

Confounded exposure metrics

Matthew R. Wiens  | Jonathan L. French  | James A. Rogers 

Metrum Research Group, Tariffville, Connecticut, USA

Correspondence

Matthew R. Wiens, Metrum Research Group, 2 Tunxis Road, Suite 112, Tariffville, CT 06081, USA.

Email: mattheww@metrumrg.com

Exposure-response (E-R) modeling frequently relies on the use of exposure metrics that summarize drug concentrations over time. We present simulations to demonstrate that certain commonly used exposure metrics, including average concentration up to an event time, are likely to lead to causal confounding under the very conditions that motivate their use.

Exposure-response (E-R) analysis modeling strategies are varied and often specific to the type of data collected in a trial.¹ One typical and important consideration is the choice of summary exposure metric. Although models for the dynamic effects of time-varying exposure generally permit a broader range of questions to be addressed, models using time-aggregated summaries of exposure are often preferred for their simplicity.

One common choice of time-aggregated exposure metric is average exposure until the event, C_{avgTE} . This metric may be computed by taking the area under the curve (AUC) up to the time of event, AUC_{TE} , and dividing by the time of the event TE, resulting in $C_{\text{avgTE}} = AUC_{\text{TE}} / \text{TE}$. This choice of exposure metric is generally motivated by a desire to leverage all relevant dosing and pharmacokinetic (PK) data until the event. In contexts involving dose adjustments and/or dose holidays, average concentration up to an event time (C_{avgTE}) may seem intuitively preferable to, for example, average concentration over the first dosing cycle (C_{avgC1}), or average concentration at steady-state ($C_{\text{avg,ss}}$), both of which are insensitive to the particularities of individual dosing histories.

Exposure metrics that depended on event times or responder status have been used in recent analyses and regulatory submissions. For example, in an analysis of cabozantinib, dose modifications throughout the trial justified the use of C_{avgTE} in a survival analysis to show E-R relationship for several safety end points as a function of cabozantinib exposure.² In a slight variation, a regulatory submission for selinexor used the average dose to the event and estimated clearance to derive a “time-averaged” AUC.³ One of the exposure metrics considered in a regulatory filing of inotuzumab ozogamicin was C_{avgTE} over the time interval of treatment, which partially depended on AEs.⁴

Unfortunately, the intuitive appeal of C_{avgTE} is misleading. As we demonstrate in the following simulation, the very conditions that motivate the use of such a metric (i.e., dosing patterns leading to higher or lower average exposures over time) are also conditions that will generate spurious associations between exposure and response.

To illustrate this point, we simulated a time-to-event response with no causal dependence on exposure, concentration, or other covariates and then analyzed (as categorical and continuous time-to-event end points) the simulated data using C_{avgTE} or average concentration in the first cycle (C_{avgC1}) exposure metrics. The PK data were simulated from a two-compartment model using *mrgsolve*⁵ with interindividual variability on clearance and no covariate effects. The PK parameters were chosen such that accumulation was negligible. The dose level was the same for all dose events and all patients, with a 3-week dosing cycle. Response data were simulated using a Weibull distribution for 75% of the patients, and the

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial](https://creativecommons.org/licenses/by-nc/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2023 Metrum Research Group, Inc. *CPT: Pharmacometrics & Systems Pharmacology* published by Wiley Periodicals LLC on behalf of American Society for Clinical Pharmacology and Therapeutics.

remaining 25% of the patients were assigned to not have the event. After six cycles (147 days), all patients were administratively censored. The simulations were analyzed graphically, with Kaplan–Meier curves, and with logistic regression, which is available on the linked GitHub repository (<https://github.com/metrumresearchgroup/confounder-exposure-metrics>).

The distribution used to simulate the time-to-event response had no covariate effects nor random interindividual variability. The distribution had the highest hazard at the beginning of the trial, and then monotonically decreased (Figure 1b). Initially, high hazard rates unrelated to study drug exposure can and do occur for many AEs, for example, when studies are designed to enroll participants following acute events, when the standard of care treatment in a trial of combination therapy entails short-term

risks, or when unobserved characteristics of patients affect the baseline hazard.

To illustrate how C_{avgTE} changes as a function of time, the average concentration from time zero through time t ($C_{\text{avg},t}$), or equivalently total area under the concentration curve divided by time ($C_{\text{avg},t} = \text{AUC}_{0-t}/t$), was derived for the typical patient at a grid of event times t (Figure 1a). For example, at $t = 21$ days, $C_{\text{avg}21}$ is $C_{\text{avg}C1}$. Two distinct trends are notable. Within each dosing cycle $C_{\text{avg},t}$ is highest near the start of the cycle and decreases over time. Overall, the highest values of $C_{\text{avg},t}$ are observed in the beginning of the first cycle.

A scatterplot of C_{avgTE} versus time to the event had a clear relationship between C_{avgTE} and event time when exposures were high, or the event time was small (Figure 2a). Continuing the hypothetical analysis,

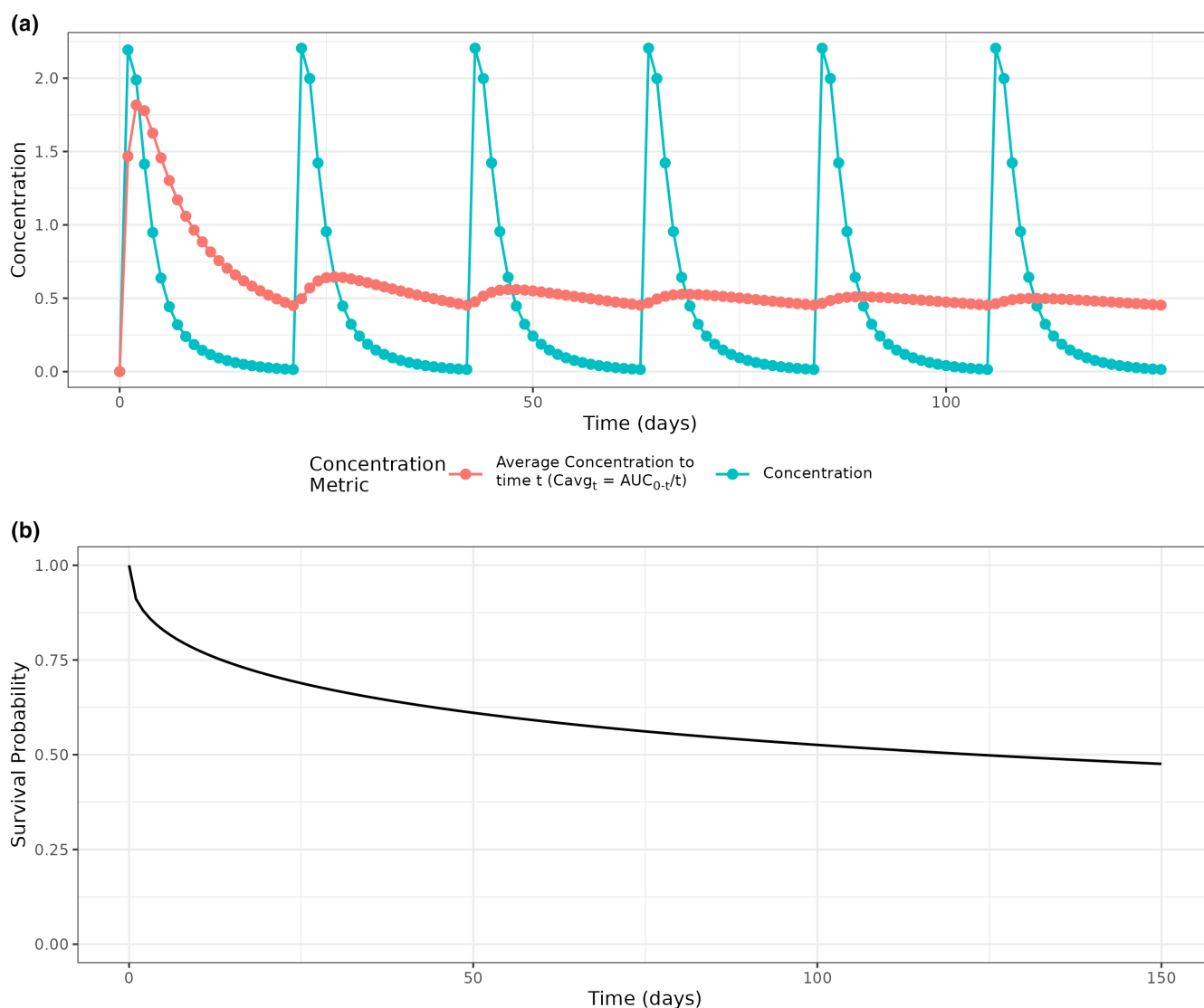


FIGURE 1 Simulation design. (a) Concentration-time curve for a typical patient, and the average concentration ($C_{\text{avg},t} = \text{AUC}_{0-t}/t$) from time zero, which is C_{avgTE} if an event were to occur at a specific time, t . (b) Survival curve for simulating event times for all patients, which is independent of concentration and has no covariate effects.

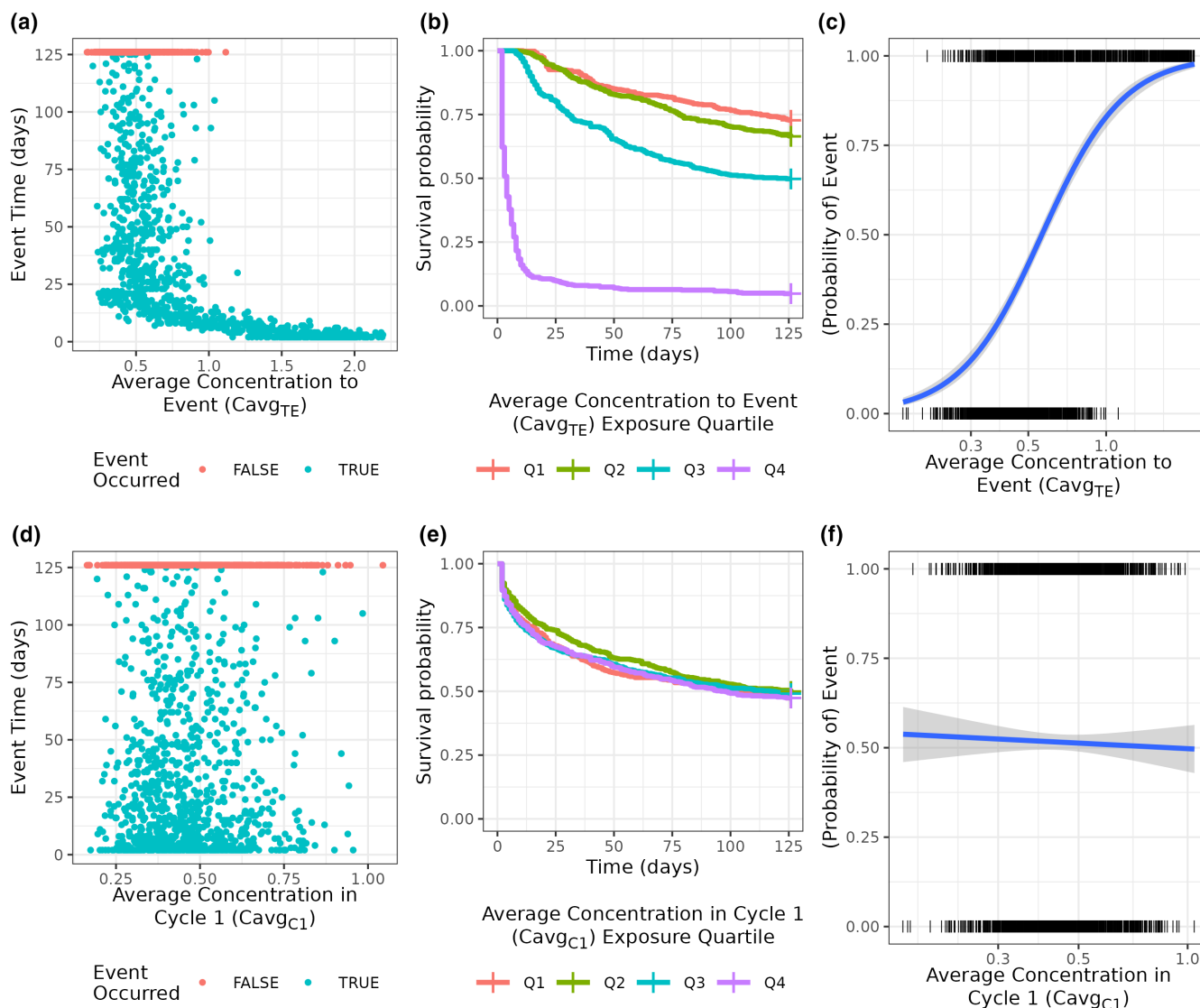


FIGURE 2 Simulation and analysis results by exposure metric used. (a–c) Analyses using average concentration to the event (C_{avgTE}) as the exposure metric, which show an exposure–response (E–R) relationship resulting from using the outcome to derive exposure. (d–f) Analyses using cycle 1 average concentration (C_{avgC1}) as the exposure metric, which correctly does not show an E–R relationship. In (d) and (f) the blue line is a logistic regression fit with a 95% confidence interval.

Kaplan–Meier plots stratified by C_{avgTE} quartile show a clear separation across the exposure groups, especially for the highest quartile of exposure (Figure 2b). A logistic regression using only C_{avgTE} as a predictor, that is, no covariates, also showed a clear relationship with the predicted probability of an event, ranging from ~0.1 at the lower range of exposure to ~0.9 at the highest exposures (Figure 2c). All three analysis strategies led to the same conclusion, that higher C_{avgTE} was associated with shorter time to the event and a higher event probability, and in 1000 replications, the true (null) causal effect was never contained inside the 95% confidence interval. Because C_{avgTE} was used as the exposure metric, this result would likely be interpreted as “higher exposures cause higher

event rates,” with corresponding consequences for future planning and regulatory interactions. Such a conclusion is incorrect because, by design, the true causal relationship was null (flat).

However, using C_{avgC1} led to an unbiased conclusion of the E–R relationship. The scatterplot between C_{avgC1} and event time correctly showed no association (Figure 2d). The Kaplan–Meier curves were essentially identical (Figure 2e), and the logistic regression had a negligible relationship between C_{avgC1} and the probability of an event (Figure 2f). These analyses would correctly lead to the conclusion that there is no E–R relationship. We note that use of C_{avgC1} is consistent with the recommendations in Dai et al.⁶ and Ruiz-Garcia et al.¹

The above-illustrated problem with C_{avgTE} cannot be diagnosed using standard statistical model fit or model comparison criteria. For example, in the presence of two competing models (one using C_{avgTE} and one using C_{avgC1}), a natural approach would be to use both qualitative and quantitative model evaluation tools. For the two logistic regression models, the model using C_{avgTE} had an Akaike information criterion (AIC) of 2046 and the model using C_{avgC1} had an AIC of 2774, indicating that using C_{avgTE} leads to a better fit of the data despite leading to the wrong causal conclusion.

Moreover, it is not even logically possible to correctly create certain simulation-based diagnostics, such as Visual Predictive Checks when using C_{avgTE} . The required simulation logic in this case would be circular: exposure depends on when the event happens, and the latter is unknown because it can only be simulated with knowledge of the exposure. The logical impossibility of constructing such a simulation is in itself an indication that the causal question has not been properly formulated.

To build an intuition of why C_{avgTE} led to a biased conclusion, one may imagine two patients with identical longitudinal concentration data, but different event times. These patients will have different C_{avgTE} because their event times are different. One would never use the outcome as a covariate in an E-R model, yet indirectly this is the logic when using C_{avgTE} . In other words, the outcome caused the exposure in the analysis instead of having the exposure cause the outcome, and the predicted probability of an outcome will be different only because their observed outcomes were different. This is a specific example of an explanatory variable that depends on the outcome, thereby introducing a spurious association between cause and effect.⁷

Furthermore, such confounding is not limited to average concentration metrics. The same principle of conditioning on the outcome applies in other circumstances. For example, consider a drug that accumulates after each dose and maximum concentration (C_{max}) is the exposure metric. C_{max} will increase cycle by cycle because of the accumulation, so longer event times will be associated with higher exposures. Again, information about the event time was used to determine the time window for computing C_{max} , and therefore bias is introduced (see the GitHub repository for a simulation example).

More broadly, it has been recognized that E-R modeling can be subject to bias due to causal confounding; for example, in the presence of unmeasured or unmodeled confounders⁶ and immortal time bias.^{8,9} The choice of exposure metric is another such way bias can be introduced into the analysis, as demonstrated by the preceding simulation.

Although non-null causal relationships were not specifically considered here, the null scenario is sufficient to convincingly demonstrate that analyses based on C_{avgTE} are problematic. As we have shown, a non-null association with C_{avgTE} is not evidence of a non-null causal effect—a sufficiently damning analysis property in itself. Neither have we considered scenarios where multiple dose levels are available. Multiple randomized dose levels would be expected to mitigate causal bias in any exposure-response analysis; nonetheless it would be ill-advised to knowingly introduce analytic bias only to hope that experimental design will provide a cure for the self-inflicted wound.

C_{avgTE} should be understood a priori as an exposure metric that will lead to biased analyses. Whereas there are conditions under which it will be unbiased, these are the very same conditions that would make C_{avgC1} or $C_{\text{avg,ss}}$ an equally valid metric, that is, scenarios with no average temporal trends in exposure. Instead, as a general approach, we suggest using exposure metrics that clearly do not depend on the outcome or intercurrent events, for example, C_{avgC1} , thinking clearly about how specific drug development questions lead to appropriate exposure metrics, or alternatively using models that avoid the use of summary measures of exposure altogether and instead model the dynamic effects of time-varying exposure, as described in Ruiz-Garcia et al.^{1,10} A simple rule of thumb is: “if you can’t in principle simulate responses using your exposure metric, choose a different exposure metric.”

FUNDING INFORMATION

No funding was received for this work.

CONFLICT OF INTEREST STATEMENT

The authors declared no competing interests for this work. As an Associate Editor for CPT: Pharmacometrics & Systems Pharmacology, Jonathan L. French was not involved in the review or decision process for this paper.

ORCID

Matthew R. Wiens  <https://orcid.org/0000-0002-0839-4425>

Jonathan L. French  <https://orcid.org/0000-0001-9649-169X>

James A. Rogers  <https://orcid.org/0009-0004-2363-1339>

REFERENCES

1. Ruiz-Garcia A, Baverel P, Bottino D, et al. A comprehensive regulatory and industry review of modeling and simulation practices in oncology clinical drug development. *J Pharmacokinet Pharmacodyn.* 2023;50:147-172. doi:10.1007/s10928-023-09850-2

2. Tran BD, Li J, Ly N, Faggioni R, Roskos L. Cabozantinib exposure-response analysis for the phase 3 CheckMate 9ER trial of nivolumab plus cabozantinib versus sunitinib in first-line advanced renal cell carcinoma. *Cancer Chemother Pharmacol*. 2023;91(2):179-189.
3. Center for Drug Evaluation and Research, US Food and Drug Administration. Clinical Pharmacology and Biopharmaceutics Review: Application Number 212306Orig1s000; selinexor [Internet]. 2019 [cited 2023 May 1]. https://www.accessdata.fda.gov/drugsatfda_docs/nda/2019/212306Orig1s000MultiDisciplineR.pdf
4. Center for Drug Evaluation and Research, US Food and Drug Administration. Clinical Pharmacology and Biopharmaceutics Review: Application Number 761040Orig1s000; Inotuzumab Ozogamicin [Internet]. 2016 [cited 2023 May 1]. https://www.accessdata.fda.gov/drugsatfda_docs/nda/2017/761040Orig1s000MultidisciplineR.pdf
5. Baron KT, Gillespie B, Margossian C, et al. mrgsolve: Simulate from ODE-Based Models [Internet]. Metrum Research Group 2021 <https://cran.r-project.org/package=mrgsolve>
6. Dai HI, Vugmeyster Y, Mangal N. Characterizing exposure-response relationship for therapeutic monoclonal antibodies in Immuno-oncology and beyond: challenges, perspectives, and prospects. *Clin Pharmacol Ther*. 2020;108(6):1156-1170.
7. Hernán MA, Robins JM. *Causal Inference: What if*. Chapman & Hall/CRC; 2020.
8. Harun R, Yang E, Kassir N, Zhang W, Lu J. Machine learning for exposure-response analysis: methodological considerations and confirmation of their importance via computational experiments. *Pharmaceutics*. 2023;15(5):1381.
9. Suissa S. Immortal time bias in observational studies of drug effects. *Pharmacoepidemiol Drug Saf*. 2007;16(3):241-249.
10. Center for Drug Evaluation and Research. E9(R1) Statistical Principles for Clinical Trials: Addendum: Estimands [Internet]. FDA.gov. 2021 [cited 2022 Jan 25]. <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/e9r1-statistical-principles-clinical-trials-addendum-estimands-and-sensitivity-analysis-clinical>

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Wiens MR, French JL, Rogers JA. Confounded exposure metrics. *CPT Pharmacometrics Syst Pharmacol*. 2024;13:187-191. doi:[10.1002/psp4.13074](https://doi.org/10.1002/psp4.13074)