Technical Appendix

A. Description of the Zero-inflated Negative Binomial (ZINB) model for modelling healthcare utilisation frequency

The ZINB model consists of two parts:

i. a logistic regression model to model the probability of excess zero, or those not incurring utilisation of a specific healthcare service, i.e. potential non-users, and
ii. a negative binomial model to model the frequency of utilisation of a specific healthcare service among patients with potential of incurring utilization, i.e. potential users.

The logistic model can be expressed by:

$$logit(p) = \beta_0 + \beta_1 INT + \beta_2 BASELINE + \beta_3 PS$$

where

$p =$ probability of excess zero for the specific healthcare service post − enrolment
$INT =$ indicator variable for patient enrolled in the intervention
$BASELINE =$ log transformed pre − enrolment utilisation with 0.5 added to zero values
$PS =$ propensity score $\left(i.e. \frac{\exp(\boldsymbol{X}'\widehat{\boldsymbol{\varphi}})}{1+\exp(\boldsymbol{X}'\widehat{\boldsymbol{\varphi}})}\right)$ where $\boldsymbol{X}$ is a matrix of matched variables and $\widehat{\boldsymbol{\varphi}}$ vector of the corresponding estimated parameters

For patients enrolled in the intervention, the odds of being 'not-at-risk' of incurring utilisation of a service would be $\boldsymbol{exp(\beta_1)}$ times the odds of patients who were not enrolled.

The negative binomial model can be expressed by:

$$log(\mu) = \gamma_0 + \gamma_1 INT + \gamma_2 BASELINE + \gamma_3 PS + \log(TIME)$$

where

$INT, BASELINE$ and PS have been defined
$\mu =$ mean frequency of the specific healthcare service post − enrolment given covariate
$TIME =$ follow − up time

For patients enrolled in the intervention, the incidence rate of utilisation frequency would be $\boldsymbol{\exp(\gamma_1)}$ times the incidence rate of patients who were not enrolled.

B. Description of the two-part model (twopm) for modelling healthcare charges

The *twopm* consists of

i. a probit regression model to model the probability of zero, or not incurring the specific healthcare charges, and
ii. a generalized linear model with log-link and gamma family to model charges of the specific healthcare service among patients incurring some charges.

The probit model can be expressed by:

$$\Phi^{-1}(q) = \beta_0 + \beta_1 INT + \beta_2 BASELINE + \beta_3 PS$$

where
$INT, BASELINE$ and PS have been defined
$\Phi^{-1} =$ the probit function (i. e. inverse cumulative distribution function for standard normal)
$q =$ probability of zero charges for the specific healthcare service post − enrolment

For patients enrolled in the intervention, the marginal effect (marginal probability) of zero charge for a specific healthcare service would be $\boldsymbol{\Phi\left(\hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_2\overline{BASELINE} + \hat{\beta}_3\overline{PS}\right)}$, where $\overline{\boldsymbol{BASELINE}}$ and $\overline{\boldsymbol{PS}}$ are mean values. For patients not enrolled in the intervention, the marginal effect would be $\boldsymbol{\Phi(\hat{\beta}_0 + \hat{\beta}_2\overline{BASELINE} + \hat{\beta}_3\overline{PS})}$.

The generalized linear model with log-link and gamma family can be expressed by:

$$log(\pi) = \gamma_0 + \gamma_1 INT + \gamma_2 BASELINE + \gamma_3 PS + \log(TIME)$$

where
$INT, BASELINE, PS$ and $TIME$ have been defined
$\pi =$ mean charges of the specific healthcare service post − enrolment given covariate

For patients enrolled in the intervention, their mean healthcare charges would be $\boldsymbol{\exp(\gamma_1)}$ times that of patients who were not enrolled.