# Appendix 1 – Simulation of survival data

Individual patient level overall survival (OS) data were simulated and included two measured baseline characteristics, one continuous () and one binary (), both of which were assumed to be confounders and are marginally independent. The survival data was simulated through a hazard function where the baseline hazard followed a Weibull distribution. Suppose *D* is a binary treatment variable, *U* a binary unmeasured confounder (), *X* the design matrix with and *t* be time, then the hazard function was given by the following multiplicative model,

|  |  |  |
| --- | --- | --- |
|  |  | (1) |

where γ is the regression coefficient of *D,*  a vector of regression coefficients, η the regression coefficient of *U* and the baseline hazard.

The overall parameter of interest that was used in the CEM was γ, and given , the binary treatment variable *D* was simulated by

|  |  |  |
| --- | --- | --- |
|  |  | (2) |

where α was a vector of the intercept and regression coefficients and the regression coefficient of *U* in the treatment model and

|  |  |  |
| --- | --- | --- |
|  |  | (3) |

Survival times were simulated from Equation (1) using the approach described in [1] with the baseline hazard function given by

|  |  |  |
| --- | --- | --- |
|  |  | (4) |

where and are the scale and shape parameters respectively. The binary treatment variable (*D*) represented a hypothetical real-word data control arm and a single-armed trial, and parameters were selected such that the RWD arm represented approximately two-thirds of the simulated patient population.

In addition, an independent censoring mechanism was assumed where censoring times were generated such that the censoring indicator *C* was defined as,

|  |  |  |
| --- | --- | --- |
|  |  | (5) |

where the parameter was varied to achieve a censoring proportion between 15-25% and is the indicator function. A maximum follow-up period was defined and any simulated patient that had not experienced an event by the maximum follow-up period were administratively censored.

Table 1. Parameters used in the simulation of patient-level data

|  |  |  |
| --- | --- | --- |
| **Parameter** | **Parameter description** | **Parameter value** |
| N | Number of simulated patients | 300 |
|  | Coefficient of the intercept in the treatment model | -1 |
|  | Coefficient of in the treatment model | 0.35 |
|  | Coefficient of in the treatment model | -0.25 |
|  | Coefficient of the intercept in the outcome model | -0.23 |
|  | Coefficient of in the outcome model | -0.2 |
|  | Coefficient of in the outcome model | 0.2 |
|  | Coefficient of the treatment (exposure) in the outcome model | -0.598 |
|  | Coefficient for the censoring mechanism | 4 |
|  | Coefficient of in the treatment model | 0.693 |
|  | Coefficient of in the outcome model | 1.253 |
| λ | Scale parameter of the Weibull baseline hazard | 1 |
| ν | Shape parameter of the Weibull baseline hazard | 1.5 |

1. Bender, R., T. Augustin, and M. Blettner, *Generating survival times to simulate Cox proportional hazards models.* Statistics in medicine, 2005. **24**(11): p. 1713-1723.