Evaluating the Feasibility of Data Simulation for Radiation Research: Estimating RBE Values for HZE Radiation

Sneha Mehta[1], Lori J. Chappell[2], and S. Robin Elgart[3]

[1]Columbia University Mailman School of Public Health, New York, NY, [2]KBR, Houston, TX , [3]University of Houston, Houston, TX

Abstract

As the drive for deep space exploration intensifies, a comprehensive understanding of the health effects of radiation exposure becomes paramount to the future of human space flight. However, epidemiological data for radiation exposure, particularly to high-energy (HZE) ions, is limited, partially due to the financial and logistical costs of radiation studies. As an alternative, this study aims to assess the viability of data simulation strategies to accurately model potential study parameters prior to utilizing laboratory conditions. This study estimates a relative biological effectiveness (RBE) factor based on the solid tumor data for outbred mice provided by Edmundson et al. 2020.[1] Excess relative risk (ERR) models for HZE-Fe ions and gamma radiation were estimated using Poisson regression with Weibull models to represent the background solid tumor hazard without radiation. RBE values were calculated from the ratio of the heavy ion linear slope to the gamma linear slope. The parameters from these models were then applied to simulate iterations of 300 datasets across HZE-Fe doses of 0.05, 0.2, 0.4, and 0.75 Gy; gamma radiation doses of 0.75, 2, and 3 Gy, as well as an unirradiated control group. Sample size per dose varied from 100-500 mice across simulations (800-4000 mice total per dataset). 1500 datasets were generated total (300 each for sample sizes 800, 1600, 3200, and 4000). For each dataset, ERR per radiation type and RBE for HZE-Fe were calculated. The RBE from the Edmondson data was calculated to be 5.55. The RBEs from the simulated data converged around this value across the increasing sample sizes. For n = 4000, the mean RBE across the 300 datasets was 5.71 (95% CI: 5.47-5.95). These results suggest that RBEs calculated from simulated data are credible. Based on this exploration, data simulation is a viable method of testing radiation studies. It provides a method of testing study assumptions and refining research questions prior to embarking on costly laboratory experiments.

Data Description

In the original data produced by Edmundson et al, 2020[1], 1850 outbred mice were exposed to either (i) 0.4 Gy of HZE ^{28}Si ions, (ii) 0.4 Gy of HZE ^{56}Fe ions, (iii) 3 Gy of Cs gamma rays, or (iv) sham irradiation. 1804 mice were used in the final analysis. Mice were monitored until they reached 800 days of age or became moribund. The outcome of interest was solid cancer tumor rates at the time of morbidity or death. Mice that became moribund from other causes, died of other causes, or reached 800 days of age were censored. Dose-response functions were limited to linear estimations due to the fact there was only one dose per radiation type.

Radiation Effect Analyses

We analyzed the radiation effects for the dataset similar to Chappell et al 2023[2] using the following ERR model:

$$\boldsymbol{MM \cdot h_0(a,s)(1 + ERR(s,a,D,r)}$$

where *MM* is the number of mouse-months of follow-up per stratum, a is age in months, s is sex, D is dose in Gy, and r is radiation type. The background hazard function was defined using a Weibull model adjusted for sex. Poisson models were then used to estimate ERR based on stratified mouse-time and number of cases by sex, age (one-month categories from <14, 14-25, and ≥25 months) and radiation type. The ERR for gamma radiation was calculated to be 0.333 and 1.839 for HZE-Fe. RBE values were then estimated for each radiation type. The RBE for HZE-Fe radiation was calculated to be 5.55.

Analysis was performed using R[3] packages 'tidyverse,'[4] 'survival,'[5] 'survminer,'[6] 'gnm,'[7] 'simsurv,'[8] and 'flexsurv'[9].

Data Simulation

The calculated ERR parameters from the preliminary data were then applied to simulate sets of 300 datasets across various doses of gamma and HZE-Fe radiation. Doses of 0.75, 2, and 3 Gy of gamma radiation and 0.05, 0.2, 0.4, 0.75 Gy of HZE-Fe radiation were simulated as well as a control 0 Gy dose. Across 5 sets of simulations, sample size was varied in increments of 100 from 100 to 500 mice per dose (800-4000 mice total per dataset). 300 datasets were generated per simulation, yielding 1500 datasets total (300 each of sample sizes 800, 1600, 2400, 3200, and 4000 mice).

Analysis of the simulated data consisted of calculating the ERR and RBE per radiation type, examining Kaplan-Meier curves to gauge cancer vs censoring rates, and evaluating the change in RBE across different sample sizes. For simplicity, only results for HZE-Fe are included below but the simulated gamma values yielded similar results. Due to model convergence issues in 5 datasets, 1495 datasets were used in the final analysis.

Results

Across different sample sizes, cancer and censoring rates behaved as expected per dose (see Fig. 1). HZE-Fe at 0.75 Gy had the lowest survival rates followed by 3 Gy of gamma radiation. At lower sample sizes, these two doses behaved similarly but became more distinct at higher sample sizes.

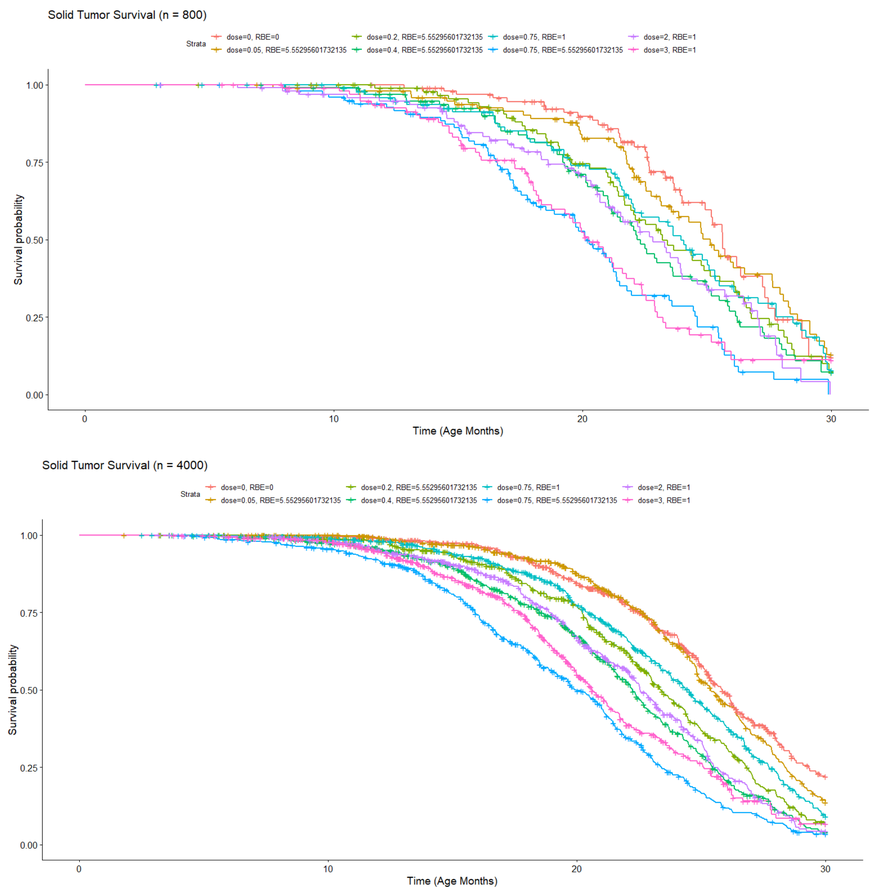

Fig 1 (a) & (b): Kaplan-Meier curves for one dataset each of n = 800 (a) and n = 4000 (b)

In the original data, at a sample size of 1804 mice, the HZE-Fe RBE was calculated to be 5.552 at a standard error of 0.880. Generally, the simulations overestimated the RBE. However, as sample sizes increased across simulations, the distributions of calculated RBEs converged around the original value (Fig 2). Additionally, the calculated standard error converged towards 0 with increasing sample sizes (Fig 3). These measures indicate the strength of the models in predicting viable results.

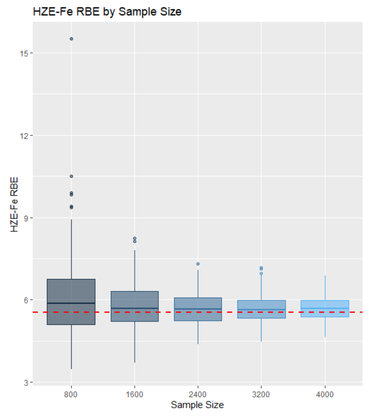

Fig 2: Distributions of all calculated RBE values per dataset across sample sizes. Red line indicates RBE from the original data.

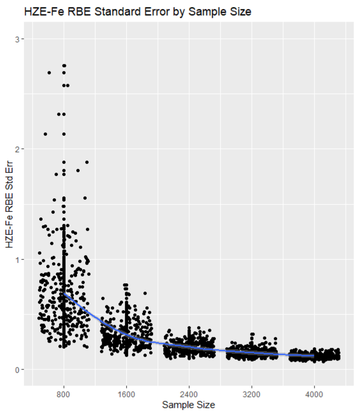

Fig 3: Plot of all calculated standard error values for RBE per dataset across sample sizes. Blue line indicates average value.

Table 1 summarizes the mean HZE-Fe RBEs and standard errors along with their 95% confidence intervals per sample size across all datasets. The original calculated RBE was within the range of all confidence intervals indicating that the assumed RBE can be estimated at a 5% level of significance. The mean standard errors of the simulated data were smaller than that of the original data most likely due to the larger number of doses available in the simulated data compared to the original data.

Sample Size	Mean RBE	95% RBE CI	Mean Std Err	95% Std Err CI
800	6.02	(4.58 - 7.45)	0.73	(0.64 - 0.82)
1600	5.79	(5.17 - 6.41)	0.32	(0.30 - 0.33)
2400	5.68	(5.28 - 6.08)	0.20	(0.19 - 0.21)
3200	5.68	(5.39 - 5.98)	0.15	(0.14 - 0.15)
4000	5.71	(5.47 - 5.95)	0.12	(0.11 - 0.12)

Table 1: Summary of mean RBE values and standard errors across sample sizes. The original HZE-Fe RBE was 5.55 with a standard error of 0.880 at a sample size of 1804.

Conclusions

Based on this exploration, statistical modeling techniques can be applied to simulate viable radiation data. Data simulation has the potential ability to test study assumptions, create more complex study parameters, and refine research questions prior to requiring the resources necessary for lab-based studies. Therefore, it can be a powerful tool in further developing the field of radiation research.

References

1) Edmondson, E. F. *et al.* (2020). Genomic mapping in outbred mice reveals overlap in genetic susceptibility for HZE ion– and γ-ray–induced tumors. *Sci. Adv.* **6**, eaax5940.
2) Chappell, L. J. *et al.* (2023). Of Men and Mice: Using Terrestrial Radiation Epidemiology Methods to Inform Analysis of Animal Models for Space Radiation Risk Assessment. *Radiat Res.* 200, 000–000.
3) R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
4) Wickham H. *et al.* (2019). "Welcome to the tidyverse." *Journal of Open Source Software*, *4*(43), 1686.
5) Therneau T. (2022). A Package for Survival Analysis in R. R package version 3.4-0.
6) Kassambara A. *et al.* (2021). survminer: Drawing Survival Curves using 'ggplot2'. R package version 0.4.9.
7) Turner H. *et al.* (2022). Generalized nonlinear models in R: An overview of the gnm package. (R package version 1.1-2).
8) Brilleman S.L. *et al.* (2020). "Simulating Survival Data Using the simsurv R Package." *Journal of Statistical Software*, *97*(3), 1-27. doi:10.18637/jss.v097.i03.
9) Jackson C. (2016). flexsurv: A Platform for Parametric Survival Modeling in R. *Journal of Statistical Software*, 70(8), 1-33. doi:10.18637/jss.v070.i08.