Quantifying and Mitigating Measurement Bias in Real-world Endpoints when constructing external control arms

Workshop on the use of external controls for evidence generation in regulatory decision-making; EMA
Amsterdam, 3rd November 2025

Benjamin Ackerman, PhD
Principal Scientist, Biostatistics, Johnson & Johnson

EFSPI
EUROPEAN FEDERATION OF STATISTICIANS IN THE PHARMACEUTICAL INDUSTRY
Representing Statistical Associations in Europe

Disclosures

The views and opinions expressed in the following slides are those of the individual presenter and should not be attributed to the respective businesses, institutions or organizations with which the presenter is employed or affiliated.

This project is supported by the Food and Drug Administration (FDA) of the U.S. Department of Health and Human Services (HHS) as part of a financial assistance award [FAIN] totaling $468,058 (32 percent) funded by FDA/HHS and approximately $1,001,454 (68 percent) funded by non-government source(s). The contents are those of the author(s) and do not necessarily represent the official views of, nor an endorsement, by FDA/HHS, or the U.S. Government.

Endpoints are often *measured differently* in trial and real-world settings.

This limits the use of real-world data (RWD) to construct comparators for single-arm trials.

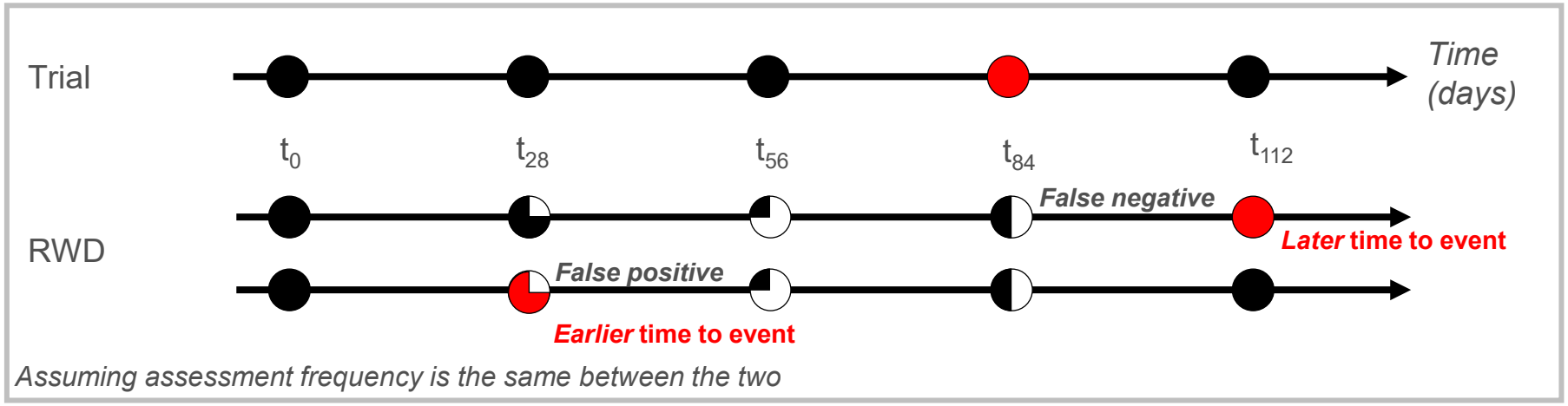
Especially challenging when considering time-to-event outcomes, like progression-free survival (PFS) in oncology.

Bias from measurement error in RWD may be due to *how* and *when* patients are assessed for disease progression

LEGEND ● Complete data ◔ ◑ ◐ Partial data ● Event occurrence

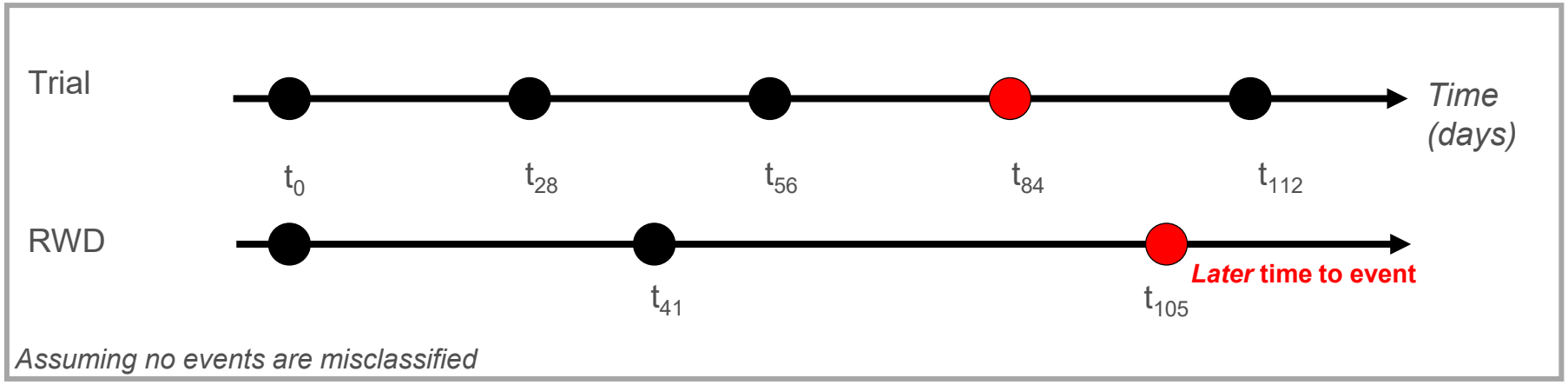
***How patients* are assessed: Misclassification bias**

Due to **missingness** of key biomarkers in RWD used to classify progression, and differences in **endpoint ascertainment approaches.**

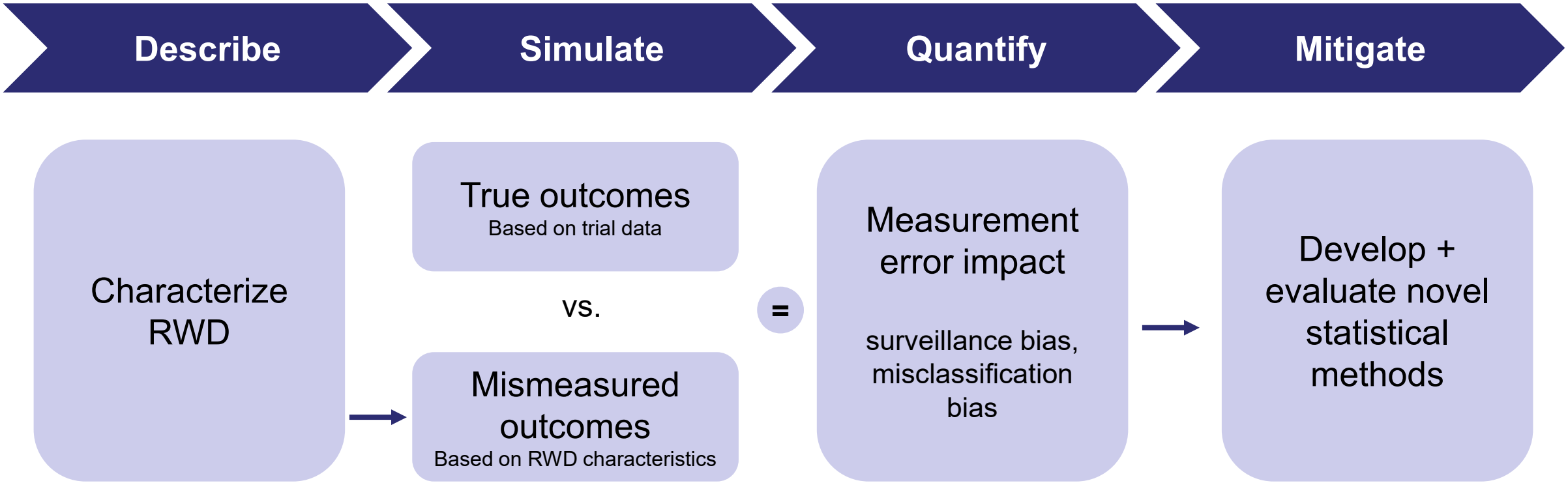


***When* patients are assessed: Surveillance bias**

Due to **variable assessment frequency** or visit cadence in routine clinical care.



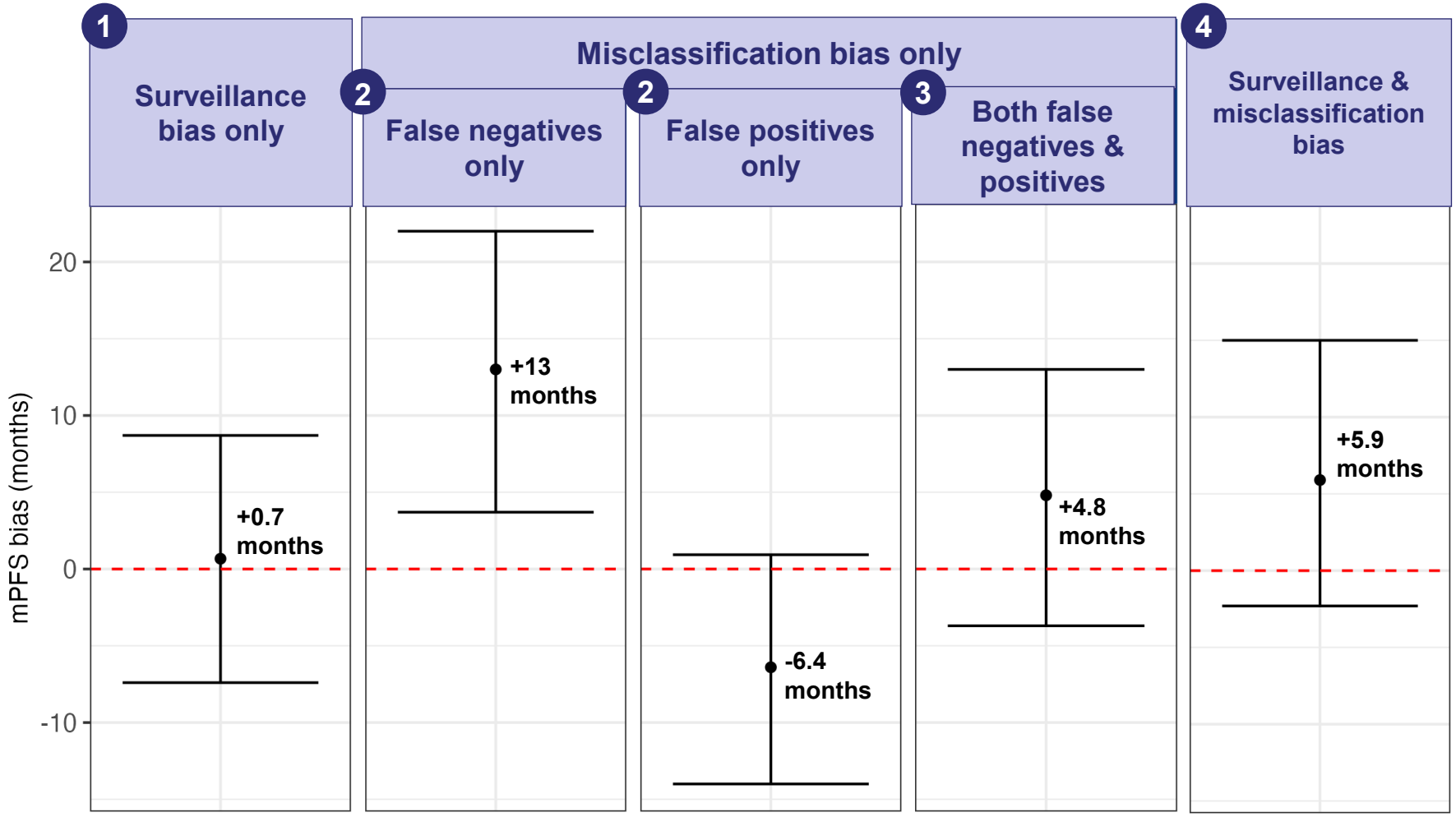
Simulation studies allow us to quantify bias and evaluate mitigation approaches



Ackerman, B., Gan, R. W., Meyer, C. S., Wang, J. R., Zhang, Y., Hayden, J., et al. (2024). *Measurement error and bias in real-world oncology endpoints when constructing external control arms*. Frontiers in Drug Safety and Regulation.

Simulation example: misclassified progression events in Newly Diagnosed Multiple Myeloma RWD can bias median PFS

1. Irregular cadence alone has minimal impact on median PFS (mPFS) bias.
2. False positives and false negatives, individually, car have a large impact on mPFS bias.
3. False positives and negatives together generate *strong opposing bias* that may downplay the impact of each error.
4. Irregular cadences and misclassified events together can generate bias greater than from each individually.



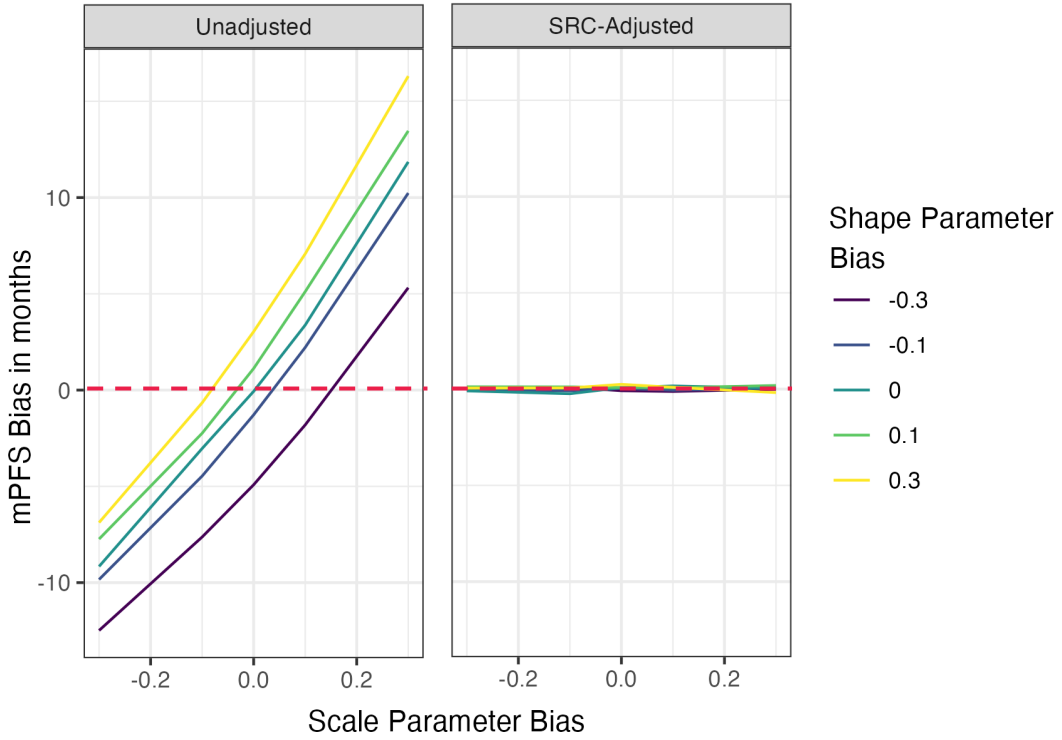
Simulation design: N = 365 patients with true mPFS = 34.2 months; false positive rate = 20%, false negative rate = 50%, 'irregular' cadence is every 28 days with higher variability than trial protocol. Error bars indicate middle 95% of point estimates under measurement error scenario, across 1,000 simulations.

Newly developed methods may mitigate endpoint measurement error bias in RWD comparators

Survival Regression Calibration (SRC)

1. **Get initial estimate** of mPFS from parametric Weibull survival model in RWD study
2. **Calculate bias** of model parameters ('scale' & 'shape') as difference between 'true' (trial-like) and 'mismeasured' (real-world) models in validation sample*
3. **Update estimate** of mPFS by calibrating model parameters according to bias factors

*Validation sample is a study where both the 'true' (trial-like) and 'mismeasured' (real-world) outcomes are collected, on a subset of the ECA RWD (internal) or a separate set of patients altogether (external) to be used for outcome measurement error modeling.



Simulation design: N = 365 patients with true mPFS = 34.2 months; 'mis-measured' outcome had bias in shape/scale Weibull parameters up to +/- 0.3; 40% (N ~ 150) patients sample for internal validation sample; 1,000 simulated datasets

Ackerman, B., Gan, R.R., Zhang, Y., Siddique, J., Roose, J., Lund, J.L., et al. (2025). "Regression calibration for time-to-event-outcomes: Mitigating bias due to measurement error in real-world endpoints." *Epidemiologic Methods.*

Applying SRC as Quantitative Bias Analysis (QBA)

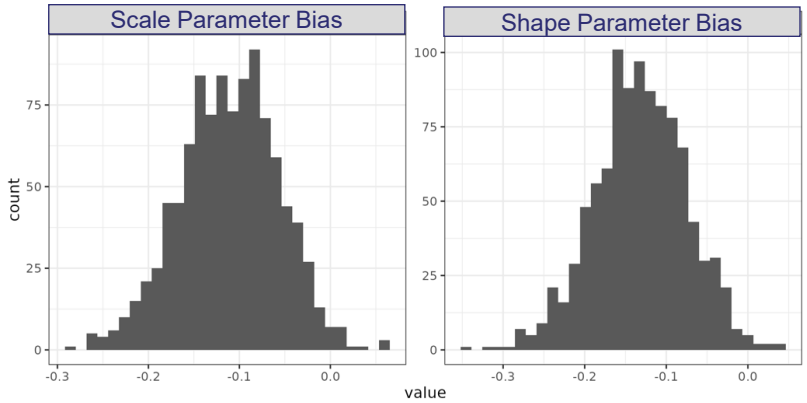
Contextualizing measurement error bias when we can't directly account for it (e.g. no validation samples)

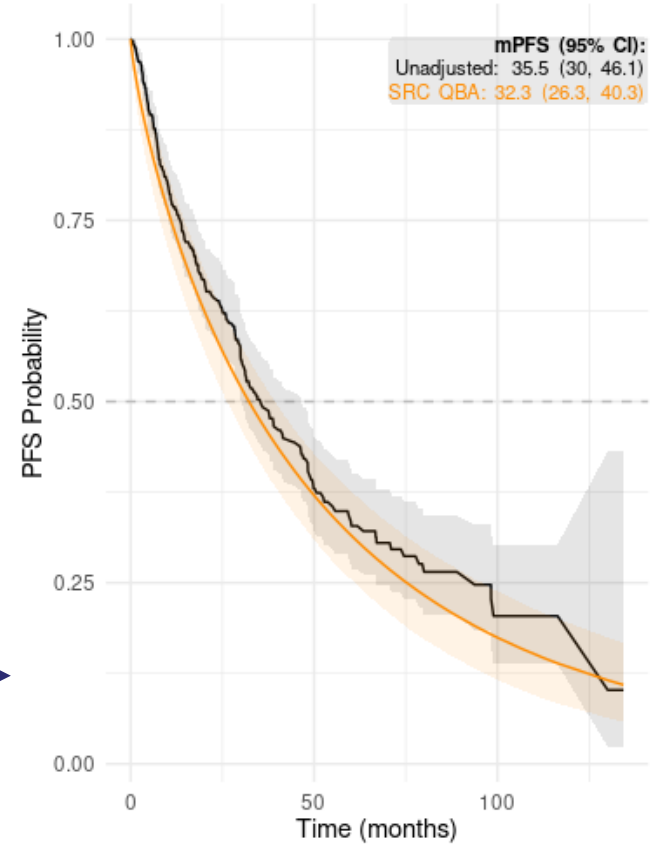
Problem: Validation sample may not be available to estimate bias correction factors needed to apply SRC

QBA approach: Set plausible ranges for bias parameters based on expert knowledge, leverage simulations, or conduct tipping point analyses

Result: More robust evidence generation that considers measurement differences between trial and real-world settings

Example: distribution of bias parameters via simulation can be used to calibrate rwPFS curve with SRC methodology





Ackerman, B., Gan, R.W., Meyer, C.S., Zhang, Y., Wang, J.R., Hayden, J., et al. (2025). "Quantitative bias analyses to address measurement error in time-to-event endpoints." *Under Review.*

It is often infeasible in ECA analyses to construct RWD comparators with endpoints aligned to trial standards, but:

Simulation studies can **demonstrate when no bias is present** due to measurement misalignments.

Methods can be applied to **quantify and reduce** statistically meaningful measurement error bias.

Quantitative Bias Analyses can **contextualize biased ECA findings** when measurement error bias is not directly estimable.

Robust statistical methods can **adequately align RWD endpoints** towards trial standards and **reduce biases for regulatory approval of ECAs**.

References

1. Ackerman, B., Gan, R.W., Meyer C.S., Wang J.R., Zhang Y., Hayden J., et al. (2024) “Measurement error and bias in real-world oncology endpoints when constructing external control arms.” Frontiers in Drug Safety and Regulation. http://doi.org/10.3389/fdsfr.2024.1423493
2. Ackerman, B., Gan, R.W., Zhang, Y., Siddique, J., Roose, J., Lund, J.L., et al. (2025). “Regression calibration for time-to-event-outcomes: Mitigating bias due to measurement error in real-world endpoints.” Epidemiologic Methods. https://doi.org/10.1515/em-2025-0009
3. Ackerman, B., Gan, R.W., Meyer, C.S., Zhang, Y., Wang, J.R., Hayden, J., et al. (2025). “Quantitative bias analyses to address measurement error in time-to-event endpoints.” Under Review.

Acknowledgements

J&J: Jennifer Hayden, Youyi Zhang, Jocelyn Wang, Joel Greshock, Kelly Reid Johnson, Jeffrey Kuhl, Laura Hester, Jason Breyer, Ashita Batavia

Academic & External Partners: Sebastian Schneeweiss, Jennifer Lund, Janick Weberpals, James Roose, Til Stürmer, Charles Poole, Juned Siddique, Issa Dahabreh, Smith Giri, Omar Nadeem, Sikander Ailawadhi, Noopur Raje

FDA colleagues

+ many more: Khaled Sarsour, Ryan Gan, Craig Meyer, Hemanth Kanakamedala, Sarah Rudolph, Yiyang Yue, Pranay Mohanty, Grace Mahoney, Nathan Chen, Ridwana Mahboob, Bryan Goldstein, John Phillips, James Range, Eric Brewer