Out of sight, out of mind? A simulation study assessing the use of Quantitative Bias Analysis for Outcome Misclassification in Single-Arm Trials with External Control Comparisons

Putnam
Inizio Advisory

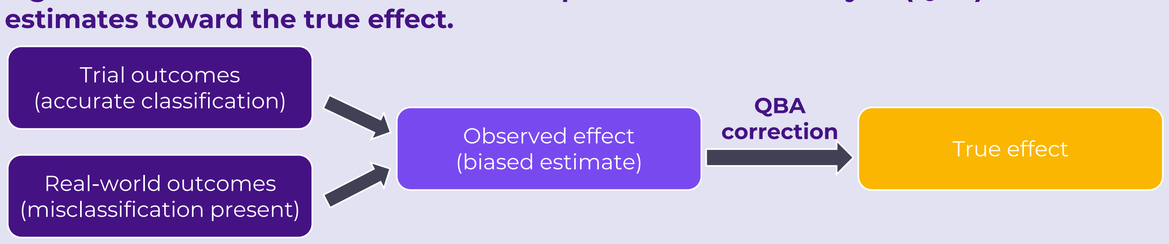
Thomas P. Leahy[1], Sylvaine Barbier[2], Alex J. Turner[3]
[1]Putnam, Toronto, ON, Canada, [2]Putnam, Lyon, France, [3]Putnam, Newcastle Upon Tyne, United Kingdom

RWD132

Introduction

- **Single-arm trials with external controls** are increasingly used in regulatory and HTA submissions.[1]
- **Key challenge**: Outcomes in real-world data may be defined or recorded differently than in trials, creating risk of **differential misclassification**, e.g. in oncology, many studies have identified concordance between physician reported response and response measured using trial definitions.[2]
- **Bias risk**: Even modest outcome misclassification can meaningfully distort estimated treatment effects.[3]
- **Quantitative bias analysis (QBA)** is more commonly used for unobserved confounding,[4] but its application to outcome misclassification remains limited.

Figure 1. Outcome misclassification and quantitative bias analysis (QBA) correct estimates toward the true effect.



- **Objective**: To assess how differential outcome misclassification affects treatment effect estimates in external control arm studies comparing single-arm trial and real-world comparator outcomes, and to evaluate whether deterministic and probabilistic QBA methods can reduce or eliminate misclassification bias.

Methods

- **Simulation setup**: A hypothetical single-arm trial with a single binary endpoint and no outcome misclassification was simulated, alongside external control datasets generated under varying outcome misclassification scenarios.
- **Misclassification scenarios**: Sensitivity and specificity varied to mimic three real-world outcome definitions with increased severity of misclassification.

Table 1. Sensitivity and specificity values used to generate misclassification scenarios for the external control datasets.

	Scenario 1	Scenario 2	Scenario 3
Sensitivity	0.95	0.85	0.60
Specificity	0.98	0.95	0.90

- **Bias correction approaches**:
  - **Deterministic QBA**: Applied correction using assumed sensitivity/specificity.[5]
  - **Probabilistic QBA**: Sensitivity/specificity treated as random variables (e.g., uniform priors); adjustment via Monte Carlo simulation (10,000 iterations).[6]
- **Performance assessment**: We compared corrected estimates against the true effect to quantify bias reduction using the Ratio of Odds Ratios

$$\mathbf{RoOR = OR_{estimate} / OR_{true}}$$

- **True effect:** The simulated "truth" was **OR = 0.53 (95% CI: 0.30–0.94)**.

Results

- **Simulated effects showed that misclassification led to bias towards the null (OR→1) in all scenarios.** Across misclassification scenarios, the observed ORs ranged 0.60–0.84 (larger OR = less apparent benefit) [**Figure 2**].
- **Deterministic QBA reduced bias:** Applying fixed sensitivity/specificity corrections produced ORs ≈ 0.43–0.57, moving estimates back toward the true effect [**Figure 2**].

Figure 2. Simulated odds ratios under increasing outcome misclassification bias compared with bias-adjusted estimates from deterministic and probabilistic quantitative bias analysis. The dashed line indicates the true effect (OR = 0.53).

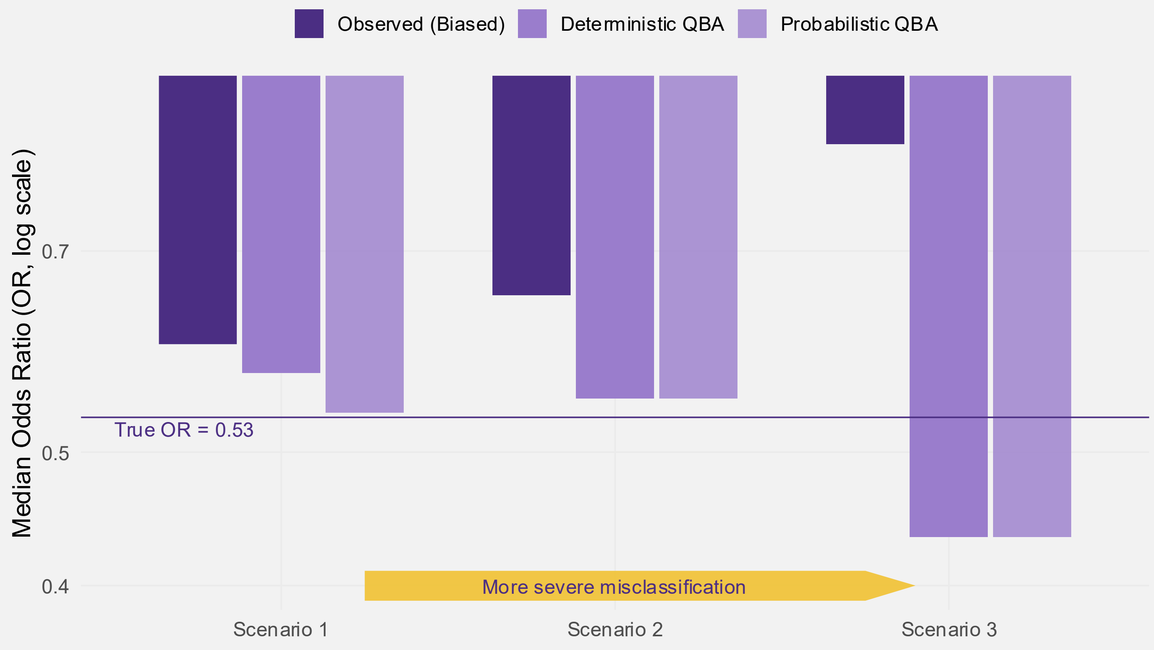
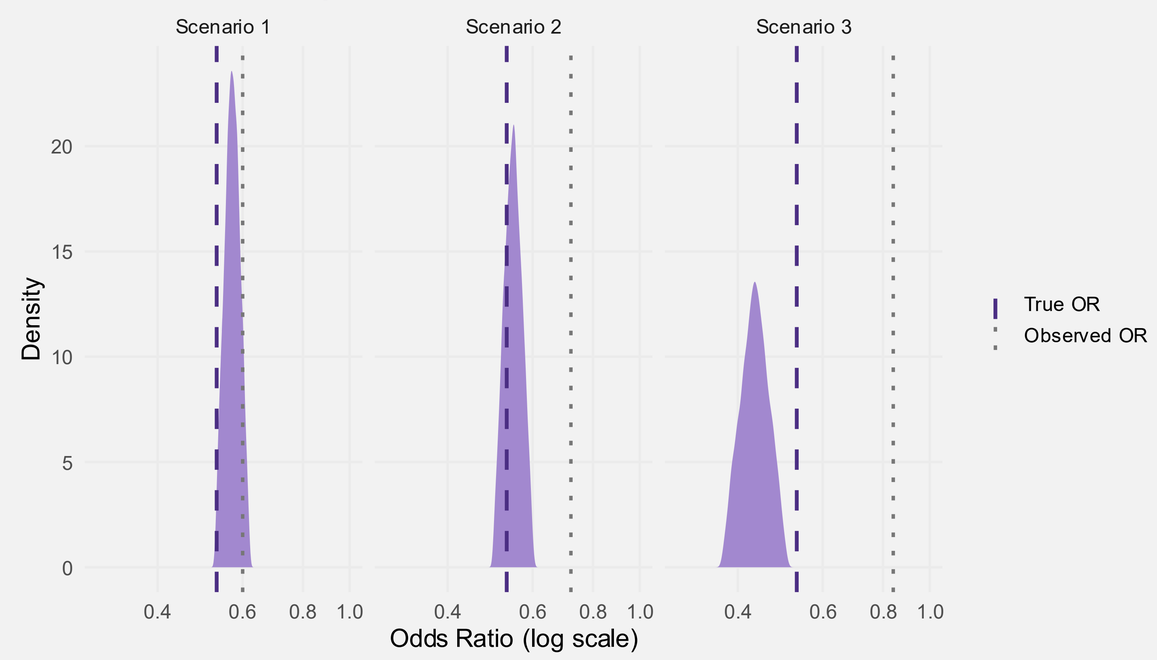


Figure 3. Probabilistic QBA (systematic error only): distributions of bias-adjusted odds ratios across misclassification scenarios. Densities are Monte Carlo draws from probsens using priors on sensitivity/specificity; dashed line = true effect (OR = 0.53), dotted line = observed OR).



- **Probabilistic QBA** produced central estimates [**Figure 2**] close to the true effect **(median ORs ≈0.43–0.57)** while also generating **scenario-specific distributions** of adjusted ORs [**Figure 3**]. These distributions show how uncertainty in sensitivity and specificity propagates into treatment effect estimates: corrected effects remain centered near the truth but with wider intervals.
- Following correction with QBA, Ratio of Odds Ratios ranged between 0.81–1.08 and 0.87–1.06 for deterministic and probabilistic QBA, respectively [**Table 2**].

Table 2. Simulated and bias-adjusted odds ratios (OR) and ratios of odds ratios (RoOR) across misclassification scenarios.

Approach	Scenario 1	Scenario 2	Scenario 3
True effect	OR = 0.53	OR = 0.53	OR = 0.53
Before correction (observed)	OR = 0.60, RoOR = 1.13	OR = 0.72, RoOR = 1.36	OR = 0.84, RoOR = 1.58
After deterministic QBA (fixed Se/Sp)	OR = 0.43, RoOR = 0.81	OR = 0.50, RoOR = 0.94	OR = 0.57, RoOR = 1.08
After probabilistic QBA (systematic only, medians)	OR = 0.46, RoOR = 0.87	OR = 0.51, RoOR = 0.96	OR = 0.56, RoOR = 1.06

Discussion & Conclusion

- **Outcome misclassification** in RW external-control comparisons **meaningfully biases estimated effects** even when misclassification is modest.
- **Deterministic QBA**, using fixed sensitivity/specificity, **substantially reduces this bias** and recovers ORs close to the simulated truth across scenarios.
- **Probabilistic QBA** treats sensitivity/specificity as distributions, producing similar central estimates while **explicitly conveying parameter uncertainty** via Monte Carlo simulation, offering a transparent way for decision-makers to assess robustness of treatment effects under parameter uncertainty.
- Probabilistic QBA can further **incorporate random error from sampling variability** beyond the systematic error correction.
- **Recommendations**: pre-specify QBA in SAPs; report observed and bias-adjusted effects and ratio-of-ORs or other relevant measures; examine systematic-only and systematic and random uncertainty to inform decision makers.

References

1. Patel D, Grimson F, Mihaylova E, Wagner P, Warren J, van Engen A, Kim J. Use of external comparators for health technology assessment submissions based on single-arm trials. Value in Health. 2021 Aug 1;24(8):1118-25.
2. Feinberg, B. A., Zettler, M. E., Klink, A. J., Lee, C. H., Gajra, A., & Kish, J. K. (2021). Comparison of solid tumor treatment response observed in clinical practice with response reported in clinical trials. JAMA Network Open, 4(2), e2036741-e2036741.
3. Desai, R. J., Levin, R., Lin, K. J., & Patorno, E. (2020). Bias Implications of Outcome Misclassification in Observational Studies Evaluating Association Between Treatments and All-Cause or Cardiovascular Mortality Using Administrative Claims. Journal of the American Heart Association, 9(17), e016906.
4 Leahy, T. P., Kent, S., Sammon, C., Groenwold, R. H., Grieve, R., Ramagopalan, S., & Gomes, M. (2022). Unmeasured confounding in nonrandomized studies: quantitative bias analysis in health technology assessment. Journal of Comparative Effectiveness Research, 11(12), 851-859.
5. Liu, J., Wang, S., & Shao, F. (2023). Quantitative bias analysis of prevalence under misclassification: evaluation indicators, calculation method and case analysis. International Journal of Epidemiology, 52(3), 942-951.
6.. Haine, Denis (2025). The episensr package: Basic sensitivity analysis of epidemiological results. R package version 2.0.0. doi: 10.5281/zenodo.4554553. https://dhaine.codeberg.page/episensr/.

Abbreviations

HTA, health technology assessment; OR, Odds Ratio; Se, sensitivity; Sp, specificity; QBA, quantitative bias analysis; SAP, statistical analysis plan; RW, real-world; RoOR, ratio of odds ratios;

Contact
Sylvaine Barbier
Sylvaine.Barbier@putassoc.com

Find out more at putassoc.com