Factor: Sample Size

Parameter Setting

I. Generate dose

Sample size: 5,000

II. Generate outcome

Disease baseline: 20%

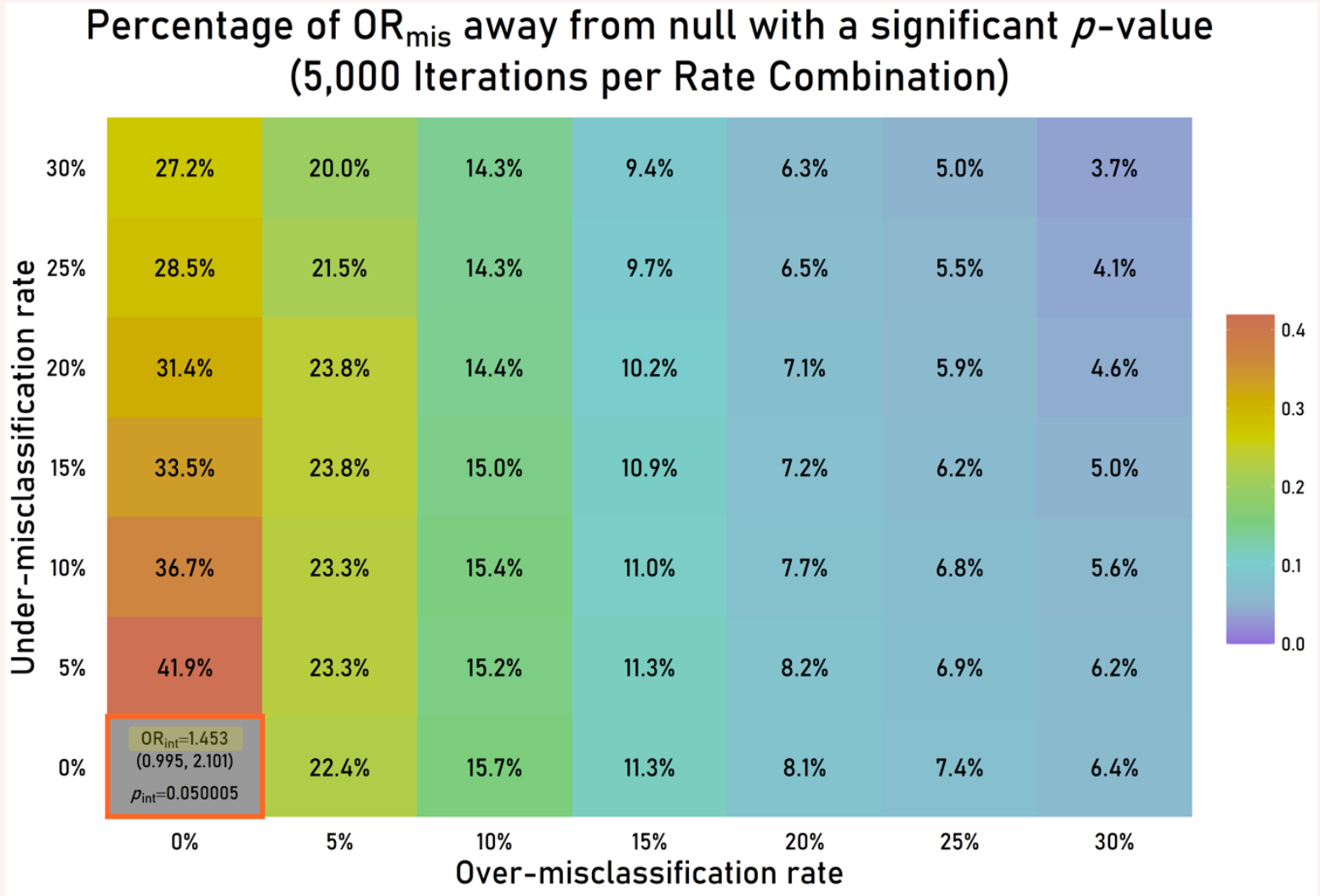
Sig threshold: 0.05

Accuracy: 0.000005

III. Simulate misclassification

Iteration times: 5,000

Heatmap Results



Phase 2 Factor: Number of Simulations

Parameter Setting

I. Generate dose

Sample size: 500

II. Generate outcome

Disease baseline: 20%

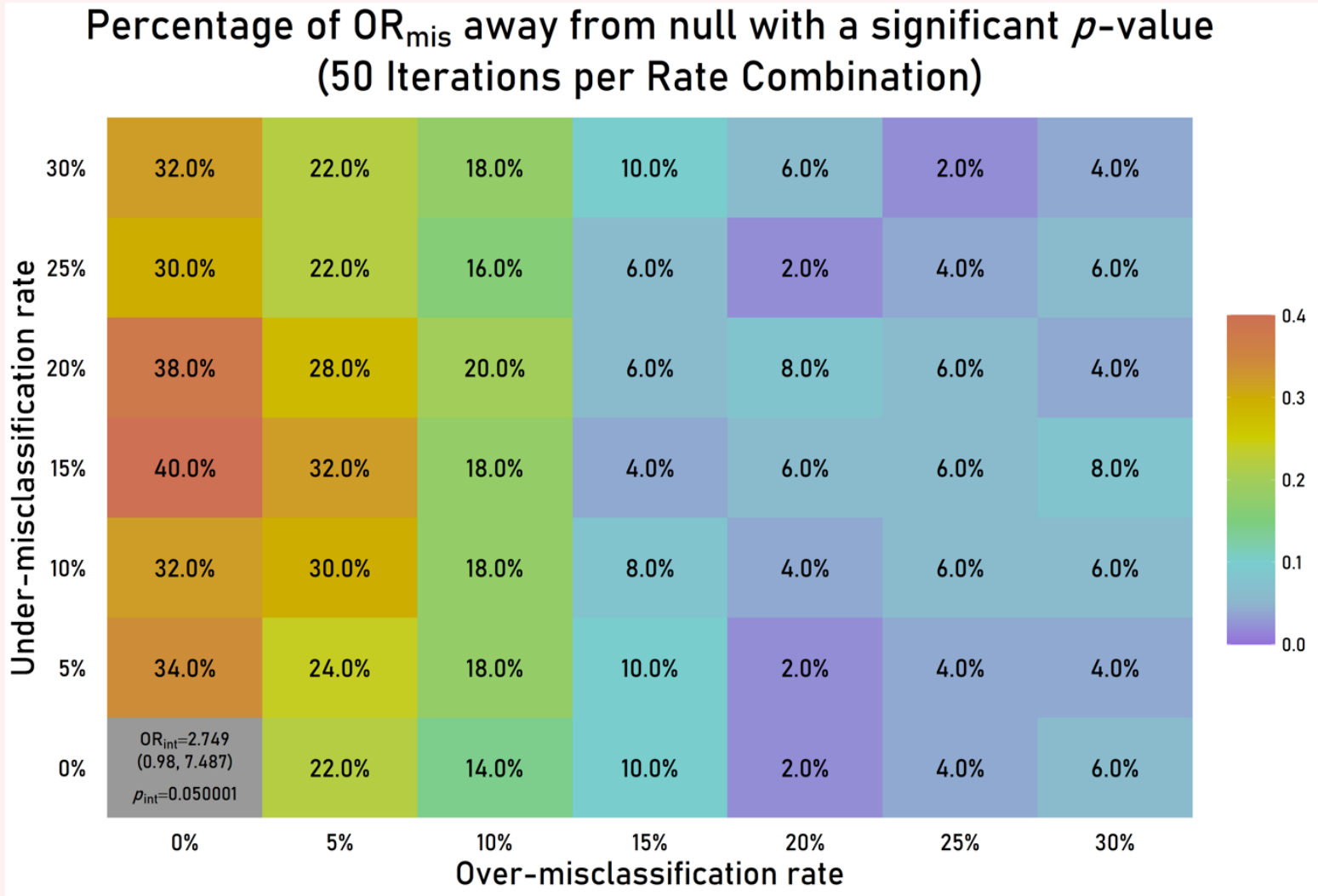
Sig threshold: 0.05

Accuracy: 0.000005

III. Simulate misclassification

Iteration times: 50

Heatmap Results



Phase 2 Factor: Number of Simulations

Parameter Setting

I. Generate dose

Sample size: 500

II. Generate outcome

Disease baseline: 20%

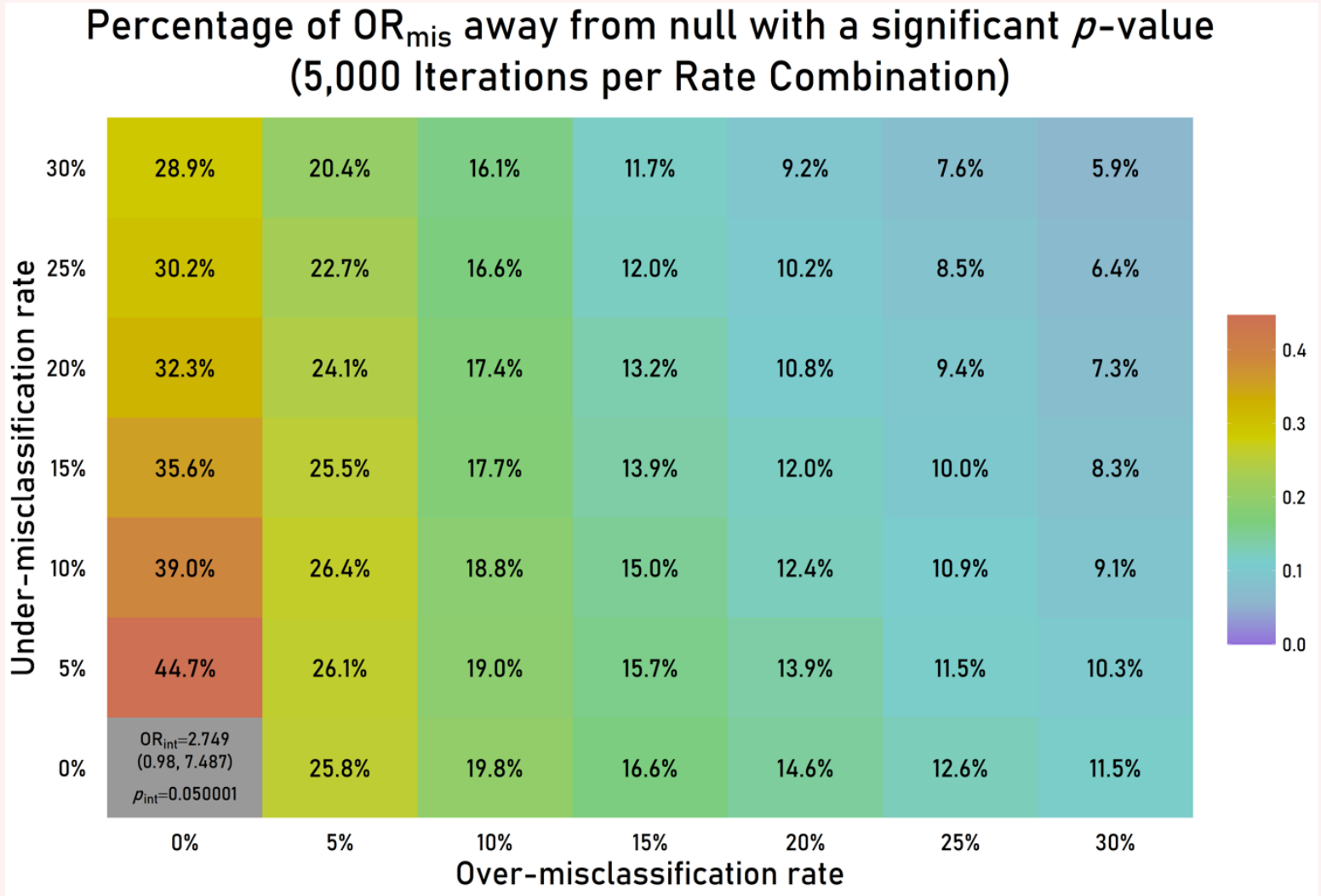
Sig threshold: 0.05

Accuracy: 0.000005

III. Simulate misclassification

Iteration times: 5,000

Heatmap Results



Phase 2 Factor: Baseline Disease Probability

Parameter Setting

I. Generate dose

Sample size: 5,000

II. Generate outcome

Disease baseline: 10%

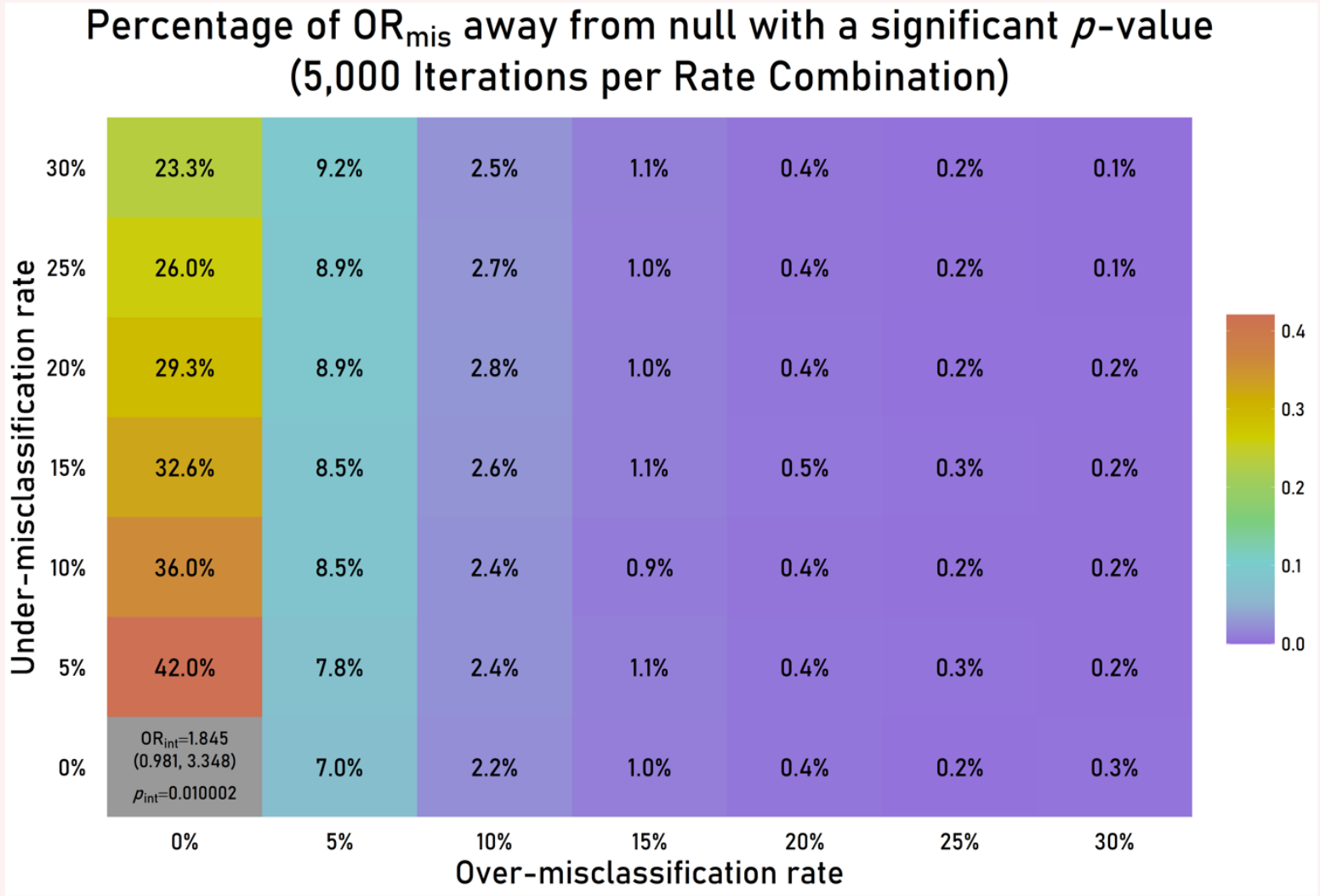
Sig threshold: 0.01

Accuracy: 0.000005

III. Simulate misclassification

Iteration times: 5,000

Heatmap Results



Phase 2 Factor: Baseline Disease Probability

Parameter Setting

I. Generate dose

Sample size: 5,000

II. Generate outcome

Disease baseline: 30%

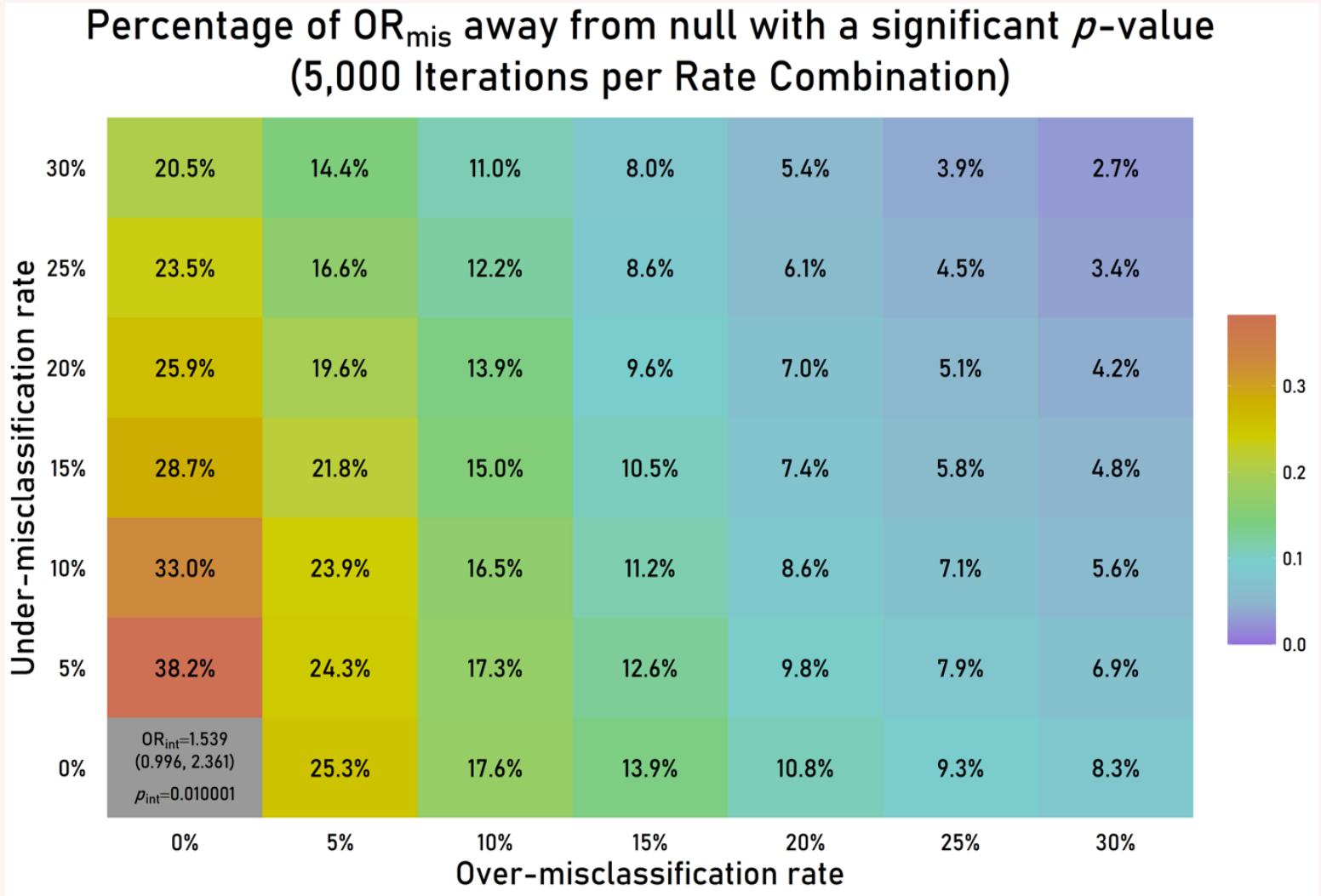
Sig threshold: 0.01

Accuracy: 0.000005

III. Simulate misclassification

Iteration times: 5,000

Heatmap Results



Phase 2 Factor: Significance Threshold

Parameter Setting

I. Generate dose

Sample size: 5,000

II. Generate outcome

Disease baseline: 10%

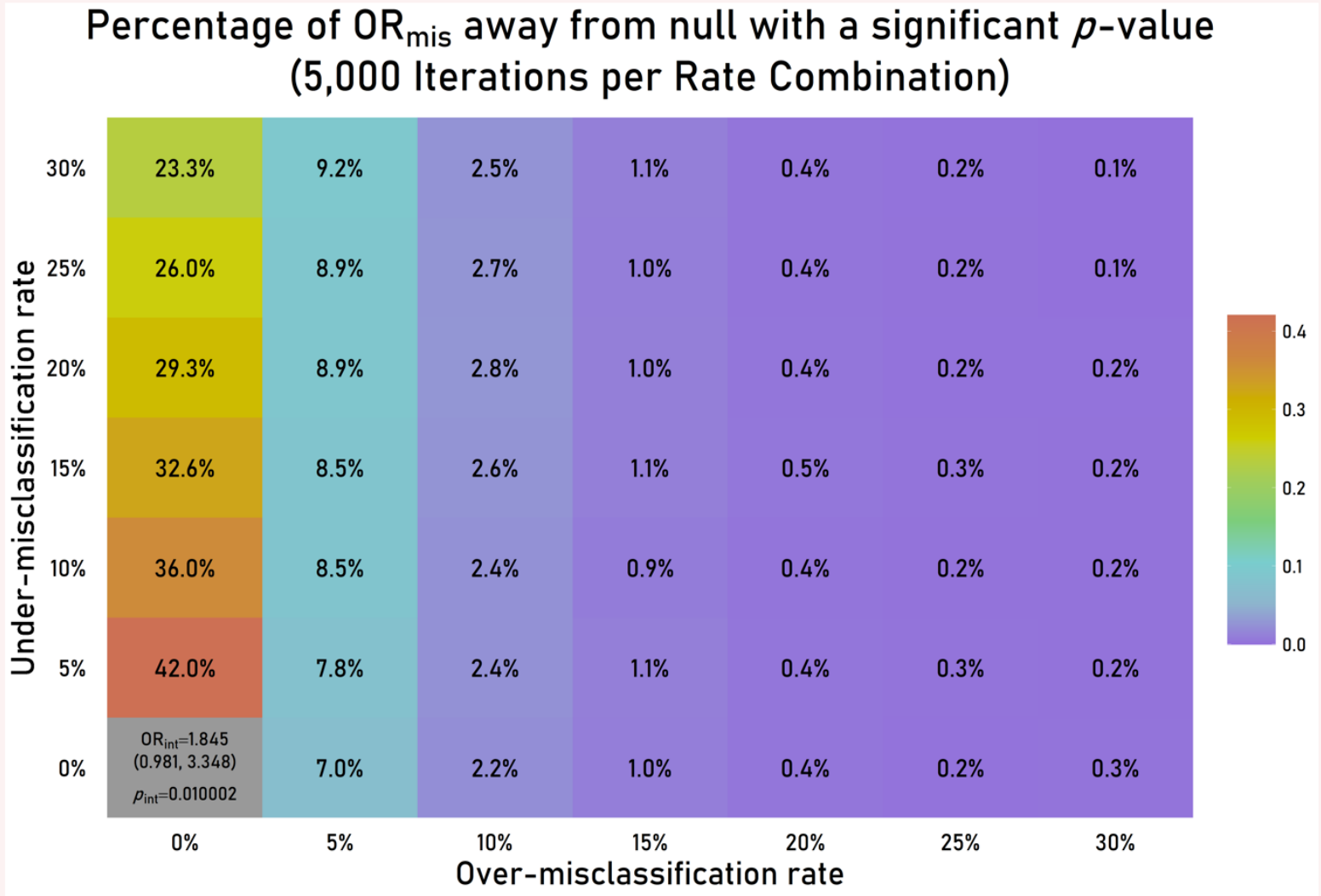
Sig threshold: 0.01

Accuracy: 0.000005

III. Simulate misclassification

Iteration times: 5,000

Heatmap Results



Phase 2 Factor: Significance Threshold

Parameter Setting

I. Generate dose

Sample size: 5,000

II. Generate outcome

Disease baseline: 10%

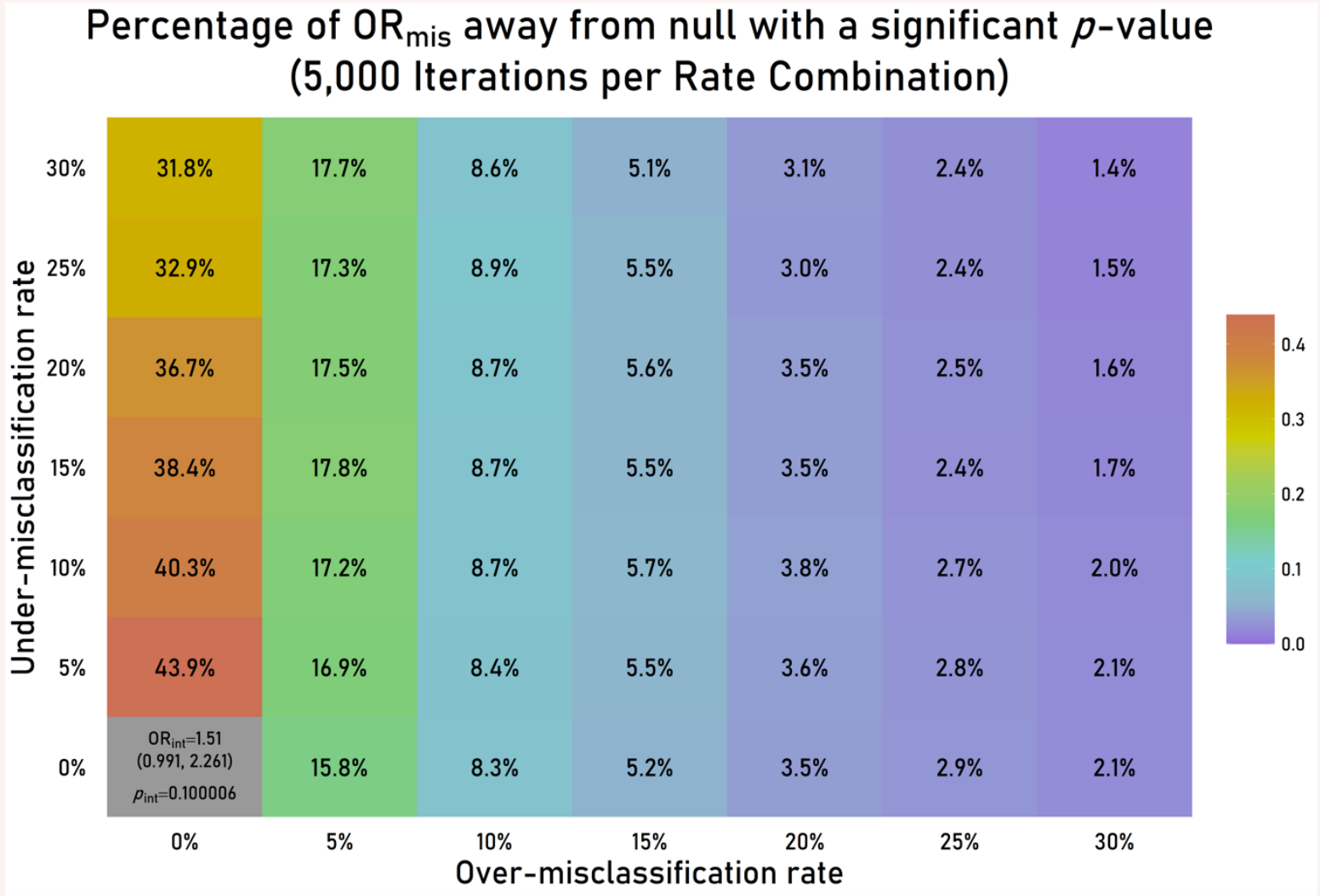
Sig threshold: 0.10

Accuracy: 0.000005

III. Simulate misclassification

Iteration times: 5,000

Heatmap Results



Phase 2: Factor Analysis

Sample Size

200 → 5,000

Initial dataset easier to find and more practically representative

Number of simulations

50 → 5,000

Pattern smoother, cleaner, and more interpretable

Disease baseline probability

10% → 20% → 30%

Higher percentage decreases
Lower percentage increases

Significance threshold

0.01 → 0.05 → 0.10

Percentages increase across all combinations of rates

Summary and Future Work

Phase 2 Impact

- Quantify the misclassification impact on epidemiological study
- Explore factors affect the results

Phase 3 Correction

- Apply correction methods
- Assess whether correction methods work as expected


Thank you! :)