**Supplemental Materials for**

**Jen et al, “*Hurdle Poisson Regression Model for Identifying Factors Related to Non-compliance and Waiting Time for Confirmatory Diagnosis in Colorectal Cancer Screening”***

**The Likelihood Function for Estimation of Parameters**

After aggregating data by relevant covariates, we assume there are *G* subsets and define confirmation with colonoscopy as a binary outcome,, which represents whether the *j*-th screenee of subset *i* had underwent colonoscopy or not so that is the number of screenees had underwent colonoscopy in subset *i* and is the total number of screenee in subset *i* ϵ {*1,…,G*}. is time to undergo colonoscopy of the *j*-th screenee in subset *i* so that represents total waiting time in subset *i*. is the probability of refusing to undergo colonoscopy (non-complier), and is the mean arrival rate of receiving colonoscopy. To easily express the likelihood function for the hurdle model, we divided each subset *i* into two groups by non-compliance or compliance, symbolizing *i1* and *i2* so that *i*\* = {*11,12} …*{*G1,G2}*.

The hurdle is crossed if a count is greater than zero. Therefore, the process generates a binary response (whether the number of screenees undergoing colonoscopy takes on the value zero or a positive value), and the probability mass function (p.m.f) is

(1)

Because is a count data, which follows Poisson distribution, the probability of zero counts is

(2)

and the probability of no zero counts equals . As result, we can obtain the p.m.f of truncated Poisson process (given the count greater than one)

(3)

Thus, the p.m.f of hurdle model can be expressed as

(4)

where .

The general form of the likelihood function for the hurdle model is

(5)

 To identify the effect of relevant covariates (), we model using the logistic regression model

(6)

and using the Poisson regression model

(7)

Thus, the log-likelihood can be written

(8)

**Supplementary Table 1.** Multivariate analysis of hurdle part on factors affecting non-compliance for colonoscopy and non-hurdle part on factors affecting waiting time for undergoing colonoscopy by screening periods

|  |  |  |
| --- | --- | --- |
| Characteristics | Hurdle part | Non-hurdle part |
| Coefficient | Risk Score | Coefficient | Risk Score |
| Gender |  |  |  |  |
| male | -2.3532a | 0 | 0.0076 | 8 |
| female | 0.0283 | 28 | -4.0401 a | 0 |
| Age |  |  |  |  |
| 50-54 | -2.3532 a | 0 | 0.0308 | 31 |
| 55-59 | 0.1023 | 102 | 0.0235 | 24 |
| 60-64 | 0.1290 | 129 | 0.0252 | 25 |
| 65-69 | 0.2309 | 231 | -4.0401 a | 0 |
| Screening round |  |  |  |  |
| Prevalent | 0.4809 | 481 | -4.0401 a | 0 |
| Subsequent | -2.3532 a | 0 | 0.0430 | 43 |
| Level of urbanization |  |  |  |  |
|  Metropolitan | -2.3532 a | 0 | - | - |
|  Sub-metropolitan | 0.0636 | 64 | - | - |
|  Non-metropolitan  | 0.1946 | 195 | - | - |
| Period |  |  |  |  |
| **2004-2009** |  |  |  |  |
| Geographic area |  |  |  |  |
| Northern | -2.3532 a | 0 | 0.0566  | 57 |
| Central | 0.0989 | 99 | 0.1530  | 153 |
| Southern | 0.0862 | 86 | 0.1021  | 102 |
| Eastern/offshore islands | 0.2103 | 210 | -4.2421 b  | 0 |
| Level of urbanization |  |  |  |  |
| Metropolitan | - | - | -4.2421 b  | 56 |
| Sub-metropolitan | - | - | -0.0412  | 15 |
| Non-metropolitan  | - | - | -0.0560  | 0 |
| Type of screening units |  |  |  |  |
| Hospital | 1.1320 | 1132 | 0.2298  | 230 |
| Public health centers | -2.3532 a | 0 | 0.6531  | 653 |
| Local clinic | 0.8772 | 877 | -4.2421 b  | 0 |
| **2010-2013** |  |  |  |  |
| Geographic area |  |  |  |  |
| Northern | -1.2174 b | 0 | 0.1052 | 106 |
| Central | 0.1238  | 124 | 0.1263 | 127 |
| Southern | 0.0137  | 14 | -0.0011 | 0 |
| Eastern/offshore islands | 0.0164  | 16 | -4.0401 a | 1 |
| Level of urbanization |  |  |  |  |
| Metropolitan | - | - | -4.0401 a | 0 |
| Sub-metropolitan | - | - | 0.0964 | 96 |
| Non-metropolitan  | - | - | 0.0333 | 33 |
| Type of screening units |  |  |  |  |
| Hospital | -0.0400  | 0 | 0.1210 | 121 |
| Public health centers | -1.2174b  | 40 | 0.0742 | 74 |
| Local clinic | 0.6256  | 666 | -4.0401 a | 0 |

a represents intercept term.

b represents intercept term plus period effect.

**Supplementary Figure 1.** Study framework of hurdle model for referral to confirmatory diagnosis in Taiwanese nationwide colorectal cancer screening program



**Supplementary Figure 2.** Illustration of two type of zero on the waiting time (WT) for confirmatory examination

