Web-based supplementary materials for "Design and analysis of cluster randomized trials with time-to-event outcomes under the additive hazards mixed model" by Blaha, Esserman, and Li.

1 | ADDITIONAL DETAILS ON DERIVING THE SANDWICH VARIANCE EXPRESSION

1.1 | Assuming a general one-parameter frailty distribution

We first elaborate on the steps to derive the explicit forms of ω_1 and ω_0 when the frailty distribution follows a general one-parameter frailty distribution such that $\xi_i \sim f(\xi;\theta)$, where θ represents the unknown frailty parameter. Obtaining this general expression allows one to extrapolate the sample size formula to accommodate an alternative frailty distribution; furthermore, the result in our main manuscript under the normal frailty assumption can also be treated as a special case.

To proceed, we first define

$$G(t;\theta) = -\log\left\{\int_0^\infty e^{-xt} f(x;\theta)dx\right\}$$

as the negative log Laplace transform, in which case the marginal hazard function induced by the AHMM is given by

$$\lambda_0(t) + G'(t;\theta) + Z_i\delta,$$

and we define $G'(t;\theta)$ and $G''(t;\theta)$ as the first and second derivative of $G(t;\theta)$. We notice that the above Laplace transform has several interesting properties which we provide below. After interchanging differentiation and integration, we have

$$\begin{aligned}
G'(t;\theta) =& \frac{\int_0^\infty xe^{-xt} f(x;\theta)dx}{\int_0^\infty e^{-xt} f(x;\theta)dx} \\
G''(t;\theta) =& \frac{-\left(\int_0^\infty x^2 e^{-xt} f(x;\theta)dx\right)\left(\int_0^\infty e^{-xt} f(x;\theta)dx\right) - \left(\int_0^\infty xe^{-xt} f(x;\theta)dx\right)^2}{\left(\int_0^\infty e^{-xt} f(x;\theta)dx\right)^2}.
\end{aligned}$$

Then simple differentiation allows us to verify that

$$\begin{aligned}
\frac{\partial G(t+s)}{\partial t} &= \frac{\partial G(t+s)}{\partial s} = \frac{\int_0^\infty xe^{-x(t+s)} f(x;\theta)dx}{\int_0^\infty e^{-x(t+s)} f(x;\theta)dx} = G'(t+s;\theta) \\
\frac{\partial^2 G(t+s)}{\partial t\partial s} &= \frac{-\left(\int_0^\infty x^2 e^{-x(t+s)} f(x;\theta)dx\right)\left(\int_0^\infty e^{-x(t+s)} f(x;\theta)dx\right) - \left(\int_0^\infty xe^{-x(t+s)} f(x;\theta)dx\right)^2}{\left(\int_0^\infty e^{-x(t+s)} f(x;\theta)dx\right)^2} = G''(t+s;\theta)
\end{aligned}$$

Recall that the sandwich variance for the intervention effect estimator in an AHMM takes the form $\sigma_\delta^2 = \omega_1/\omega_0^2$, where

$$\omega_0 = E\left(\sum_{j=1}^{m_i}\int_0^\tau Y_{ij}(t)\left(Z_i - \mu(t)\right)^2 dt\right) \quad \text{and} \quad \omega_1 = E\left(\sum_{j=1}^{m_i}\int_0^\tau Y_{ij}(t)\left(Z_i - \mu(t)\right) dM_{ij}(t)\right)^2.$$

For ω_0, under cluster randomization, we can still follow the main paper to show that

$$\mu(t) = \frac{E\left\{\sum_{j=1}^{m_i} Z_i Y_{ij}(t)\right\}}{E\left\{\sum_{j=1}^{m_i} Y_{ij}(t)\right\}} = \frac{\pi S^{(1)}(t)}{(1-\pi)S^{(0)}(t) + \pi S^{(1)}(t)}. \tag{1}$$

and

$$\omega_0 = \overline{m}\int_0^\tau \mathcal{G}(t)\left\{\frac{1}{(1-\pi)S^{(0)}(t)} + \frac{1}{\pi S^{(1)}(t)}\right\}^{-1} dt.$$

Next, for ω_1, we use the Law of Iterated Expectation and write $\omega_1 = E_{m_i}(\omega_1^*)$ where

$$\omega_1^* = m_i E_{Z_i}\left(\int_0^\tau \mathcal{G}(t)(Z_i-\mu(t))^2 f^{(Z_i)}(t)dt\right) \tag{2}$$

$$+ m_i(m_i-1)E_{Z_i}\left(\int_0^\tau\int_0^\tau \mathcal{G}(t,s)(Z_i-\mu(t))(Z_i-\mu(s))f^{(Z_i)}(t,s)dtds\right) \tag{3}$$

$$- m_i(m_i-1)E_{Z_i}\left(\int_0^\tau\int_0^\tau \mathcal{G}(t,s)(Z_i-\mu(t))(Z_i-\mu(s))\frac{\partial S^{(Z_i)}(t,s)}{\partial t}\{-\lambda_0(s)-G'(s;\theta)-Z_i\delta\}dtds\right) \tag{4}$$

$$- m_i(m_i-1)E_{Z_i}\left(\int_0^\tau\int_0^\tau \mathcal{G}(t,s)(Z_i-\mu(t))(Z_i-\mu(s))\frac{\partial S^{(Z_i)}(t,s)}{\partial s}\{-\lambda_0(t)-G'(t;\theta)-Z_i\delta\}dtds\right) \tag{5}$$

$$+ m_i(m_i-1)E_{Z_i}\left(\int_0^\tau\int_0^\tau \mathcal{G}(t,s)(Z_i-\mu(t))(Z_i-\mu(s))S^{(Z_i)}(t,s)\{-\lambda_0(t)-G'(t;\theta)-Z_i\delta\}\{-\lambda_0(s)-G'(s;\theta)-Z_i\delta\}dtds\right). \tag{6}$$

By taking the expectation of each term, we can define that

$$(2) = m_i g_1 = m_i\left\{(1-\pi)\int_0^\tau \mathcal{G}(t)\mu(t)^2 f^{(0)}(t)dt + \pi\int_0^\tau \mathcal{G}(t)\,(1-\mu(t))^2\, f^{(1)}(t)dt\right\},$$

$$(3) = m_i(m_i-1)g_2$$

$$= m_i(m_i-1)\left\{(1-\pi)\int_0^\tau\int_0^\tau \mathcal{G}(t,s)\mu(t)\mu(s)f^{(0)}(t,s)dtds + \pi\int_0^\tau\int_0^\tau \mathcal{G}(t,s)(1-\mu(t))(1-\mu(s))f^{(1)}(t,s)dtds\right\},$$

$$(4)+(5) = m_i(m_i-1)(g_3+g_4)$$

$$= m_i(m_i-1)(1-\pi)\int_0^\tau\int_0^\tau \mathcal{G}(t,s)\mu(t)\mu(s)\frac{\partial S^{(0)}(t,s)}{\partial t}\left(\lambda_0(s)+G'(s;\theta)\right)dtds$$

$$+ m_i(m_i-1)\pi\int_0^\tau\int_0^\tau \mathcal{G}(t,s)(1-\mu(t))(1-\mu(s))\frac{\partial S^{(1)}(t,s)}{\partial t}\left(\lambda_0(s)+G'(s;\theta)+\delta\right)dtds$$

$$+ m_i(m_i-1)(1-\pi)\int_0^\tau\int_0^\tau \mathcal{G}(t,s)\mu(t)\mu(s)\frac{\partial S^{(0)}(t,s)}{\partial s}\left(\lambda_0(t)+G'(t;\theta)\right)dtds$$

$$+ m_i(m_i-1)\pi\int_0^\tau\int_0^\tau \mathcal{G}(t,s)(1-\mu(t))(1-\mu(s))\frac{\partial S^{(1)}(t,s)}{\partial s}\left(\lambda_0(t)+G'(t;\theta)+\delta\right)dtds,$$

$$(6) = m_i(m_i-1)(g_5+g_6)$$

$$= m_i(m_i-1)(1-\pi)\int_0^\tau\int_0^\tau \mathcal{G}(t,s)\mu(t)\mu(s)S^{(0)}(t,s)\left(\lambda_0(s)+G'(s;\theta)\right)\left(\lambda_0(t)+G'(t;\theta)\right)dtds$$

$$+ m_i(m_i-1)\pi\int_0^\tau\int_0^\tau \mathcal{G}(t,s)(1-\mu(t))(1-\mu(s))S^{(1)}(t,s)\left(\lambda_0(s)+G'(s;\theta)+\delta\right)\left(\lambda_0(t)+G'(t;\theta)+\delta\right)dtds$$

By putting these terms together and taking the expectation with respect to m_i, we get

$$\omega_1 = E(\omega_1^*) = \overline{m} g_1 + \left(\overline{m^2} - \overline{m}\right)\left(g_2 + g_3 + g_4 + g_5 + g_6\right).$$

By definition of the CV of cluster sizes (square-root of variance over mean), we then have

$$\sigma_\delta^2 = \frac{\omega_1}{\omega_0^2} = \frac{g_1 + \left\{(1+\mathrm{CV}^2)\overline{m} - 1\right\}\left(g_2 + g_3 + g_4 + g_5 + g_6\right)}{\overline{m} h^2}.$$

In the above integral terms, notice that explicit forms of the integrand can be obtained under the AHMM when the frailty follows a general one-parameter distribution with parameter θ First of all, the arm-specific marginal survival function can be derived as

$$S^{(z)}(t) = P(T_{ij} \geq t | Z_i = z) = \exp\left\{-\int_0^t \lambda_0(u)du - G(t;\theta) - z\delta t\right\}.$$

Similarly, the arm-specific marginal bivariate survival function can be derived as

$$S^{(z)}(t,s) = P(T_{ij} \geq t, T_{ij} \geq s | Z_i = z) = \exp\left\{-\int_0^t \lambda_0(u)du - \int_0^s \lambda_0(u)du - G(t+s;\theta) - z\delta(t+s)\right\}.$$

Therefore, we can compute the mean function (1) as

$$\mu(t) = \frac{\pi \exp\left\{-\int_0^t \lambda_0(u)du - G(t;\theta) - \delta t\right\}}{(1-\pi)\exp\left\{-\int_0^t \lambda_0(u)du - G(t;\theta)\right\} + \pi \exp\left\{-\int_0^t \lambda_0(u)du - G(t;\theta) - \delta t\right\}} = \frac{\pi}{\pi + (1-\pi)\exp(\delta t)}.$$

The arm-specific univariate density function under the AHMM is given by

$$f^{(z)}(t) = -\frac{\partial S^{(z)}(t)}{\partial t} = \exp\left\{-\int_0^t \lambda_0(u)du - G(t;\theta) - z\delta t\right\}\left(\lambda_0(t) + G'(t;\theta) + z\delta\right).$$

Finally, the arm-specific bivariate density function is given by

$$\begin{aligned} f^{(z)}(t,s) =& \frac{\partial^2}{\partial t \partial s}\left(S^{(z)}(t,s) - S^{(z)}(t) - S^{(s)}(t) + 1\right) = \frac{\partial^2 S^{(z)}(t,s)}{\partial t \partial s} \\ =& \exp\left\{-\int_0^t \lambda_0(u)du - \int_0^s \lambda_0(u)du - G(t+s;\theta) - z\delta(t+s)\right\}\left(\lambda_0(t) + G'(t+s;\theta) + z\delta\right)\left(\lambda_0(s) + G'(t+s;\theta) + z\delta\right) \\ & - \exp\left\{-\int_0^t \lambda_0(u)du - \int_0^s \lambda_0(u)du - G(t+s;\theta) - z\delta(t+s)\right\} G''(t+s;\theta). \end{aligned}$$

1.2 | Sandwich variance expressions under the normal or exponential frailty distribution

Under the assumption of the normal frailty as in our main manuscript, we can set $\theta = \sigma^2$ as the parameter controlling for the frailty variance, in which case $f(\xi;\theta = \sigma^2)$ is the normal density. It is straightforward to verify that $G(t;\theta) = -t^2\theta/2$, $G'(t;\theta) = -t\theta$ and $G''(t;\theta) = -\theta$, and the expression in our main manuscript is then obtained.

If the frailty follows a mean-zero exponential frailty such that $\xi_i + \theta \sim \mathrm{Exp}(\mathrm{mean} = \theta)$, then we also have $G(t;\theta) = \log(1+\theta t)$, $G'(t;\theta) = \theta(1+\theta t)^{-1}$ and $G''(t;\theta) = -\theta^2(1+\theta t)^{-2}$. Plugging these expressions in the above general formula results in the sample size formula when the frailty follows an exponential distribution.

In Web Figure 5 , we compare the predicted power under the normal frailty and the mean-zero exponential frailty distribution for common values of the design parameters (for example, by using the same value of the variance parameters for both frailty distributions).

2 | WEB TABLES ON TYPE I ERROR RATE

WEB TABLE 1 Empirical type I error rates for the z-test and randomization test from a simulation study based on different parameter configurations of the baseline hazard ($\lambda_0(t)$), frailty variance (σ^2), average cluster size ($\overline{m}$), and estimated number of clusters ($\widehat{n}$) when the coefficient of variation of cluster size is 0. BC0: Uncorrected sandwich variance estimator; BC1: bias-corrected sandwich variance estimator extending Kauermann and Carroll; BC2: bias-corrected sandwich variance estimator extending Mancl and DeRouen; BC3: bias-corrected sandwich variance estimator extending Fay and Graubard; BC4: bias-corrected sandwich variance estimator extending Morel et al; Rand: randomization-based test. Bold values indicate values close to nominal empirical type I error rate (from 3.6% to 6.4%).

Configuration				Type I error				
$\lambda_0(t)$	σ^2	$\overline{m}$	$\widehat{n}$	BC0 (z-test)	BC1 & BC3 (z-test)	BC2 (z-test)	BC4 (z-test)	Rand test
4	0.5	10	48	0.068	0.066	0.066	0.065	**0.059**
4	0.5	30	22	0.093	0.081	0.074	0.077	**0.061**
4	0.5	50	16	0.098	0.090	0.082	0.083	**0.055**
2	0.5	10	24	0.077	0.071	0.068	0.068	**0.054**
2	0.5	30	14	0.091	0.073	0.068	0.065	**0.047**
2	0.5	50	12	0.113	0.100	0.083	0.087	**0.054**
2	0.5	10	46	**0.064**	**0.057**	**0.050**	**0.052**	**0.046**
2	0.5	30	26	0.081	0.074	0.069	0.070	**0.057**
2	0.5	50	24	0.089	0.085	0.078	0.079	0.068
2	0.25	10	40	0.083	0.080	0.076	0.075	**0.061**
2	0.25	30	20	0.072	**0.063**	**0.055**	**0.058**	**0.041**
2	0.25	50	16	0.106	0.092	0.079	0.085	**0.049**
2	0.05	10	34	0.070	0.067	**0.063**	**0.064**	**0.053**
2	0.05	30	14	0.097	0.079	0.069	0.073	**0.055**
2	0.05	50	8	0.143	0.111	0.088	0.090	**0.047**
1	0.05	10	10	0.129	0.110	0.093	0.088	**0.054**
1	0.05	10	18	0.078	0.068	**0.061**	**0.062**	**0.050**
1	0.05	30	8	0.126	0.108	0.084	0.084	**0.041**
1	0.05	50	6	0.183	0.144	0.111	0.103	0.002
1	0.05	10	52	**0.046**	**0.045**	**0.042**	**0.042**	**0.036**
1	0.05	30	22	0.091	0.086	0.075	0.075	**0.062**
1	0.05	50	16	0.101	0.092	0.086	0.084	**0.062**

WEB TABLE 2 Empirical type I error rates for the z-test and randomization test from a simulation study based on different parameter configurations of the baseline hazard ($\lambda_0(t)$), frailty variance (σ^2), average cluster size ($\overline{m}$), and estimated number of clusters ($\widehat{n}$) when the coefficient of variation of cluster size is 0.25. BC0: Uncorrected sandwich variance estimator; BC1: bias-corrected sandwich variance estimator extending Kauermann and Carroll; BC2: bias-corrected sandwich variance estimator extending Mancl and DeRouen; BC3: bias-corrected sandwich variance estimator extending Fay and Graubard; BC4: bias-corrected sandwich variance estimator extending Morel et al; Rand: randomization-based test. Bold values indicate values close to nominal empirical type I error rate (from 3.6% to 6.4%).

Configuration				Type I error				
$\lambda_0(t)$	σ^2	$\overline{m}$	$\widehat{n}$	BC0 (z-test)	BC1 & BC3 (z-test)	BC2 (z-test)	BC4 (z-test)	Rand test
4	0.5	10	48	0.079	0.077	0.071	0.075	0.065
4	0.5	30	22	0.080	0.072	**0.061**	**0.063**	**0.044**
4	0.5	50	16	0.087	0.074	0.068	0.069	**0.048**
2	0.5	10	24	0.095	0.089	0.083	0.085	**0.062**
2	0.5	30	14	0.106	0.098	0.087	0.090	**0.056**
2	0.5	50	12	0.114	0.104	0.082	0.094	**0.046**
2	0.5	10	46	**0.061**	**0.057**	**0.057**	**0.057**	**0.045**
2	0.5	30	28	0.076	0.070	0.069	0.069	**0.056**
2	0.5	50	24	0.091	0.085	0.081	0.080	**0.063**
2	0.25	10	40	**0.062**	**0.060**	**0.054**	**0.056**	**0.049**
2	0.25	30	20	0.082	0.074	0.072	0.071	**0.056**
2	0.25	50	16	0.097	0.083	0.071	0.075	**0.053**
2	0.05	10	34	0.071	0.068	0.066	0.066	**0.055**
2	0.05	30	14	0.110	0.095	0.081	0.084	**0.058**
2	0.05	50	10	0.113	0.092	0.068	0.076	**0.044**
1	0.05	10	10	0.104	0.090	0.068	0.069	**0.047**
1	0.05	10	18	0.080	0.072	**0.063**	**0.064**	**0.047**
1	0.05	30	8	0.149	0.124	0.111	0.111	**0.057**
1	0.05	50	6	0.187	0.153	0.121	0.117	0.002
1	0.05	10	52	0.067	0.065	**0.064**	**0.064**	**0.057**
1	0.05	30	22	0.067	**0.063**	**0.057**	**0.056**	**0.046**
1	0.05	50	16	0.093	0.080	0.070	0.076	**0.054**

WEB TABLE 3 Empirical type I error rates for the z-test and randomization test from a simulation study based on different parameter configurations of the baseline hazard ($\lambda_0(t)$), frailty variance (σ^2), average cluster size ($\overline{m}$), and estimated number of clusters ($\widehat{n}$) when the coefficient of variation of cluster size is 0.5. BC0: Uncorrected sandwich variance estimator; BC1: bias-corrected sandwich variance estimator extending Kauermann and Carroll; BC2: bias-corrected sandwich variance estimator extending Mancl and DeRouen; BC3: bias-corrected sandwich variance estimator extending Fay and Graubard; BC4: bias-corrected sandwich variance estimator extending Morel et al; Rand: randomization-based test. Bold values indicate values close to nominal empirical type I error rate (from 3.6% to 6.4%).

Configuration				Type I error				
$\lambda_0(t)$	σ^2	$\overline{m}$	$\widehat{n}$	BC0 (z-test)	BC1 & BC3 (z-test)	BC2 (z-test)	BC4 (z-test)	Rand test
4	0.5	10	50	**0.064**	**0.062**	**0.060**	**0.062**	**0.053**
4	0.5	30	24	0.092	0.083	0.071	0.079	**0.054**
4	0.5	50	18	0.121	0.107	0.096	0.102	**0.061**
2	0.5	10	26	0.084	0.076	**0.063**	0.071	**0.053**
2	0.5	30	16	0.124	0.102	0.090	0.097	**0.062**
2	0.5	50	14	0.122	0.103	0.076	0.089	**0.046**
2	0.5	10	50	**0.060**	**0.058**	**0.052**	**0.056**	**0.047**
2	0.5	30	32	0.076	0.072	0.065	0.071	**0.049**
2	0.5	50	28	0.099	0.088	0.075	0.085	**0.053**
2	0.25	10	42	0.068	**0.063**	**0.061**	**0.062**	**0.049**
2	0.25	30	22	0.087	0.080	0.075	0.080	**0.052**
2	0.25	50	18	0.123	0.105	0.093	0.102	**0.064**
2	0.05	10	34	0.066	**0.061**	**0.059**	**0.060**	**0.049**
2	0.05	30	14	0.103	0.089	0.077	0.079	**0.051**
2	0.05	50	10	0.116	0.098	0.082	0.090	**0.043**
1	0.05	10	10	0.132	0.120	0.100	0.101	**0.057**
1	0.05	10	18	0.082	0.073	**0.063**	0.065	**0.041**
1	0.05	30	8	0.171	0.135	0.104	0.111	**0.039**
1	0.05	50	6	0.215	0.160	0.120	0.123	0.004
1	0.05	10	54	**0.057**	**0.054**	**0.051**	**0.054**	**0.045**
1	0.05	30	24	0.085	0.077	0.067	0.075	**0.050**
1	0.05	50	18	0.097	0.082	0.076	0.080	**0.056**

WEB TABLE 4 Empirical type I error rates for the z-test and randomization test from a simulation study based on different parameter configurations of the baseline hazard ($\lambda_0(t)$), frailty variance (σ^2), average cluster size ($\overline{m}$), and estimated number of clusters ($\widehat{n}$) when the coefficient of variation of cluster size is 0.75. BC0: Uncorrected sandwich variance estimator; BC1: bias-corrected sandwich variance estimator extending Kauermann and Carroll; BC2: bias-corrected sandwich variance estimator extending Mancl and DeRouen; BC3: bias-corrected sandwich variance estimator extending Fay and Graubard; BC4: bias-corrected sandwich variance estimator extending Morel et al; Rand: randomization-based test. Bold values indicate values close to nominal empirical type I error rate (from 3.6% to 6.4%).

Configuration				Type I error				
$\lambda_0(t)$	σ^2	$\overline{m}$	$\widehat{n}$	BC0 (z-test)	BC1 & BC3 (z-test)	BC2 (z-test)	BC4 (z-test)	Rand test
4	0.5	10	52	0.089	0.081	0.075	0.078	**0.058**
4	0.5	30	26	0.086	0.074	0.065	0.074	**0.052**
4	0.5	50	20	0.097	0.083	0.072	0.081	**0.048**
2	0.5	10	28	0.078	0.068	**0.060**	0.067	**0.039**
2	0.5	30	18	0.128	0.119	0.105	0.116	0.069
2	0.5	50	16	0.146	0.133	0.119	0.129	**0.064**
2	0.5	10	56	**0.062**	**0.058**	**0.054**	**0.057**	**0.046**
2	0.5	30	36	0.074	0.065	**0.058**	0.067	**0.045**
2	0.5	50	32	0.106	0.092	0.079	0.091	**0.060**
2	0.25	10	44	0.071	**0.063**	**0.057**	**0.062**	**0.047**
2	0.25	30	24	0.103	0.093	0.079	0.090	**0.051**
2	0.25	50	20	0.094	0.082	0.071	0.082	**0.044**
2	0.05	10	36	0.068	**0.059**	**0.055**	**0.058**	**0.045**
2	0.05	30	14	0.134	0.114	0.098	0.107	**0.062**
2	0.05	50	10	0.154	0.128	0.103	0.124	**0.047**
1	0.05	10	10	0.133	0.109	0.079	0.088	0.035
1	0.05	30	6	0.246	0.183	0.125	0.135	0.004
1	0.05	10	18	0.097	0.082	0.070	0.078	**0.044**
1	0.05	30	8	0.186	0.155	0.120	0.128	**0.041**
1	0.05	50	6	0.261	0.197	0.151	0.161	0.004
1	0.05	10	56	**0.057**	**0.056**	**0.052**	**0.056**	**0.045**
1	0.05	30	26	0.087	0.074	0.066	0.071	**0.048**
1	0.05	50	20	0.095	0.088	0.077	0.082	**0.048**

WEB TABLE 5 Empirical type I error rates for the Wald t-test and randomization test from a simulation study based on different parameter configurations of the baseline hazard ($\lambda_0(t)$), frailty variance (σ^2), average cluster size ($\overline{m}$), and estimated number of clusters ($\widehat{n}$) when the coefficient of variation of cluster size is 0.25. BC0: Uncorrected sandwich variance estimator; BC1: bias-corrected sandwich variance estimator extending Kauermann and Carroll; BC2: bias-corrected sandwich variance estimator extending Mancl and DeRouen; BC3: bias-corrected sandwich variance estimator extending Fay and Graubard; BC4: bias-corrected sandwich variance estimator extending Morel et al; Rand: randomization-based test. Bold values indicate values close to nominal empirical type I error rate (from 3.6% to 6.4%).

Configuration				Type I error				
$\lambda_0(t)$	σ^2	$\overline{m}$	$\widehat{n}$	BC0 (t-test)	BC1 & BC3 (t-test)	BC2 (t-test)	BC4 (t-test)	Rand test
4	0.5	10	50	**0.060**	**0.057**	**0.055**	**0.057**	**0.054**
4	0.5	30	24	0.073	**0.062**	**0.057**	**0.058**	**0.058**
4	0.5	50	18	0.083	0.072	0.068	0.069	0.066
2	0.5	10	26	**0.061**	**0.057**	**0.052**	**0.054**	**0.058**
2	0.5	30	16	0.086	0.075	0.068	0.073	0.065
2	0.5	50	14	0.070	**0.059**	**0.054**	**0.050**	**0.046**
2	0.5	10	48	**0.060**	**0.059**	**0.056**	**0.056**	**0.054**
2	0.5	30	30	**0.059**	**0.056**	**0.054**	**0.054**	**0.049**
2	0.5	50	26	0.068	**0.064**	**0.058**	**0.061**	**0.053**
2	0.25	10	42	**0.052**	**0.048**	**0.045**	**0.046**	**0.045**
2	0.25	30	22	**0.056**	**0.051**	**0.047**	**0.049**	**0.044**
2	0.25	50	18	0.081	0.08	0.068	0.072	0.067
2	0.05	10	36	**0.053**	**0.049**	**0.047**	**0.047**	**0.047**
2	0.05	30	16	0.072	**0.061**	**0.054**	**0.056**	**0.057**
2	0.05	50	12	0.070	**0.055**	**0.049**	**0.047**	**0.045**
1	0.05	10	12	0.087	0.073	**0.054**	**0.060**	**0.053**
1	0.05	30	6	0.112	0.084	**0.059**	**0.043**	0.002
1	0.05	50	6	0.117	0.087	**0.061**	**0.052**	0.002
1	0.05	10	20	0.069	**0.061**	**0.06**	**0.055**	**0.061**
1	0.05	30	10	0.081	0.072	**0.062**	**0.059**	**0.058**
1	0.05	50	8	0.092	0.072	**0.052**	**0.051**	**0.036**
1	0.05	10	54	**0.062**	**0.059**	**0.057**	**0.057**	**0.058**
1	0.05	30	24	0.067	**0.060**	**0.059**	**0.059**	**0.057**
1	0.05	50	18	0.067	**0.059**	**0.056**	**0.057**	**0.054**

WEB TABLE 6 Empirical type I error rates for the Wald t-test and randomization test from a simulation study based on different parameter configurations of the baseline hazard ($\lambda_0(t)$), frailty variance (σ^2), average cluster size ($\overline{m}$), and estimated number of clusters ($\widehat{n}$) when the coefficient of variation of cluster size is 0.5. BC0: Uncorrected sandwich variance estimator; BC1: bias-corrected sandwich variance estimator extending Kauermann and Carroll; BC2: bias-corrected sandwich variance estimator extending Mancl and DeRouen; BC3: bias-corrected sandwich variance estimator extending Fay and Graubard; BC4: bias-corrected sandwich variance estimator extending Morel et al; Rand: randomization-based test. Bold values indicate values close to nominal empirical type I error rate (from 3.6% to 6.4%).

Configuration				Type I error				
$\lambda_0(t)$	σ^2	$\overline{m}$	$\widehat{n}$	BC0 (t-test)	BC1 & BC3 (t-test)	BC2 (t-test)	BC4 (t-test)	Rand test
4	0.5	10	52	**0.040**	**0.039**	**0.038**	**0.038**	**0.037**
4	0.5	30	26	**0.055**	**0.047**	**0.041**	**0.042**	**0.037**
4	0.5	50	20	0.067	**0.062**	**0.047**	**0.056**	**0.037**
2	0.5	10	28	0.066	**0.059**	**0.051**	**0.052**	**0.046**
2	0.5	30	18	0.081	0.074	0.069	0.072	**0.050**
2	0.5	50	16	0.101	0.089	0.078	0.085	**0.055**
2	0.5	10	52	**0.053**	**0.050**	**0.046**	**0.049**	**0.046**
2	0.5	30	34	0.072	**0.064**	**0.055**	**0.062**	**0.047**
2	0.5	50	30	0.073	0.068	**0.064**	0.069	**0.058**
2	0.25	10	44	**0.058**	**0.055**	**0.051**	**0.054**	**0.048**
2	0.25	30	24	0.069	**0.060**	**0.054**	**0.061**	**0.051**
2	0.25	50	20	0.068	**0.061**	**0.053**	**0.057**	**0.048**
2	0.05	10	36	**0.064**	**0.059**	**0.050**	**0.053**	**0.054**
2	0.05	30	16	0.077	0.066	**0.060**	**0.059**	**0.053**
2	0.05	50	12	0.090	0.066	**0.059**	**0.058**	**0.054**
1	0.05	10	12	**0.061**	**0.051**	**0.045**	**0.044**	**0.043**
1	0.05	30	6	0.123	0.104	0.072	0.067	0.001
1	0.05	50	6	0.120	0.094	**0.062**	**0.062**	0.004
1	0.05	10	20	**0.060**	**0.057**	**0.050**	**0.050**	**0.052**
1	0.05	30	10	0.098	0.088	0.077	0.068	**0.049**
1	0.05	50	8	0.119	0.096	0.074	0.076	**0.047**
1	0.05	10	56	**0.064**	**0.060**	**0.055**	**0.058**	**0.056**
1	0.05	30	26	0.066	**0.053**	**0.047**	**0.050**	**0.043**
1	0.05	50	20	0.075	0.071	**0.057**	**0.063**	**0.048**

WEB TABLE 7 Empirical type I error rates for the Wald t-test and randomization test from a simulation study based on different parameter configurations of the baseline hazard ($\lambda_0(t)$), frailty variance (σ^2), average cluster size ($\overline{m}$), and estimated number of clusters ($\widehat{n}$) when the coefficient of variation of cluster size is 0.75. BC0: Uncorrected sandwich variance estimator; BC1: bias-corrected sandwich variance estimator extending Kauermann and Carroll; BC2: bias-corrected sandwich variance estimator extending Mancl and DeRouen; BC3: bias-corrected sandwich variance estimator extending Fay and Graubard; BC4: bias-corrected sandwich variance estimator extending Morel et al; Rand: randomization-based test. Bold values indicate values close to nominal empirical type I error rate (from 3.6% to 6.4%).

Configuration				Type I error				
$\lambda_0(t)$	σ^2	$\overline{m}$	$\widehat{n}$	BC0 (t-test)	BC1 & BC3 (t-test)	BC2 (t-test)	BC4 (t-test)	Rand test
4	0.5	10	54	**0.054**	**0.049**	**0.046**	**0.047**	**0.045**
4	0.5	30	28	0.066	**0.063**	**0.055**	**0.061**	**0.044**
4	0.5	50	22	0.084	0.073	**0.061**	0.074	**0.044**
2	0.5	10	30	0.075	**0.063**	**0.056**	0.065	**0.047**
2	0.5	30	20	0.093	0.079	0.070	0.077	**0.053**
2	0.5	50	18	0.102	0.093	0.071	0.088	**0.058**
2	0.5	10	58	0.066	**0.056**	**0.050**	**0.054**	**0.046**
2	0.5	30	38	0.080	0.075	0.069	0.077	**0.055**
2	0.5	50	34	0.078	0.071	0.066	0.071	**0.050**
2	0.25	10	46	0.070	0.066	**0.060**	**0.064**	**0.051**
2	0.25	30	26	0.070	**0.059**	**0.052**	**0.059**	**0.050**
2	0.25	50	22	0.083	0.073	**0.060**	0.072	**0.050**
2	0.05	10	38	**0.048**	**0.039**	0.035	**0.038**	**0.039**
2	0.05	30	16	0.086	0.075	0.068	0.073	**0.048**
2	0.05	50	12	0.105	0.090	0.070	0.083	**0.054**
1	0.05	10	12	0.079	0.070	**0.054**	**0.059**	0.035
1	0.05	30	8	0.128	0.097	0.079	0.075	**0.041**
1	0.05	50	6	0.155	0.122	0.082	0.088	0.004
1	0.05	10	20	0.071	0.065	**0.053**	**0.059**	**0.045**
1	0.05	30	10	0.120	0.099	0.073	0.080	**0.057**
1	0.05	50	8	0.140	0.113	0.080	0.094	**0.049**
1	0.05	10	58	**0.055**	**0.054**	**0.050**	**0.052**	**0.056**
1	0.05	30	28	0.068	**0.064**	**0.056**	**0.062**	**0.045**
1	0.05	50	22	0.084	0.068	**0.062**	0.069	**0.044**

3 | WEB TABLES ON POWER

WEB TABLE 8 Predicted and empirical power for the z-test and randomization test from a simulation study based on different parameter configurations of the baseline hazard ($\lambda_0(t)$), effect size (δ), frailty variance (σ^2), average cluster size ($\overline{m}$), and estimated number of clusters ($\widehat{n}$) when the coefficient of variation of cluster size is 0. Predicted: prediction power by the proposed formula; BC0: Uncorrected sandwich variance estimator; BC1: bias-corrected sandwich variance estimator extending Kauermann and Carroll; BC2: bias-corrected sandwich variance estimator extending Mancl and DeRouen; BC3: bias-corrected sandwich variance estimator extending Fay and Graubard; BC4: bias-corrected sandwich variance estimator extending Morel et al; Rand: randomization-based test. Bold values indicate the empirical power differs at most by 2.6% from the predicted power.

Configuration					Power					
$\lambda_0(t)$	β_0	θ_0	$\overline{m}$	$\widehat{n}$	Predicted	BC0 (z-test)	BC1 & BC3 (z-test)	BC2 (z-test)	BC4 (z-test)	Rand test
4	1.5	0.5	10	48	0.814	**0.811**	**0.803**	**0.795**	**0.801**	**0.789**
4	1.5	0.5	30	22	0.832	0.868	**0.853**	**0.831**	**0.839**	0.798
4	1.5	0.5	50	16	0.825	0.866	**0.850**	**0.825**	**0.833**	0.779
2	1.5	0.5	10	24	0.822	0.855	**0.839**	**0.821**	**0.832**	**0.805**
2	1.5	0.5	30	14	0.845	0.880	**0.862**	**0.831**	**0.852**	0.779
2	1.5	0.5	50	12	0.854	0.887	**0.865**	**0.833**	**0.846**	0.769
2	1	0.5	10	46	0.810	0.841	**0.828**	**0.816**	**0.824**	**0.814**
2	1	0.5	30	26	0.801	**0.810**	**0.795**	**0.782**	**0.791**	0.759
2	1	0.5	50	24	0.830	0.858	**0.843**	**0.819**	**0.836**	0.797
2	1	0.25	10	40	0.817	**0.821**	**0.812**	**0.803**	**0.810**	**0.799**
2	1	0.25	30	20	0.838	0.881	0.865	**0.842**	**0.852**	**0.822**
2	1	0.25	50	16	0.848	0.877	**0.859**	**0.834**	**0.848**	0.786
2	1	0.05	10	34	0.809	**0.826**	**0.818**	**0.811**	**0.816**	**0.807**
2	1	0.05	30	14	0.853	0.891	**0.870**	**0.844**	**0.861**	0.806
2	1	0.05	50	8	0.805	0.872	**0.831**	0.775	**0.810**	0.627
1	1.5	0.05	10	10	0.820	0.873	**0.825**	0.761	**0.814**	0.732
1	1	0.05	10	18	0.828	0.866	**0.851**	**0.829**	**0.847**	0.806
1	1	0.05	30	8	0.868	0.906	**0.876**	0.831	**0.868**	0.678
1	1	0.05	50	6	0.892	0.940	**0.896**	0.835	**0.884**	0.035
1	0.5	0.05	10	52	0.803	**0.823**	**0.818**	**0.814**	**0.816**	**0.803**
1	0.5	0.05	30	22	0.811	**0.820**	**0.807**	**0.795**	**0.802**	0.762
1	0.5	0.05	50	16	0.815	**0.838**	**0.815**	**0.795**	**0.804**	0.757

WEB TABLE 9 Predicted and empirical power for the z-test and randomization test from a simulation study based on different parameter configurations of the baseline hazard ($\lambda_0(t)$), effect size (δ), frailty variance (σ^2), average cluster size ($\overline{m}$), and estimated number of clusters ($\widehat{n}$) when the coefficient of variation of cluster size is 0.25. Predicted: prediction power by the proposed formula; BC0: Uncorrected sandwich variance estimator; BC1: bias-corrected sandwich variance estimator extending Kauermann and Carroll; BC2: bias-corrected sandwich variance estimator extending Mancl and DeRouen; BC3: bias-corrected sandwich variance estimator extending Fay and Graubard; BC4: bias-corrected sandwich variance estimator extending Morel et al; Rand: randomization-based test. Bold values indicate the empirical power differs at most by 2.6% from the predicted power.

Configuration					Power					
$\lambda_0(t)$	β_0	θ_0	$\overline{m}$	$\widehat{n}$	Predicted	BC0 (z-test)	BC1 & BC3 (z-test)	BC2 (z-test)	BC4 (z-test)	Rand test
4	1.5	0.5	10	48	0.810	**0.811**	**0.805**	**0.798**	**0.803**	**0.789**
4	1.5	0.5	30	22	0.824	**0.843**	**0.831**	**0.807**	**0.821**	0.778
4	1.5	0.5	50	16	0.814	0.875	0.855	**0.832**	0.849	0.775
2	1.5	0.5	10	24	0.815	0.849	**0.831**	**0.808**	**0.827**	**0.798**
2	1.5	0.5	30	14	0.832	0.888	0.861	**0.824**	**0.852**	0.766
2	1.5	0.5	50	12	0.839	0.883	**0.860**	**0.834**	**0.852**	0.763
2	1	0.5	10	46	0.801	**0.813**	**0.801**	**0.791**	**0.795**	**0.784**
2	1	0.5	30	28	0.814	0.850	**0.833**	**0.817**	**0.826**	**0.797**
2	1	0.5	50	24	0.813	**0.838**	**0.825**	**0.813**	**0.821**	0.771
2	1	0.25	10	40	0.812	0.849	0.842	**0.827**	**0.837**	**0.817**
2	1	0.25	30	20	0.828	0.866	0.858	**0.844**	**0.852**	**0.809**
2	1	0.25	50	16	0.835	0.883	0.869	**0.851**	**0.858**	0.800
2	1	0.05	10	34	0.808	0.836	**0.827**	**0.815**	**0.822**	**0.799**
2	1	0.05	30	14	0.850	0.893	**0.866**	**0.845**	**0.860**	0.811
2	1	0.05	50	10	0.879	0.906	**0.889**	**0.856**	**0.884**	0.780
1	1.5	0.05	10	10	0.818	0.897	0.855	0.782	0.854	0.767
1	1	0.05	10	18	0.826	0.886	0.862	**0.832**	0.855	**0.827**
1	1	0.05	30	8	0.863	0.919	**0.884**	0.826	**0.874**	0.671
1	1	0.05	50	6	0.885	0.933	**0.892**	0.818	**0.877**	0.039
1	0.5	0.05	10	52	0.800	0.846	0.843	0.834	0.837	**0.821**
1	0.5	0.05	30	22	0.804	0.837	**0.822**	**0.806**	**0.813**	**0.779**
1	0.5	0.05	50	16	0.806	0.841	**0.827**	**0.808**	**0.816**	0.751

WEB TABLE 10 Predicted and empirical power for the z-test and randomization test from a simulation study based on different parameter configurations of the baseline hazard ($\lambda_0(t)$), effect size (δ), frailty variance (σ^2), average cluster size ($\overline{m}$), and estimated number of clusters ($\widehat{n}$) when the coefficient of variation of cluster size is 0.5. Predicted: prediction power by the proposed formula; BC0: Uncorrected sandwich variance estimator; BC1: bias-corrected sandwich variance estimator extending Kauermann and Carroll; BC2: bias-corrected sandwich variance estimator extending Mancl and DeRouen; BC3: bias-corrected sandwich variance estimator extending Fay and Graubard; BC4: bias-corrected sandwich variance estimator extending Morel et al; Rand: randomization-based test. Bold values indicate the empirical power differs at most by 2.6% from the predicted power.

Configuration					Power					
$\lambda_0(t)$	β_0	θ_0	$\overline{m}$	$\widehat{n}$	Predicted	BC0 (z-test)	BC1 & BC3 (z-test)	BC2 (z-test)	BC4 (z-test)	Rand test
4	1.5	0.5	10	50	0.815	**0.836**	**0.827**	**0.821**	**0.828**	**0.815**
4	1.5	0.5	30	24	0.833	0.873	0.860	**0.846**	**0.856**	**0.819**
4	1.5	0.5	50	18	0.827	0.865	**0.850**	**0.829**	**0.842**	0.780
2	1.5	0.5	10	26	0.823	0.874	0.857	**0.838**	0.853	**0.809**
2	1.5	0.5	30	16	0.844	0.894	**0.865**	**0.838**	**0.865**	0.807
2	1.5	0.5	50	14	0.851	0.914	0.880	**0.854**	0.880	0.799
2	1	0.5	10	50	0.808	**0.834**	**0.828**	**0.824**	**0.828**	**0.798**
2	1	0.5	30	32	0.824	0.869	0.857	**0.838**	0.856	**0.816**
2	1	0.5	50	28	0.824	0.868	**0.850**	**0.834**	**0.847**	**0.820**
2	1	0.25	10	42	0.817	0.859	**0.843**	**0.833**	**0.840**	**0.818**
2	1	0.25	30	22	0.834	0.867	**0.851**	**0.829**	**0.845**	0.789
2	1	0.25	50	18	0.842	0.871	**0.861**	**0.838**	**0.857**	0.806
2	1	0.05	10	34	0.805	0.858	0.846	**0.831**	0.843	**0.823**
2	1	0.05	30	14	0.842	0.871	**0.852**	**0.808**	**0.848**	0.763
2	1	0.05	50	10	0.868	0.914	**0.884**	**0.842**	**0.877**	0.740
1	1.5	0.05	10	10	0.813	0.867	**0.819**	0.758	**0.828**	0.736
1	1	0.05	10	18	0.819	0.876	0.860	**0.831**	0.856	**0.804**
1	1	0.05	30	8	0.848	0.901	**0.857**	0.782	**0.847**	0.603
1	1	0.05	50	6	0.866	0.935	**0.888**	0.786	**0.882**	0.039
1	0.5	0.05	10	54	0.806	0.848	0.840	**0.832**	0.837	**0.820**
1	0.5	0.05	30	24	0.817	0.856	**0.834**	**0.810**	**0.823**	0.771
1	0.5	0.05	50	18	0.823	**0.843**	**0.825**	**0.807**	**0.820**	0.775

WEB TABLE 11 Predicted and empirical power for the z-test and randomization test from a simulation study based on different parameter configurations of the baseline hazard ($\lambda_0(t)$), effect size (δ), frailty variance (σ^2), average cluster size ($\overline{m}$), and estimated number of clusters ($\widehat{n}$) when the coefficient of variation of cluster size is 0.75. Predicted: prediction power by the proposed formula; BC0: Uncorrected sandwich variance estimator; BC1: bias-corrected sandwich variance estimator extending Kauermann and Carroll; BC2: bias-corrected sandwich variance estimator extending Mancl and DeRouen; BC3: bias-corrected sandwich variance estimator extending Fay and Graubard; BC4: bias-corrected sandwich variance estimator extending Morel et al; Rand: randomization-based test. Bold values indicate the empirical power differs at most by 2.6% from the predicted power.

Configuration					Power					
$\lambda_0(t)$	β_0	θ_0	$\overline{m}$	$\widehat{n}$	Predicted	BC0 (z-test)	BC1 & BC3 (z-test)	BC2 (z-test)	BC4 (z-test)	Rand test
4	1.5	0.5	10	52	0.812	**0.821**	**0.810**	**0.800**	**0.809**	**0.795**
4	1.5	0.5	30	26	0.827	**0.850**	**0.832**	**0.814**	**0.835**	0.792
4	1.5	0.5	50	20	0.819	0.896	0.877	0.852	0.876	**0.805**
2	1.5	0.5	10	28	0.817	0.859	**0.838**	**0.813**	0.847	**0.804**
2	1.5	0.5	30	18	0.832	0.893	0.868	**0.835**	0.874	0.799
2	1.5	0.5	50	16	0.838	0.908	0.880	**0.838**	0.888	**0.814**
2	1	0.5	10	56	0.813	0.850	**0.836**	**0.826**	**0.839**	**0.815**
2	1	0.5	30	36	0.808	0.862	0.846	0.836	0.853	**0.817**
2	1	0.5	50	32	0.806	0.863	0.847	**0.823**	0.849	**0.816**
2	1	0.25	10	44	0.811	0.852	0.840	**0.824**	0.840	**0.811**
2	1	0.25	30	24	0.823	0.870	0.856	**0.829**	0.857	**0.801**
2	1	0.25	50	20	0.828	0.900	0.884	0.862	0.886	**0.825**
2	1	0.05	10	36	0.821	0.859	**0.844**	**0.826**	**0.845**	**0.814**
2	1	0.05	30	14	0.828	0.856	**0.825**	0.783	**0.826**	0.722
2	1	0.05	50	10	0.848	0.894	**0.858**	0.797	**0.861**	0.674
1	1.5	0.05	10	10	0.804	0.874	**0.820**	0.730	0.837	0.683
1	1.5	0.05	30	6	0.928	**0.950**	0.892	0.782	**0.917**	0.019
1	1	0.05	10	18	0.808	0.862	0.836	**0.787**	0.844	0.767
1	1	0.05	30	8	0.824	0.888	**0.835**	0.742	**0.836**	0.550
1	1	0.05	50	6	0.833	0.907	**0.851**	0.735	**0.853**	0.033
1	0.5	0.05	10	56	0.806	0.842	**0.828**	**0.820**	**0.828**	**0.798**
1	0.5	0.05	30	26	0.816	**0.837**	**0.819**	**0.793**	**0.820**	0.775
1	0.5	0.05	50	20	0.821	0.888	0.866	**0.839**	0.867	**0.811**

WEB TABLE 12 Predicted and empirical power for the Wald t-test and randomization test from a simulation study based on different parameter configurations of the baseline hazard ($\lambda_0(t)$), effect size (δ), frailty variance (σ^2), average cluster size ($\overline{m}$), and estimated number of clusters ($\widehat{n}$) when the coefficient of variation of cluster size is 0.25. Predicted: prediction power by the proposed formula; BC0: Uncorrected sandwich variance estimator; BC1: bias-corrected sandwich variance estimator extending Kauermann and Carroll; BC2: bias-corrected sandwich variance estimator extending Mancl and DeRouen; BC3: bias-corrected sandwich variance estimator extending Fay and Graubard; BC4: bias-corrected sandwich variance estimator extending Morel et al; Rand: randomization-based test. Bold values indicate the empirical power differs at most by 2.6% from the predicted power.

Configuration					Power					
$\lambda_0(t)$	β_0	θ_0	$\overline{m}$	$\widehat{n}$	Predicted	BC0 (t-test)	BC1 & BC3 (t-test)	BC2 (t-test)	BC4 (t-test)	Rand test
4	1.5	0.5	10	50	0.810	0.851	0.843	**0.836**	0.841	0.843
4	1.5	0.5	30	24	0.824	0.862	**0.843**	**0.826**	**0.835**	**0.822**
4	1.5	0.5	50	18	0.814	0.860	0.849	**0.817**	**0.837**	**0.829**
2	1.5	0.5	10	26	0.815	0.853	**0.832**	**0.814**	**0.836**	**0.824**
2	1.5	0.5	30	16	0.831	0.876	0.859	**0.830**	**0.843**	**0.825**
2	1.5	0.5	50	14	0.838	0.895	0.875	**0.838**	**0.858**	**0.843**
2	1	0.5	10	48	0.801	**0.822**	**0.814**	**0.800**	**0.807**	**0.805**
2	1	0.5	30	30	0.814	0.857	0.843	**0.833**	**0.838**	**0.840**
2	1	0.5	50	26	0.813	**0.832**	**0.822**	**0.805**	**0.815**	**0.804**
2	1	0.25	10	42	0.813	0.849	**0.839**	**0.833**	**0.835**	**0.838**
2	1	0.25	30	22	0.828	**0.854**	**0.836**	**0.819**	**0.829**	**0.822**
2	1	0.25	50	18	0.835	0.876	0.865	**0.847**	**0.856**	**0.846**
2	1	0.05	10	36	0.808	0.854	0.840	**0.828**	0.838	0.839
2	1	0.05	30	16	0.850	0.889	**0.870**	**0.843**	**0.860**	**0.856**
2	1	0.05	50	12	0.878	0.913	**0.891**	**0.856**	**0.879**	**0.868**
1	1.5	0.05	10	12	0.817	0.881	**0.834**	0.762	**0.821**	**0.838**
1	1.5	0.05	30	6	0.820	0.900	**0.814**	0.702	**0.805**	0.037
1	1.5	0.05	50	6	0.931	0.979	**0.951**	0.875	**0.943**	0.034
1	1	0.05	10	20	0.826	0.871	0.858	**0.833**	**0.852**	**0.850**
1	1	0.05	30	10	0.861	0.910	**0.881**	0.831	**0.870**	**0.843**
1	1	0.05	50	8	0.881	0.948	**0.903**	0.845	**0.891**	0.831
1	0.5	0.05	10	54	0.801	0.835	**0.827**	**0.820**	**0.823**	**0.827**
1	0.5	0.05	30	24	0.804	**0.817**	**0.809**	**0.793**	**0.802**	**0.798**
1	0.5	0.05	50	18	0.805	0.854	0.838	**0.820**	**0.828**	**0.819**

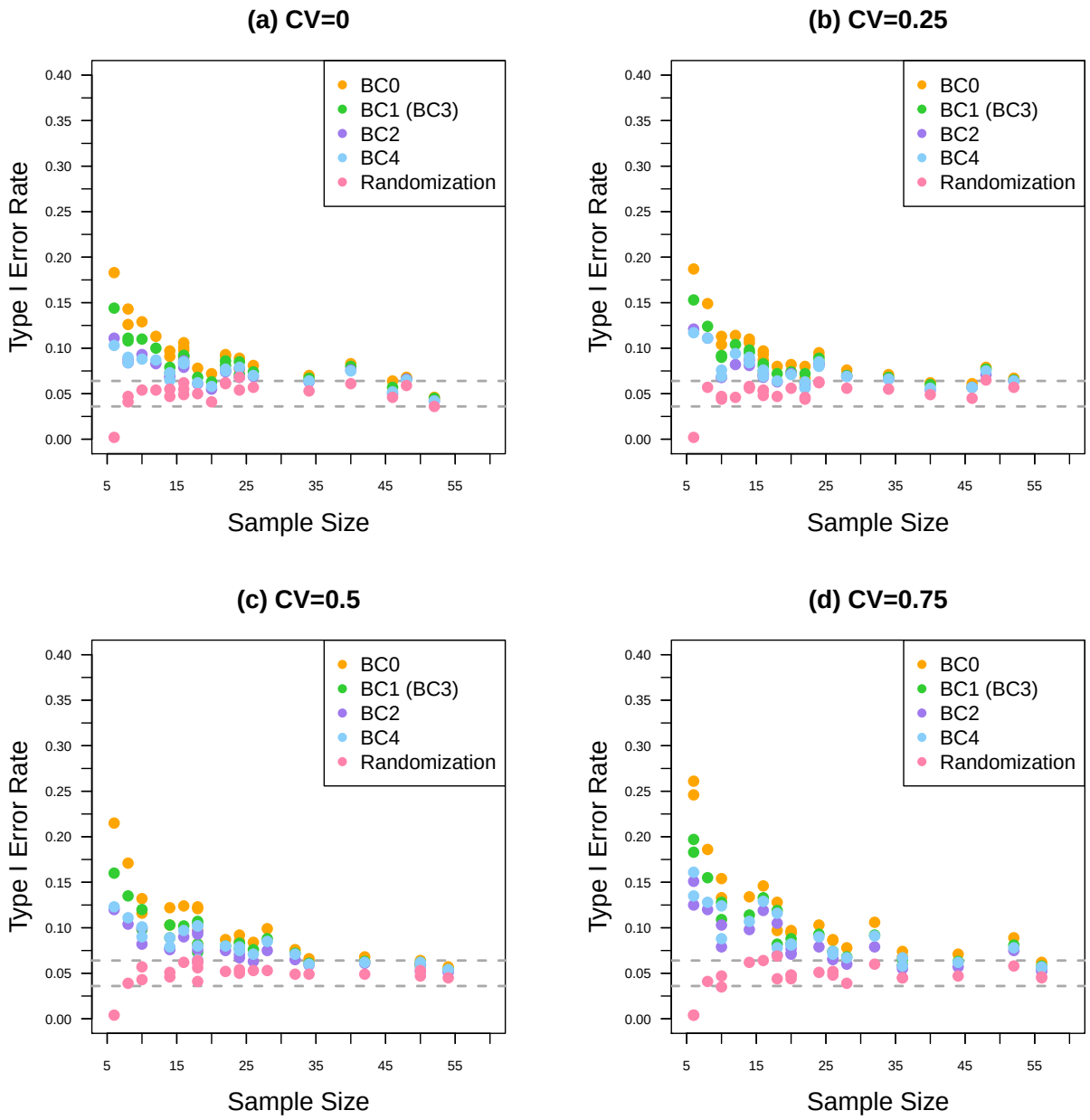
WEB TABLE 13 Predicted and empirical power for the Wald t-test and randomization test from a simulation study based on different parameter configurations of the baseline hazard ($\lambda_0(t)$), effect size (δ), frailty variance (σ^2), average cluster size ($\overline{m}$), and estimated number of clusters ($\widehat{n}$) when the coefficient of variation of cluster size is 0.5. Predicted: prediction power by the proposed formula; BC0: Uncorrected sandwich variance estimator; BC1: bias-corrected sandwich variance estimator extending Kauermann and Carroll; BC2: bias-corrected sandwich variance estimator extending Mancl and DeRouen; BC3: bias-corrected sandwich variance estimator extending Fay and Graubard; BC4: bias-corrected sandwich variance estimator extending Morel et al; Rand: randomization-based test. Bold values indicate the empirical power differs at most by 2.6% from the predicted power.

Configuration					Power					
$\lambda_0(t)$	β_0	θ_0	$\overline{m}$	$\widehat{n}$	Predicted	BC0 (t-test)	BC1 & BC3 (t-test)	BC2 (t-test)	BC4 (t-test)	Rand test
4	1.5	0.5	10	52	0.815	0.862	0.852	0.846	0.849	0.848
4	1.5	0.5	30	26	0.833	0.877	0.864	**0.841**	0.861	**0.843**
4	1.5	0.5	50	20	0.827	0.883	0.862	**0.838**	0.856	**0.838**
2	1.5	0.5	10	28	0.823	0.876	0.862	**0.843**	0.861	0.851
2	1.5	0.5	30	18	0.843	0.915	0.895	**0.868**	0.894	0.875
2	1.5	0.5	50	16	0.851	0.919	0.900	**0.876**	0.902	0.878
2	1	0.5	10	52	0.808	0.843	0.839	**0.832**	0.838	0.835
2	1	0.5	30	34	0.824	0.864	0.857	**0.846**	0.856	**0.844**
2	1	0.5	50	30	0.824	0.880	0.868	0.854	0.867	0.856
2	1	0.25	10	44	0.817	0.868	0.855	**0.843**	0.853	0.851
2	1	0.25	30	24	0.834	0.896	0.880	**0.858**	0.877	**0.860**
2	1	0.25	50	20	0.842	0.894	0.882	**0.862**	0.880	**0.864**
2	1	0.05	10	36	0.805	0.834	**0.823**	**0.814**	**0.824**	**0.814**
2	1	0.05	30	16	0.842	0.880	**0.862**	**0.825**	**0.856**	**0.836**
2	1	0.05	50	12	0.866	0.907	**0.888**	**0.844**	**0.877**	**0.855**
1	1.5	0.05	10	12	0.811	0.888	**0.829**	0.774	**0.834**	**0.827**
1	1.5	0.05	30	6	0.806	0.913	0.837	0.696	**0.828**	0.040
1	1.5	0.05	50	6	0.921	0.973	**0.926**	0.823	**0.934**	0.032
1	1	0.05	10	20	0.819	0.872	0.847	**0.815**	**0.843**	**0.839**
1	1	0.05	30	10	0.846	0.904	**0.867**	0.805	**0.867**	**0.827**
1	1	0.05	50	8	0.861	0.913	**0.877**	0.809	**0.867**	0.784
1	0.5	0.05	10	56	0.806	0.836	**0.829**	**0.824**	**0.829**	**0.832**
1	0.5	0.05	30	26	0.817	0.845	**0.831**	**0.818**	**0.824**	**0.819**
1	0.5	0.05	50	20	0.823	0.867	**0.847**	**0.821**	**0.845**	**0.831**

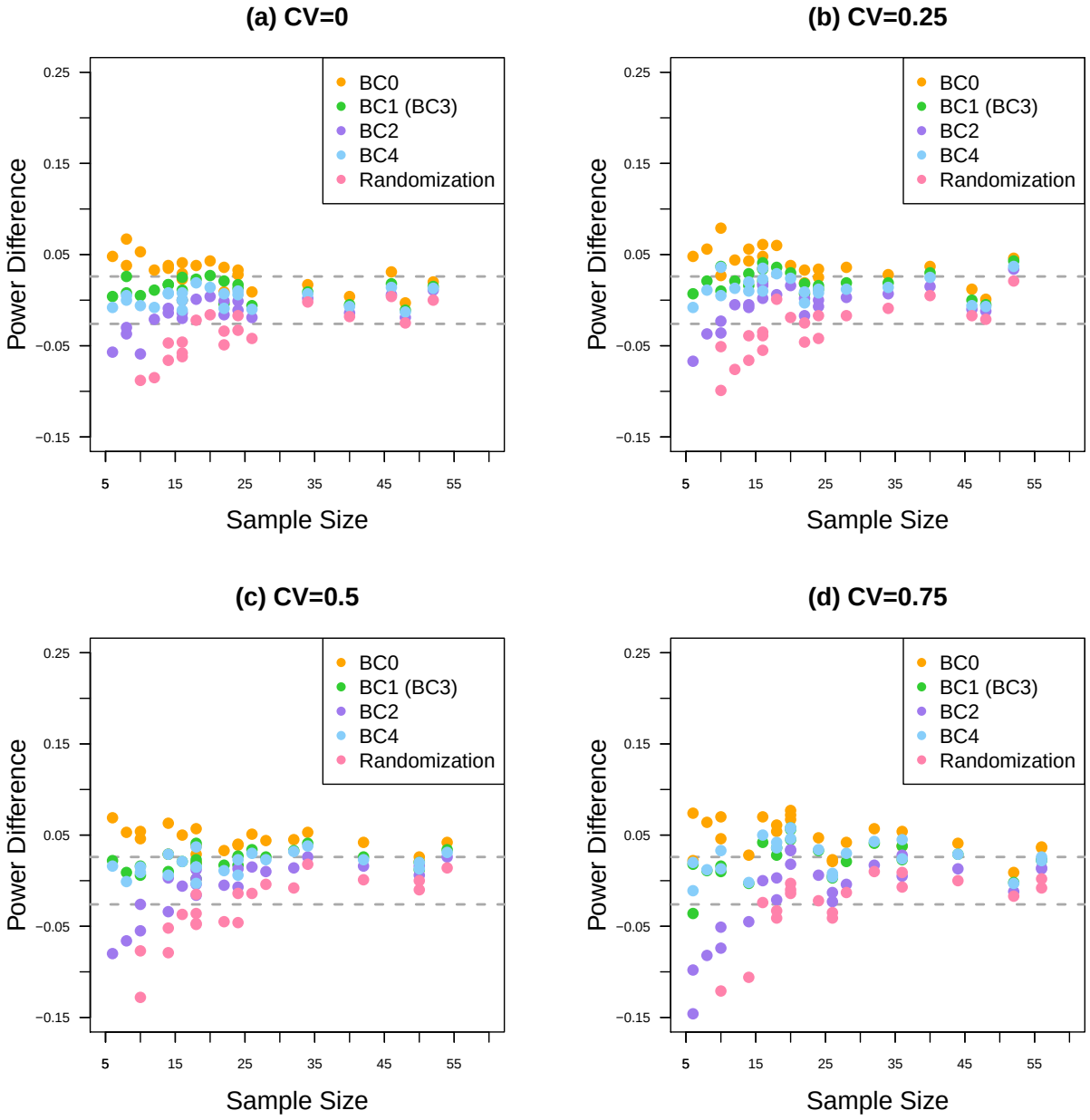
WEB TABLE 14 Predicted and empirical power for the Wald t-test and randomization test from a simulation study based on different parameter configurations of the baseline hazard ($\lambda_0(t)$), effect size (δ), frailty variance (σ^2), average cluster size ($\overline{m}$), and estimated number of clusters ($\widehat{n}$) when the coefficient of variation of cluster size is 0.75. Predicted: prediction power by the proposed formula; BC0: Uncorrected sandwich variance estimator; BC1: bias-corrected sandwich variance estimator extending Kauermann and Carroll; BC2: bias-corrected sandwich variance estimator extending Mancl and DeRouen; BC3: bias-corrected sandwich variance estimator extending Fay and Graubard; BC4: bias-corrected sandwich variance estimator extending Morel et al; Rand: randomization-based test. Bold values indicate the empirical power differs at most by 2.6% from the predicted power.

Configuration					Power					
$\lambda_0(t)$	β_0	θ_0	$\overline{m}$	$\widehat{n}$	Predicted	BC0 (t-test)	BC1 & BC3 (t-test)	BC2 (t-test)	BC4 (t-test)	Rand test
4	1.5	0.5	10	54	0.813	0.843	**0.836**	**0.825**	**0.835**	**0.828**
4	1.5	0.5	30	28	0.827	0.884	0.868	**0.850**	0.873	0.863
4	1.5	0.5	50	22	0.819	0.869	**0.845**	**0.826**	0.853	**0.818**
2	1.5	0.5	10	30	0.818	0.875	0.853	**0.832**	0.854	**0.842**
2	1.5	0.5	30	20	0.832	0.910	0.891	**0.858**	0.891	**0.846**
2	1.5	0.5	50	18	0.837	0.916	0.895	0.875	0.896	0.866
2	1	0.5	10	58	0.813	**0.831**	**0.814**	**0.804**	**0.816**	**0.816**
2	1	0.5	30	38	0.808	0.857	0.845	**0.829**	0.850	**0.834**
2	1	0.5	50	34	0.807	0.870	0.853	**0.830**	0.851	0.841
2	1	0.25	10	46	0.811	0.841	**0.830**	**0.821**	**0.834**	**0.824**
2	1	0.25	30	26	0.823	0.857	**0.838**	**0.818**	**0.842**	**0.812**
2	1	0.25	50	22	0.828	0.881	0.867	**0.838**	0.864	**0.831**
2	1	0.05	10	38	0.821	0.870	0.851	**0.832**	0.849	0.849
2	1	0.05	30	16	0.828	0.891	0.859	**0.824**	0.861	**0.811**
2	1	0.05	50	12	0.847	0.894	**0.858**	0.814	**0.861**	0.819
1	1.5	0.05	10	12	0.802	0.878	**0.815**	0.740	0.830	**0.790**
1	1.5	0.05	30	8	0.921	0.961	**0.908**	0.831	**0.931**	0.780
1	1.5	0.05	50	6	0.902	0.954	**0.892**	0.761	**0.909**	0.060
1	1	0.05	10	20	0.808	0.861	**0.831**	**0.796**	0.836	**0.817**
1	1	0.05	30	10	0.821	0.893	**0.842**	0.749	**0.846**	0.755
1	1	0.05	50	8	0.828	0.928	0.867	0.771	0.879	0.728
1	0.5	0.05	10	58	0.806	0.838	**0.831**	**0.823**	0.834	**0.821**
1	0.5	0.05	30	28	0.816	0.862	0.849	**0.830**	0.848	**0.836**
1	0.5	0.05	50	22	0.821	0.857	**0.837**	**0.816**	**0.834**	**0.804**

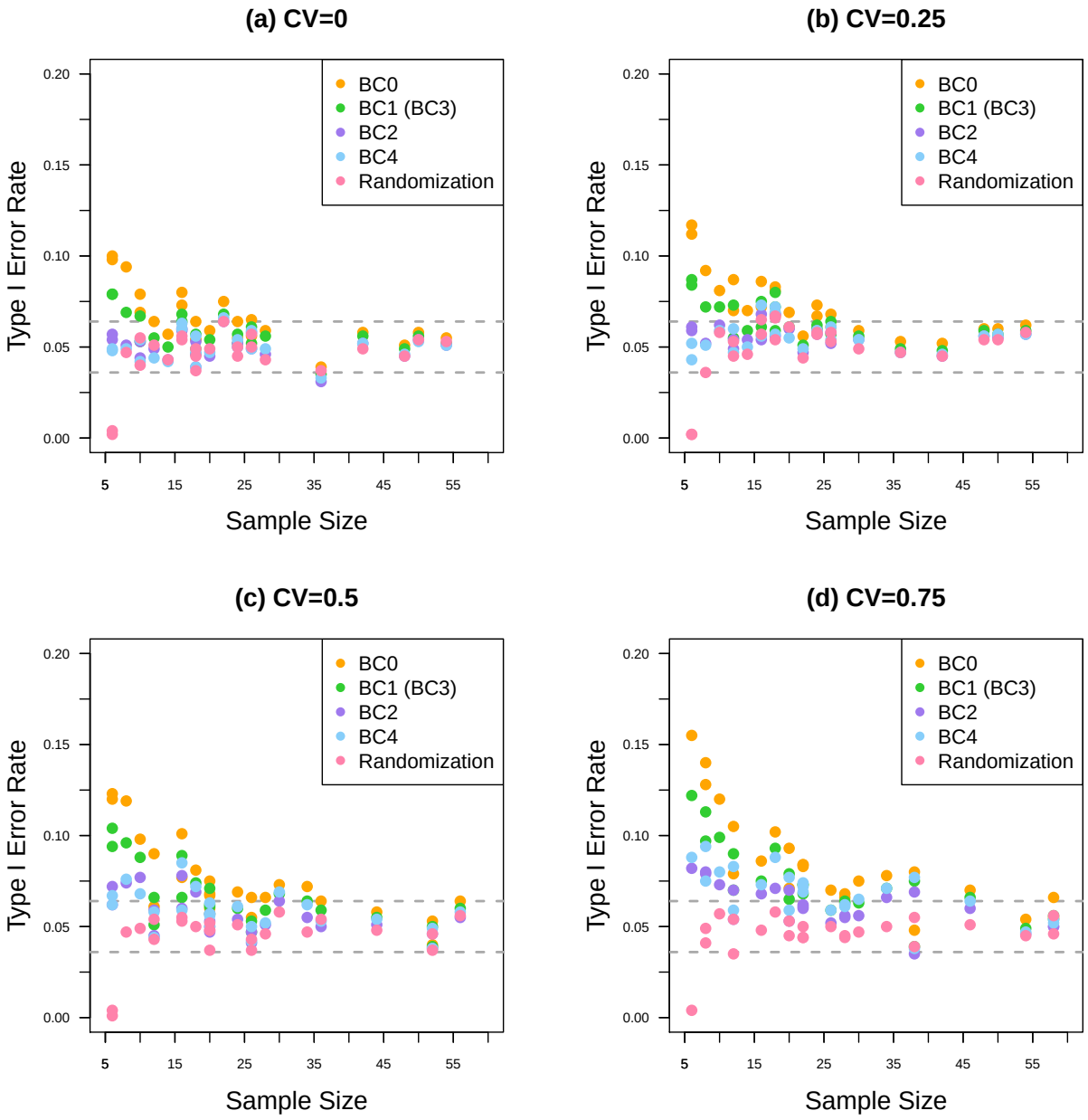
4 | WEB FIGURES AS ALTERNATIVE SUMMARIES OF SIMULATION RESULTS



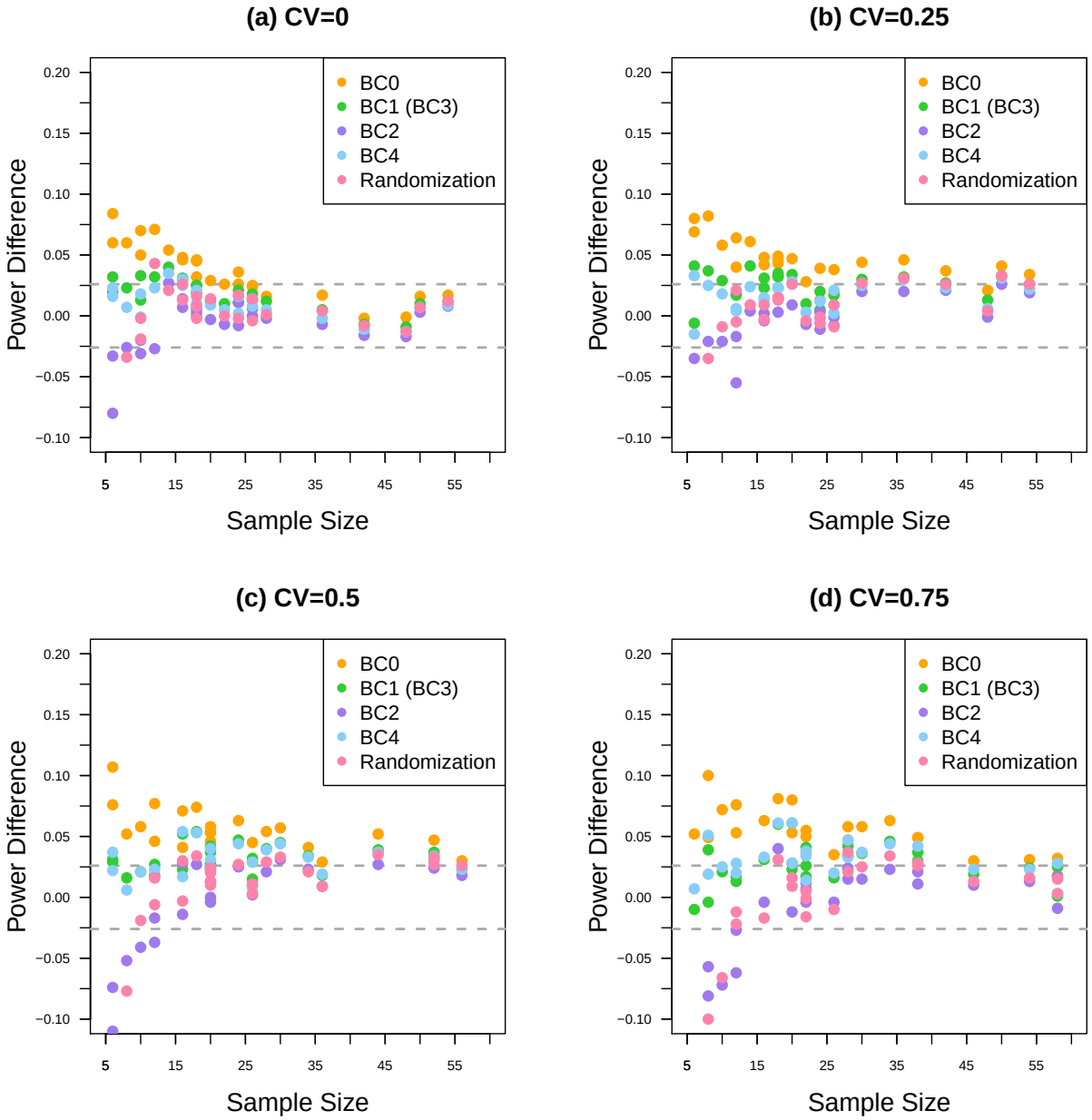
WEB FIGURE 1 Graphical summary of all simulation results on type I error rate based on the z-test and the randomization test. The dashed lines represent close to nominal empirical type I error rates (from 3.6% to 6.4%).



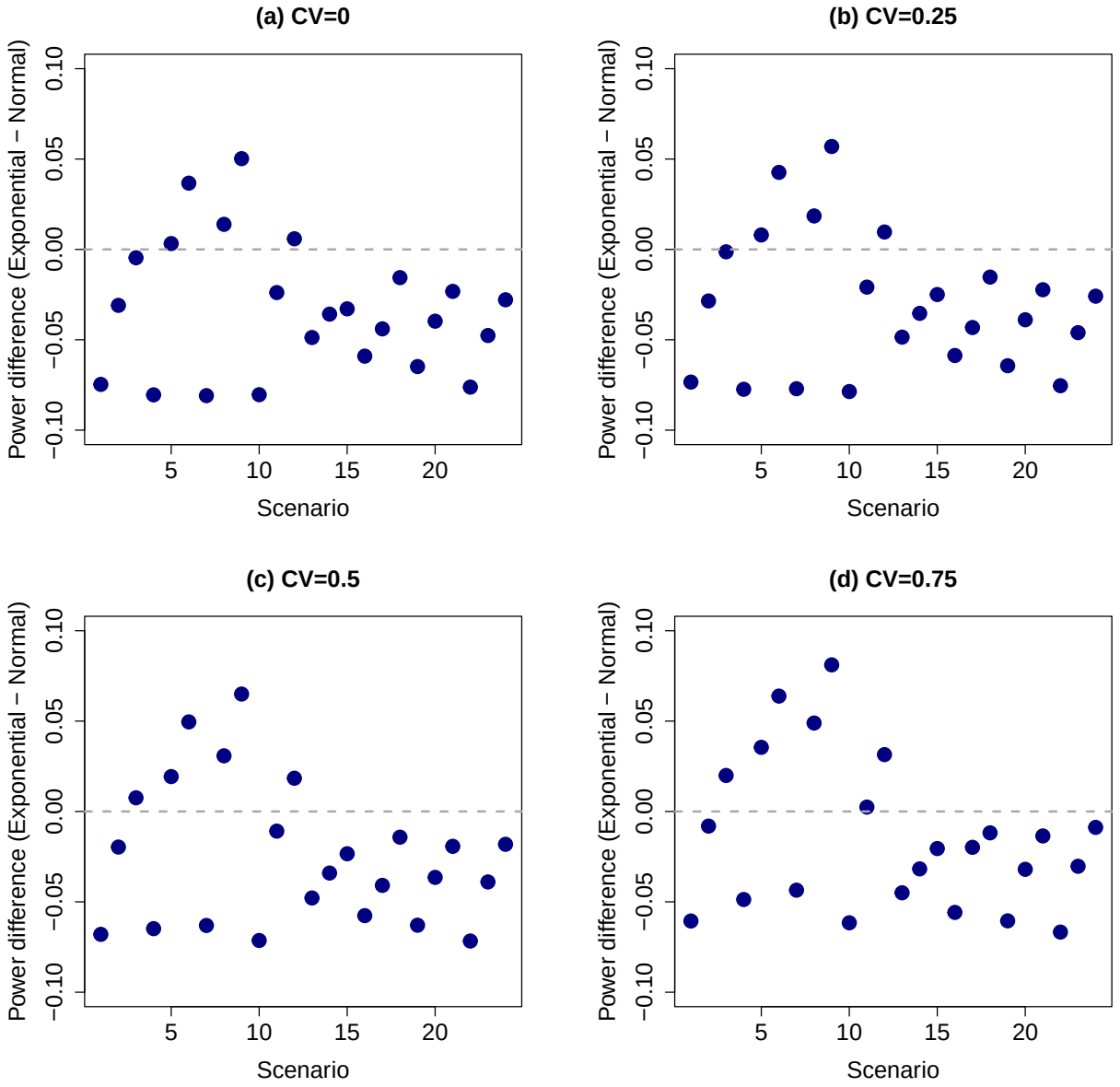
WEB FIGURE 2 Graphical summary of all simulation results on the difference between empirical and predicted power based on the z-test and the randomization test. The dashed lines indicate the empirical power differs by 2.6% from the predicted power.



WEB FIGURE 3 Graphical summary of all simulation results on type I error rate based on the *t*-test and the randomization test. The dashed lines represent close to nominal empirical type I error rates (from 3.6% to 6.4%).



WEB FIGURE 4 Graphical summary of all simulation results on the difference between empirical and predicted power based on the *t*-test and the randomization test. The dashed lines indicate the empirical power differs by 2.6% from the predicted power.



WEB FIGURE 5 A comparison of the predicted power under the AHMM with an exponential frailty versus a normal frailty for fixed sample size and design parameters (the variance of the normal and exponential frailty parameters are set to be the same for each scenario). Panels (a-d) correspond to sample size configurations and design parameters listed in Table 3 and Web Tables 12-14, respectively, for four different CV of cluster sizes. The scenario index number refers to the row number in the corresponding simulation results table. The gray dashed line corresponds to equality in power.