SUPPLEMENTARY MATERIAL

Supplement D: Comprehensive Plots of Type I Error and power
(doi: [10.1214/17-AOAS1028SUPPD](https://doi.org/10.1214/17-AOAS1028SUPPD); .pdf).

We provide additional supporting plots (see Figures 1-8) that show the type I error and power of all tests with equal/unequal sample sizes using generated data of independent observations or under provider-level clustering. Figures 1-5, and 7-8 have three columns, which correspond to Poisson, Negative Binomial, and Zero-inflated Poisson data respectively. Each row in these figures corresponds to a testing method.

In particular, Fig 1-2 show the type I error using independent observations with equal/unequal sample sizes, whereas Fig 3-4 show the type I error using correlated data that are generated using a mean-preserving random variable that follows a Normal/Gamma distribution. Fig 5 demonstrates coverage of confidence interval from ridge regression, and Fig 6 shows the type I error of the dynamic test. Fig 7-8 display the power using independent observations with equal/unequal sample sizes.

FIG 1. Type I error rates of all CPT code-specific tests with *unequal* sample sizes ($n_0 = 1000$, $n_1 = 3000$) using Poisson data, Negative Binomial data, or Zero-inflated Poisson data, each with a group mean of λ ranging from 10^{-6} to 10^2 , plotted on $\log_{10}$ scale.

FIG 2. Type I error rates of all CPT code-specific tests using Poisson data, Negative Binomial data, or Zero-inflated Poisson data with *equal* sample sizes ($n_0 = 1000$, $n_1 = 1000$), each with a group mean of λ ranging from 10^{-6} to 10^2 , plotted on $\log_{10}$ scale.

FIG 3. Type I error rates of t-test, dynamic test, and ridge test with provider-level clustering using Poisson data, Negative Binomial data, or Zero-inflated Poisson data. To introduce association, a mean-preserving random variable $\gamma_p^\beta \sim \text{Gamma}(\text{shape} = \frac{1}{\beta}, \text{scale} = \beta)$ with $E[\gamma_p^\beta] = 1$ and $\text{Var}[\gamma_p^\beta] = \beta$ is shared by patients treated by provider p . The cohorts have unequal sample sizes ($n_0 = 1,000$, $n_1 = 3,000$), each with a group mean of λ ranging from 10^{-6} to 10^2 , plotted on $\log_{10}$ scale.

FIG 4. Type I error rates of t-test, dynamic test, and ridge test with provider-level clustering using Poisson data, Negative Binomial data, or Zero-inflated Poisson data. To introduce association, a mean-preserving random variable $\gamma_p^\sigma \sim \text{Normal}(0, \sigma^2)$ with $E[\gamma_p^\sigma] = 0$ and $\text{Var}[\gamma_p^\sigma] = \sigma^2$ is shared by patients treated by provider p . The cohorts have unequal sample sizes ($n_0 = 1,000$, $n_1 = 3,000$), each with a group mean of λ ranging from 10^{-6} to 10^2 , plotted on $\log_{10}$ scale.

FIG 5. Coverage of confidence interval for ridge Poisson regression using Poisson data, Negative Binomial data, or Zero-inflated Poisson data with *unequal* sample sizes ($n_0 = 1000$, $n_1 = 3000$), each with a group mean of λ ranging from 10^{-6} to 10^2 , plotted on $\log_{10}$ scale.

FIG 6. Type I error rates of the dynamic test which is a mixture of the Negative Binomial exact test and the t-test, with unequal sample sizes ($n_0 = 1000$, $n_1 = 3000$) using Poisson data, Negative Binomial data, or Zero-inflated Poisson data, each with a group mean of λ ranging from 10^{-6} to 10^2 , plotted on $\log_{10}$ scale. Colored lines correspond to Negative Binomial ET (—); t-test (---); Dynamic test (—).

FIG 7. Power of all CPT code-specific tests with *unequal* sample sizes ($n_0 = 1000$, $n_1 = 3000$) using Negative Binomial data, Poisson data, or Zero-inflated Poisson data, each with a mean of $\lambda_0 = 1$ or 0.1 in cohort 0, and a rate ratio on log 2 scale ($\log_2 \frac{\lambda_1}{\lambda_0}$) ranging from 0 to 1.5. Colored lines correspond to Negative Binomial LRT (—); Negative Binomial ET (---); Poisson LRT (---); Poisson ET (---); Fisher's ET (---); Binomial LRT (---); t-test (---).

FIG 8. Power of all CPT code-specific tests with *equal* sample sizes ($n_0 = 1000$, $n_1 = 1000$) using Negative Binomial data, Poisson data, or Zero-inflated Poisson data, each with a mean of $\lambda_0 = 1$ or 0.01 in cohort 0, and a rate ratio on log 2 scale ($\log_2 \frac{\lambda_1}{\lambda_0}$) ranging from 0 to 1.5. Colored lines correspond to Negative Binomial LRT (—); Negative Binomial ET (---); Poisson LRT (---); Poisson ET (---); Fisher's ET (---); Binomial LRT (---); t-test (—).