MODELING COUNT DATA FOR HEALTHCARE UTILIZATION: AN EMPIRICAL STUDY OF OUTPATIENT VISITS AMONG VIETNAMESE OLDER PEOPLE

Abstract

Background

Vietnam is undergoing an unprecedented pace of aging process and is expected to experience the fastest aging process in region. Association between increasing age and health deterioration has been well-documented across settings. Consequently, demand for healthcare utilization is rising among older people. However, healthcare utilization, here measured as count data, creates challenges for modeling because such data typically has distributions that are skewed with a large mass at zero. This study compares empirical econometric strategies for the modeling of healthcare utilization (measured as the number of outpatient visits in the last 12 months), and identifies the determinants of healthcare utilization among Vietnamese older people based on the best-fitting model identified.

Methods

Using the Vietnam Household Living Standard Survey in 2006 (N=2426), nine econometric regression models for count data were examined to identify the best-fitting one. We used model selection criteria; statistical tests; and goodness-of-fit for in-sample model selection. In addition, we conducted 10-fold cross-validation checks to examine reliability of in-sample model selection. Finally, we utilized marginal effects to identify the factors associated with number of outpatient visits among Vietnamese older people based on the best-fitting model identified.

Results

We found strong evidence in favor of hurdle negative binomial model 2 (HNB2) for both in-sample selection and 10-fold cross-validation checks. The marginal effect results of the HNB2 showed that predisposing, enabling, need, and lifestyle factors were significantly associated with number of outpatient visits. The predicted probabilities for each count event showed the distinct trends of healthcare utilization among specific groups: at low count events, women and people in younger age group used more healthcare utilization than did men and their counterparts in older age groups, but a reversed trend was found at higher count events.

Conclusions

The findings here suggest that the HNB2 model should be considered for use in modeling counts of healthcare use. This study's findings lay the groundwork for future research on the modeling of healthcare utilization in developing countries and those findings could be used to forecast on healthcare demand and making provisions for healthcare costs.

***Keywords*:** count data, Vietnam, modeling healthcare utilization, older people, outpatient visits, hurdle models, overdispersed data.

INTRODUCTION

Count data (observations that have only nonnegative integer values) frequently arise in healthcare utilization data such as number of outpatient visits to hospitals. Data on healthcare utilization are typically characterized by a substantial point mass at zero, a long right tail of individuals who make heavy use of healthcare, and a tendency for variances to increase with the mean. Consequently, such datasets pose challenges for modeling healthcare utilization. Modeling healthcare utilization has received considerable attention in the field of health economics since an understanding of the factors that drive healthcare utilization is essential for policy-making. In addition, the choice of econometric models has substantial implications for a number of empirical purposes such as predicted probability of use of healthcare services and the likelihood of being extensive users of such services.

The Poisson regression model (PRM), a basic model for count data, is considered here as a starting point of analysis. The Poisson distribution has a special property, equality of the conditional mean and the conditional variance, which is known as equidispersion. However, that property has been evaluated as too restrictive in the literature on modeling healthcare utilization because in real data, count variables often have a variance greater than the mean, a condition known as overdispersion. The negative binomial regression model (NBRM) has a built-in parameter that accounts for the overdispersion problem, so estimates from the NBRM appear to be substantially more efficient than those of the PRM.

Two alternative regression models considered for count data, zero-inflated regression model (ZIM) and hurdle regression model (HRM), which allow zeros and positive observations are generated by two different processes. In particular, the HRM reflects two different decision-making processes: whether to use healthcare or not; and (conditional on the decision of use of

healthcare) how much care to consume. The HRM can be viewed as a principal-agent model, where the principal (the patient) initiates the first visit to a hospital to seek healthcare, but the physician (the agent) and patient jointly decide the second and subsequent visits (1). While the HRM allows for the possibility that zeros are generated by a different process from positive observations, the ZIM, introduced by Mullahy (2) and Lambert (3), allows zeros to be generated by two distinct processes: structural and sampling processes. The strategy behind the ZIM reflects the intuition that there are two latent groups in the population: potential users and nonusers. For example, in the context of outpatient visits to hospitals, it might be reasonable to think that the population comprises two types of groups: individuals who would never seek outpatient services in hospitals and individuals who would. There are two possible scenarios: first, there were individuals in the latter group who did not use outpatient services during the period when the survey was conducted (sampling zeros); and second, an individual observed to have zero outpatient visits either just happened not to seek outpatient services during the period when the survey was conducted (sampling zeros) or would never do so (structural zeros).

Although the HRM and ZIM can be viewed as two-component finite mixture models, such mixture is of a limited form because the zeros are treated in separate processes in those count models. Latent class model (LCM) provides a more general finite mixture model which has powerful properties for the modeling of healthcare utilization. Unlike the HRM and ZIM, the LCM makes no distinction between users and non-users of care; rather in the case of two latent sub-populations, it distinguishes two groups, “healthy” and “ill” (4). The LCM allows for heterogeneity along the outcome distribution by means of complex configurations of either observed or unobserved characteristics. The LCM for unobserved heterogeneity rests on the assumption that the unobserved heterogeneity which divides the population into latent classes is

based on individuals' latent long-term health status. Therefore, population heterogeneity may not be well captured by proxy variables such as self-rated health or chronic health conditions (5).

Researchers are typically interested in the distinction between extensive margins: that is, zero counts versus positive counts (e.g., no outpatient visit versus at least one outpatient visit) and intensive margins: that is, how many positive counts if nonzero counts (e.g., how many subsequent outpatient visits after the first visit is made). To date, most studies on modeling healthcare utilization for count data have been conducted in developed countries and very little is known in the context of developing ones. In developing countries, studies on healthcare expenditure, i.e. catastrophic payments for healthcare or healthcare payments and poverty, have attracted more attention than those of healthcare utilization, measured as count events. To the best of our knowledge, there has been no study on modeling healthcare utilization in developing countries, especially for the case of older people. This study, utilizing currently appropriate econometric practices in modeling count outcomes, contributes to empirical evidence on the best choice of econometric models for count data in developing countries. The aim of this study was: i) to identify the model that best explains variability in number of outpatient visits by comparing empirical econometric strategies for the modeling of healthcare utilization, measured as the number of outpatient visits in the last 12 months; and ii) to determine the determinants of healthcare utilization among Vietnamese older people based on the results of the best-fitting model identified. This study examined the effectiveness of the PRM, the NBRM, the HNB model and its extensions, the ZIM and its extensions, and the LCM.

INSTITUTIONAL BACKGROUND

In line with a rapid demographic transition towards an aging society in the world, Vietnam is undergoing an unprecedented pace of aging process and is expected to experience the fastest aging process in region (6). Impressive economic growth since *Doi Moi* (economic renovation) in the late 1980s has resulted in considerable improvement in socio-economic status and healthcare system in Vietnam. Life expectancy at age 60 of Vietnamese older people is relatively high, with an expectation of, on average, 25 more years for women and 19 more years for men (7). However, those number of expected years to live consist of an average of seven years living with illness/disability for women and the corresponding number for men is five years (7), leading to the rise in demand for healthcare utilization among older people. Particularly, healthcare utilization among Vietnamese older people was mostly outpatient visits (91.0%, governmental, private, and other health institutions combined) (7). Social health insurance (SHI) in Vietnam was introduced in 1992, followed by a series of reforms to provide access to healthcare services and reduce out-of-pocket (OOP) spending on healthcare due to fee-for-service. As a result, SHI coverage among older people was significantly increased from 43.5% in 2006 to 75.0% in 2014 and OOP was reduced over time (7).

METHODOLOGY

Data

The data used in this paper was drawn from the 2006 VHLSS, conducted by the Vietnam General Statistics Office. That survey, similar to the Living Standard Measurement Study, is one of the most commonly used household surveys in developing countries. VHLSS is conducted every two years; the information collected is used for assessment of the living standards of

populations in all regions and localities across the country. The survey gathers data on a variety of topics: demographic characteristics of household members; household income; household expenditure; education; health; employment; assets; housing; facilities; and participation in hunger elimination and poverty reduction. The 2006 VHLSS sampling framework was based on that of the Viet Nam's 1999 Housing Population Census. A three-stage stratified design method was adopted for the survey sampling. To date, of all waves of the VHLSS conducted, the 2006 VHLSS contains the richest information on the health conditions, i.e. disability and non-communicable diseases (NCDs) and lifestyle behavior, i.e. smoking of household members. That marks an advantage of the 2006 VHLSS over other waves. The final sample size of the 2006 survey was 45945 households, including an income survey of 36756 households and expenditure survey of 9189 households. At household level, the survey collected information on household income, household expenditure, household size, and so on. At the individual level, various information on individual characteristics was collected: age; gender; ethnicity; education; marital status, working status, health conditions. In this study, older people (defined as those aged 60 and older) were of interest, so we restricted our analysis to a sample of 2624 people.

Measurement of variables

Count outcome variable

The count outcome variable is the number of outpatient visits in the last 12 months. Summary statistics and frequency distributions of that variable were reported in Table 1 and Figure 1, respectively.

Explanatory variables

Empirical studies on healthcare utilization informed the selection of explanatory variables in this study. Particularly, we included four types of explanatory variables: predisposing; enabling; need; and lifestyle variables. Predisposing variables reflect demographic characteristics of respondents: age; age square; sex; marital status; and ethnicity. Enabling variables refer to differences in access to healthcare: log of household size; place of residence; region of residence; education; employment status; log of household income; SHI; and health subsidy. Need variables capture the need for healthcare: disability and NCDs. Finally, lifestyle variable included smoking. Definitions of the explanatory variables were presented in Table 1.

Regression models for count data

Poisson regression model (PRM)

The PRM assumes that a discrete random dependent variable, y_i, following a Poisson distribution, indicates the actual number of times that an event occurs, with mean μ_i indicating the expected number of times that the event will occur during a given period of time. The PRM model is defined by the density

$$f(y_i|\boldsymbol{x}_i) = \frac{e^{-\mu}\mu_i^{y_i}}{y_i!} \qquad for\ y_i = 0, 1, 2, \ldots \qquad (1)$$

where the conditional mean is defined as

$$\mu_i = E[y_i|\boldsymbol{x}_i] = \exp(\boldsymbol{x}_i'\boldsymbol{\beta})$$

and $\boldsymbol{x}_i$ is the vector of covariates, $\boldsymbol{\beta}$ is a (k x 1) parameter vector of unknown coefficients, and $y_i!$ is the factorial operator (5).

Negative binomial regression model (NBRM)

Following Cameron and Trivedi (5), the NB density for a random discrete count outcome $y = 0, 1, 2, \ldots$ can be written as

$$f(y_i|\, u_i;\ \alpha) = \frac{\Gamma(\alpha^{-1}+y_i)}{\Gamma(\alpha^{-1})\Gamma(y_i+1)}\left(\frac{\alpha^{-1}}{\alpha^{-1}+u_i}\right)^{\alpha^{-1}}\left(\frac{u_i}{\alpha^{-1}+u_i}\right)^{y_i} \quad for\ \alpha > 0 \quad (2)$$

where $\Gamma(.)$ denotes the gamma function and α is a constant dispersion parameter to be estimated. The first two conditional moments of the NBRM are

$$E[y_i|u_i;\ \alpha] = u_i = \exp(\boldsymbol{x}_i'\boldsymbol{\beta})$$

$$V[y_i|u_i;\ \alpha] = u_i + \alpha {u_i}^2.$$

The above specification corresponds to the most commonly used version of the NBRM; the negative binomial 2 (NB2). A less used version of the NBRM, known as the negative binomial 1 (NB1), specifies that the conditional variance is linear in the mean, while the conditional mean specified is similar to that in the NB2.

$$V[y_i|u_i;\ \alpha] = u_i + \alpha u_i.$$

Hurdle regression model (HRM)

In the HRM for count data, proposed by Mullahy (8), the two parts can be estimated separately using two different densities: $f_1(.)$ and $f_2(.)$. More specifically, the zero part is determined by $f_1(.)$, such that $\Pr(y_i = 0) = f_1(0)$. The positive count part, determining the amount of use of healthcare, is specified by $f_2(.)$, such that the probability of observing *y*, conditional on *y*>0, is $f_2(y_i|y_i > 0) = f_2(y_i)/\{1 - f_2(0)\}$. In practice, the most common choice for $f_1(.)$ is logit model, which is used here. The typical choice for $f_2(.)$ is usually either a truncated-at-zero Poisson or negative binomial (NB). The probability function of the HRM can be written as

$$\Pr(y_i = j_i|\boldsymbol{x}_i) = \begin{cases} f_1(0|\boldsymbol{x}_i) & if\ j_i = 0 \\ \frac{1-f_1(0|\boldsymbol{x}_i)}{1-f_2(0|\boldsymbol{x}_i)} f_2(j_i|\boldsymbol{x}_i) & if\ j_i > 0 \end{cases} \quad (3)$$

where $f_1(0|\boldsymbol{x}_i) = \frac{\exp(\boldsymbol{x}_i\boldsymbol{\beta})}{1+\exp(\boldsymbol{x}_i\boldsymbol{\beta})}$. As for the hurdle Poisson model (HPM) specification, $f_2(0|\boldsymbol{x}_i) = \exp(-\mu_i) = \exp(-\boldsymbol{x}_i'\boldsymbol{\beta})$, and $f_2(j_i|\boldsymbol{x}_i)$ is specified as the standard PRM described in equation (1). In the case of the hurdle negative binomial model (HNB), $f_2(0|\boldsymbol{x}_i) = (1 + \alpha\mu_i)^{-1/\alpha}$, and $f_2(j_i|\boldsymbol{x}_i)$ corresponds to the NBRM described in equation (2). The conditional mean of equation (3) is given by

$$E[y_i|\boldsymbol{x}_i] = \frac{1 - f_1(0|\boldsymbol{x}_i)}{1 - f_2(0|\boldsymbol{x}_i)}\mu_2$$

where μ_2 is the conditional mean of the second part.

Zero-inflated regression models (ZIM)

If the probability of being potential nonusers is *q*, then (1-*q*) is the probability of being potential users. The probability function of the ZIM can be defined as (9)

$$\Pr[y_i = j_i|\boldsymbol{x}_i] = \begin{cases} q + (1-q)f_2(0|\boldsymbol{x}_i) & if\ j_i = 0 \\ (1-q)f_2(j_i|\boldsymbol{x}_i) & if\ j_i > 0 \end{cases} \quad (4)$$

where $f_2(.)$ is the density of either the PRM or NBRM. In equation (4), positive counts arise only from the process that generates users, while zeros arise from both processes: generating both users and nonusers. For zero-inflated Poisson (ZIP) specification, $f_2(0) = \exp(-\mu_i) = \exp(-\boldsymbol{x}_i'\boldsymbol{\beta})$, and $f_2(j_i|\boldsymbol{x}_i)$ corresponds to the standard PRM described in equation (1). In the case of zero-inflated negative binomial 2 (ZINB2) specification, $f_2(0) = (1 + \alpha\mu)^{-1/\alpha}$, and $f_2(j_i|\boldsymbol{x}_i)$ corresponds to the NBRM described in equation (2).

Latent Class Models (LCM)

A random outcome variable, *y*, is drawn from one of *C* distributions, with probability π_c of being drawn from that distribution, such that $0 < \pi_c < 1$ and $\sum_{c=1}^{C} \pi_c = 1$. Then, the density function for a *C*-component finite mixture is defined as:

$$f(y|\boldsymbol{x};\ \theta_1, \theta_2, \dots, \theta_C;\ \pi_1, \pi_2, \dots, \pi_C) = \sum_{c=1}^{C} \pi_c f_c(y|\boldsymbol{x};\ \theta_c) \quad (5)$$

where $f_c(y|\boldsymbol{x};\ \theta_c)$ are the density for class or component c (c = 1, 2, …, C); and θ_c are the parameters of the distributions $f_c(.)$ (10–12).

Common choices for distributions of count data are the NB, which is used here. The latent class NB2 (LCNB2) model assumes that each of the component distributions follows a NB2 model, with mean $\mu_{c,i}$ and overdispersion α_c. For an individual in class c, the LCNB2 with gamma density for an outcome *y* can be expressed as a density function, as follows:

$$f_c(y_i|\boldsymbol{x}_i;\ \theta_c) = \frac{\Gamma(\alpha_c{}^{-1}+y_i)}{\Gamma(\alpha_c{}^{-1})\Gamma(y_i+1)} \left(\frac{\alpha_c{}^{-1}}{\alpha_c{}^{-1}+u_{c,i}}\right)^{\alpha_c{}^{-1}} \left(\frac{u_{c,i}}{\alpha_c{}^{-1}+u_{c,i}}\right)^{y_i} \quad (6)$$

where $\theta_c = (\alpha_c, \beta_c)$, $\mu_{c,i} = \exp(\boldsymbol{x}_i'\boldsymbol{\beta_c})$ (9). In this model, (α_c, β_c) are unrestricted across latent classes. The expected value of the outcome, y_i, given covariates $\boldsymbol{x}_i$, is

$$E(y_i|\boldsymbol{x}_i) = \sum_{c}^{C} \pi_c \mu_{c,i}.$$

Model specification tests

Deb et al. (10) show that choosing wrong model specifications might lead to inconsistent estimates of parameters and misleading results. We used Ramsey's RESET test to examine specification of the explanatory variables in the context of the PRM. In brief, the logic of that test is to regress the dependent variable (here, number of outpatient visits) on its predicted values and the powers of its predicted values. Such a test provides information on whether important variables correlated with high-order terms are omitted or not (10). More specifically, Ramsey's RESET test assesses whether the coefficients on the squared, cubed, and fourth-order terms are

jointly significant different from zero. Detailed information on Ramsey's RESET test has been presented elsewhere (10,13). The Ramsey's RESET test results show that the PRM used in this study were correctly specified, since the test was statistically insignificant at a *p*-value of 5% (not shown here).

In-sample model selection

In general, two common approaches are used to evaluate performance of count models: (i) compare mean predicted probabilities and observed proportions for each count of outpatient visits; and (ii) use statistical tests and goodness-of-fit measures for performance evaluation among the count models considered. Adopting the former approach, we computed average predicted probabilities for counts 0-20, since those count events accommodated most observations of outpatient visits. Then, we compared those average predicted probabilities estimated with the corresponding observed proportions of assigned counts in each count model considered. Adopting the latter approach, we used likelihood ratio (LR) and Vuong tests for model discrimination among nested and non-nested models, respectively. Basically, an LR test uses −2 times the difference in the fitted log-likelihoods of the two nested models. Among the selected count models, the NB1 nests the HNB1, the NB2 nests the HNB2, the PRM nests the HPM, the PRM nests the NBRM, and the ZIP nests the ZINB2. In addition, the Vuong (14) test was performed to evaluate efficiency among non-nested models. The Vuong test can be computed as

$$V = \frac{\bar{m}\sqrt{N}}{s_m} \qquad (7)$$

where $m_i = \ln\left\{\frac{\widehat{Pr_1}(y_i|x_i)}{\widehat{Pr_2}(y_i|x_i}\right\}$; $\sqrt{N}$ is the square root of the sample size; $\bar{m}$ and s_m are the mean and standard deviation of m_i, respectively; $\widehat{Pr_1}(y_i|x_i)$ and $\widehat{Pr_2}(y_i|x_i)$ are the predicted probability of

observing y_i in the first and the second models, respectively. The Vuong test asymptotically follows a normal distribution, so the first model is favored if V is greater than 1.96 and the second model is favored if V is smaller than -1.96 (14). Wilson (15) has shown that using the Vuong test to examine performance of the ZIM is invalid, therefore we simply used model selection criteria and goodness-of-fit measures for the ZIM.

Regarding model diagnostics, we computed two commonly used model selection criteria: the Akaike information criteria (AIC) (16) and the Bayesian information criteria (BIC) (17), for comparison among the selected count models. Those two criteria, found to be robust to model misspecification (18), can be computed as

$$AIC = -2\ln(L) + 2k \tag{8}$$

$$BIC = -2\ln(L) + \ln(N)\,k \tag{9}$$

where $\ln(L)$ is the maximized log likelihood; k is the number of parameters in the model. Smaller values in both AIC and BIC are preferable.

We also computed measures of goodness-of-fit, measured as root mean square error (RMSE) and mean absolute prediction error (MAPE), to evaluate whether the preferred model provides a good fit to the data. Those two measures of goodness-of-fit capture bias between predicted probabilities and observed proportions for each count of the count models considered here, thus the smaller the bias, the better the model is. RMSE and MAPE can be computed as

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{N}} \tag{10}$$

$$MAPE = \frac{\sum_{i=1}^{N}|(y_i - \hat{y}_i)|}{N} \tag{11}$$

where $\hat{y}_i$ is the predicted probability for each count.

K-fold cross-validation

A potential drawback of heavily parameterized models, especially where those models have been examined by a given dataset, is that they might be overfitting a particular sample of data and performing poorly in terms of out-of-sample forecasts. This implies that in-sample model performance may not always be reliable. Alternatively, *K*-fold cross-validation checks provide a useful guide to out-of-sample testing (19,20). In *K*-fold cross-validation, the original dataset is randomly divided into *K* sub-datasets of approximately equal size. Among the *K* sub-datasets, a single sub-dataset is taken as a validation dataset for model testing, and the remaining $K - 1$ sub-datasets are used as training. It is important to note that each observation in the original dataset is randomly assigned to a single sub-dataset and stays in that sub-dataset during the cross-validation examination. The cross-validation procedure is that *K*-1 models are first trained on the training datasets, and then the estimates of those models are evaluated on the validation dataset. The cross-validation process is repeated *K* times (the folds), with each of the *K* sub-datasets used exactly once as the validation data. This means that each sub-dataset has a chance to be used one time in the validation part and used to train models *K*-1 times. In practice, there is no formal rule for choosing values of *K*; the choice of *K* is usually 5 or 10. In this exercise, we used 10-fold cross-validation because its common use in practice.

RESULTS

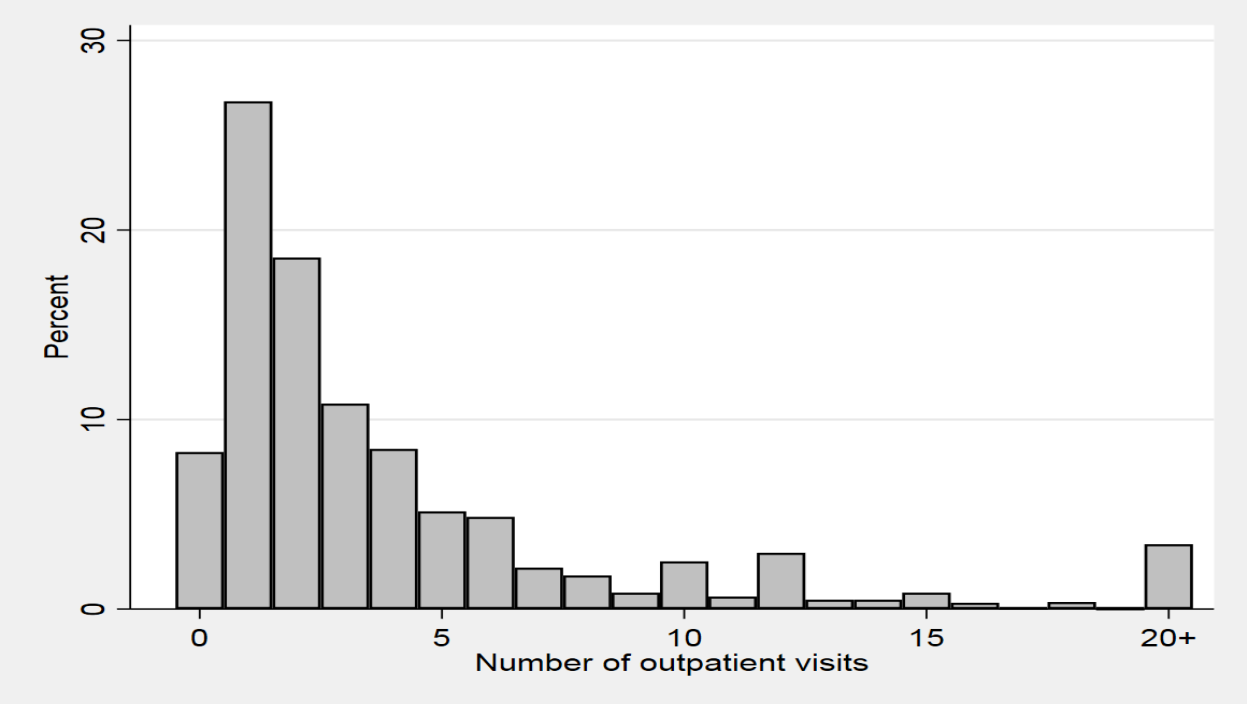
Descriptive results

Table 1 presents definitions and summary statistics for the dependent and explanatory variables. Vietnamese older people used outpatient services an average of four times in 2006. Variance of outpatient visits is $6.437^2 = 41.441$, roughly 10 times the mean of 4.336, suggesting that the data is very highly overdispersed.

[Table 1 about here]

Figure 1 presents the frequency distributions of outpatient visits truncated at 20 visits, implying that there are some excess zeros. It can be seen that the 2006 VHLSS had probability mass concentrated on a few values and was severely skewed to the right tail. Particularly, the proportion of 0 to 20 visits accounted for about 97.0% of outpatient visits. Among those visits, the most concentrated values were in the range of 1 to 6 (about 83.0%, taken together), while the zero counts accounted for only about 8.0% of the visits. The right tail of distributions of outpatient visit was very long, with a maximum value of 104.

Figure 1 The frequency distribution of outpatient visits



In-sample model selection

Figure 2 presents histograms of mean predicted probabilities and observed proportions for 0 to 20 counts of outpatient visits amongst the selected count models. The blue bars depict the actual count frequencies in each count cell, while the orange bars depict the mean predicted probabilities. The figure highlights the extent to which the probability of each count is over- or under-predicted, especially for the zeros. In this study, we used only NB2 density for the LCM analysis. Also, we did not consider either NB1 density or three-component finite mixture models because they had convergence problems, though we attempted to use built-in options in Stata (e.g., *difficult* option). In addition, the results of the ZINB2 model should be taken with caution because the number of iterations was limited to 30 due to convergence problems. It can be seen that the HNB1, HNB2, and HPR models produced exactly the same mean predicted probabilities at zero counts as those of the actual frequencies, while other count models over-predicted probabilities of the zero counts, except for the PRM, which was under-predicted. Regarding other count events, the PRM and its hurdle showed a worse fit, while the NBRM and its hurdles showed great improvement in fit relative to the PRM and its hurdle. The LCNB2 and the ZINB2 models also appeared to be a better fit than the PRM and its hurdle. Overall, it appears that the HNB1 and HNB2 were the preferred models. However, such visualization simply gives us an overall picture of the selected model performance at each count event, thus there is further discussion on model selection below, using the information criteria and statistical tests.

Figure 2 Mean predicted probabilities and observed proportions for each count of outpatient visits among count models

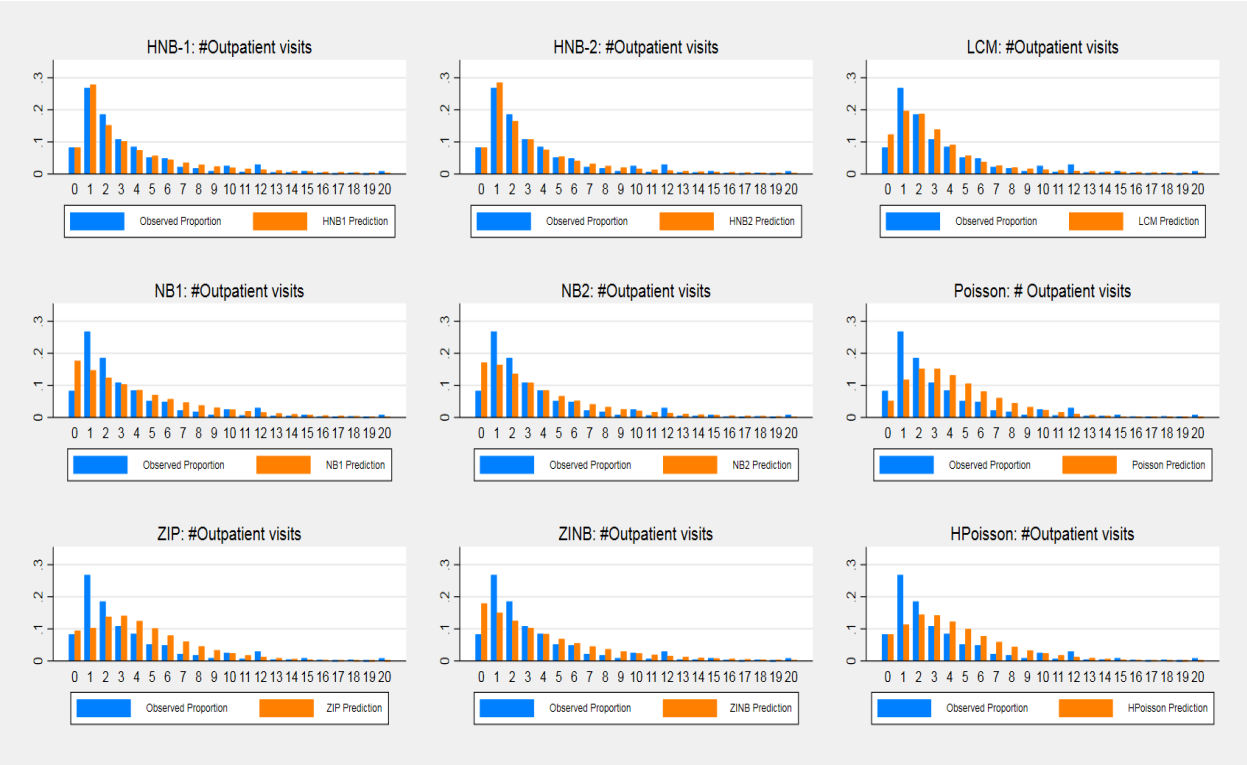


Table 2 summarizes the results of log likelihood, information criteria, and goodness-of-fit measures of each count model considered. It can be seen that the PRM produced the smallest values in log likelihood (log(L)=8550), making it the worst model. Among the count models considered, the HNB2 model provided the largest values in log likelihood, suggesting that the HNB2 is the preferred model. The results of information criteria show strong evidence in favor of the HNB2 model because its values in the AIC and BIC were the smallest among the compared models, followed by those in the HNB1 and LCNB2 models, respectively. The results of goodness-of-fit also favored the HNB2 model, since that model produced the smallest bias between the predicted probabilities and the observed proportions among the count models considered.

Table 2 Results of log likelihood, information criteria, and goodness-of-fit measures

Model	*K*	Log (*L*)	AIC	BIC	RMSE	MAPE
PRM	34	-8549.928	17167.856	17364.852	1.856	0.531
NB1	35	-6085.813	12241.626	12444.416	1.680	0.433
NB2	35	-5942.007	11954.016	12156.806	1.469	0.372
HNB1	69	-5751.954	11641.908	12041.694	0.491	0.184
HNB2	69	-5698.735[a]	11535.471[a]	11935.257[a]	0.405[a]	0.160[a]
HPM	68	-8314.001	16764.001	17157.993	1.818	0.485
ZIP	68	-8332.979	16801.958	17195.95	1.931	0.518
ZINB2	69	-6218.109	12574.219	12974.005	1.667	0.426
LCNB2	72	-5769.265	11682.53	12099.698	0.905	0.253

Note: K was the number of parameters estimated for each model; Log(L) denotes log likelihood; [a] indicates the preferred model; and RMSE and MAPE denote root mean square error and mean absolute prediction error, respectively.

Table 3 presents the results of LR tests of the NBRM versus the HNB models, the ZINB2 model versus the ZIP model, and the HPR model versus the PRM. The NBRM was rejected in favor of the HNB models for both NB1 and NB2 models. The ZIP model and the PRM were rejected in favor of the ZINB2 and HPR models, respectively.

Table 3 Results of the LR tests among nested count models

Pair model	Differences in LR	1% critical value
HNB1[b] vs. NB1	667.718***	$\chi^2(34)=56.1$
HNB2[b] vs. NB2	486.544***	$\chi^2(34)=56.1$
ZINB2[b] vs ZIP	4229.739***	$\chi^2(1)=6.6$
HPR[b] vs. PRM	471.855***	$\chi^2(34)=56.1$

*Note: *p<0.05, **p<0.01, and *** p<0.001. LR denotes likelihood ratio; [b] indicates the preferred model in pair comparison; and χ^2(.) means chi-square test and the number in the bracket refers to degree of freedom of each model considered.*

Table 4 summarizes the results of Vuong tests for non-nested models. There were 10 pairs of non-nested models, however we were particularly interested in comparisons of the HNB2 and other models because the HNB2 model appears preferable to other count models considered here, based on the results of previous information criteria and LR tests. The Vuong test results show that the HNB2 model was favored because the test statistic for the HNB2 model against the HNB1 model was 3.6, against 12.9 for the HPM, and against 4.2 for the LCNB2 model. Furthermore, those test statistics exceeded the critical value of 1.96, suggesting that the HNB2 model is a better fit than the compared models.

Table 4 Results of the Vuong tests among non-nested count models

Pair model	Vuong tests
HNB2[c] vs HNB1	3.584
HNB2[c] vs. HPM	12.883
HNB2[c] vs LCNB2	4.233
HNB1[c] vs. HPM	12.848
HNB1[d] vs. LCNB2	0.807
LCNB2[c] vs. HPM	12.428
LCNB2[c] vs. NB1	10.043
LCNB2[c] vs. NB2	9.239
NB1[c] vs. HPM	11.682
NB2[c] vs. HPM	12.147

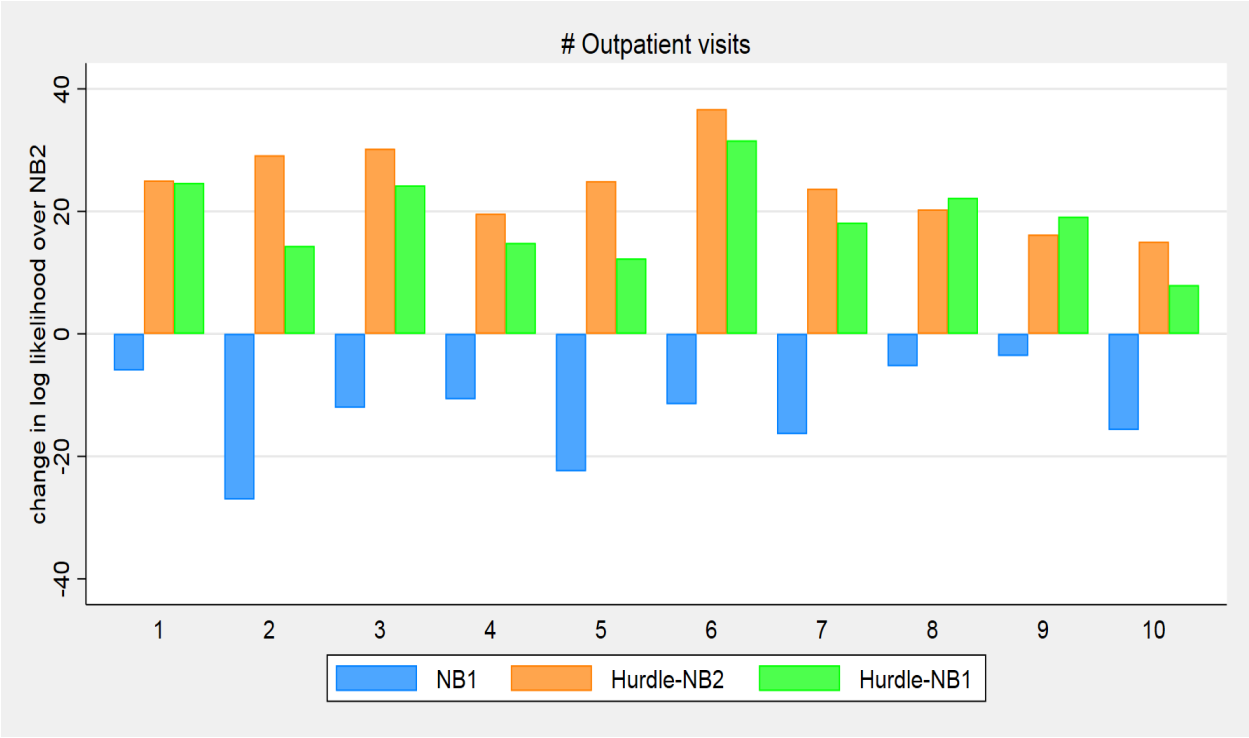
Note: [c] indicates the preferred models based on the Vuong test results; and [d] denotes no evidence of one is superior than the other.

The 10-fold cross-validation

The PRM and its hurdle and the ZIM fit considerably worse than the NBRM and its hurdles. For easy of interpretation, therefore, we did not report them. Figure 3 shows a comparison of the NBRM and its hurdles. In this exercise, the NB2 was used as the base model. The vertical bars depict the difference in log likelihood of the validation sub-dataset with respect to the NB2 model, while the horizontal bars depict the 10 replications of the selected models. Because the 10-fold cross-validation used log likelihood to compare models in each replication,

models with the highest log likelihood relative to the NB2 model were preferred. It can be seen that the NB1 model was the worst performer in each replication, however, its hurdle performed better than the NB2 model. Among the 10 replications, the HNB2 model's performance outnumbered the compared models (eight out of 10 replications).

Figure 3 Results of 10-fold cross validation among the NB1, NB2, HNB1, and HNB2



Marginal effects of the best-fitting model

The results of in-sample model selection and 10-fold cross validation show the HNB2 to be the best-fitting model. In this section, we computed marginal effects of the HNB2 model to determine the effect of each explanatory variable on outpatient visits. Marginal effects from the HRM, as a whole, require putting the part estimating zero counts and the part estimating positive counts together. More specifically, the unconditional rate is computed by combining the mean rate for those with zero counts and the mean rate for those with positive counts. By the

unconditional rate, we meant that both zero and positive counts were estimated, conditional on the explanatory variables.

$$E(y_i|\boldsymbol{x}_i) = (1 - \pi_i) * E(y_i|y_i > 0, \boldsymbol{x}_i) \quad (12)$$

where $E(y_i|y_i > 0, \boldsymbol{x}_i) = \frac{\mu_i}{1-(1+\alpha\mu_i)^{-1/\alpha}}$; and $\Pr(y_i = 0|\boldsymbol{x}_i) = \pi_i$. We used the *suest* and the *expression ()* option in *margins* command to obtain overall marginal effects from equation (12). The *suest* command provides correct standard errors for the HRM model, since that command takes into account the fact that although the two parts are independently estimated, they are dependent. The results of marginal effects of the HNB2 model are summarized in Table 5.

The results show a significantly positive effect of ethnicity, SHI, non-communicable diseases, and disability, and a significantly negative effect of log household size and smoking on the probability of visiting hospitals for outpatient services. The results of region variable were mixed. Particularly, the sample average incremental effect of being Kinh people was 1.09: Kinh people averaged 1.09 more outpatient visits than non-Kinh people, with other variables held constant. Similarly, those with SHI, on average, had 0.58 more outpatient visits than individuals without SHI. Individuals with either NCDs or disability had 2.17 and 1.16 more outpatient visits than those without NCDs and without disability, respectively. As for negative effect covariates on the probability of using outpatient services, an additional member of household was estimated to decrease number of outpatient visits by 0.8. Smoking people had 0.94 less outpatient visits than did non-smoking people.

[Table 5 about here]

Predicted probabilities of using outpatient visits at specific values

Policy-makers and researchers are typically interested in key variables that have strong impact on the health outcomes, as well as inform policy. Although the findings of this study show a significant effect of SHI on number of outpatient visits, we are particularly interested in examining the healthcare utilization trend among specific groups. It is possible that healthcare needs could be varied by age and gender. We find predicted probabilities at specific values to be particularly illustrative for interpretation of each count event for specific groups. In this regard, we examined the predicted probabilities of using outpatient visits for two main hypothetical groups: with and without SHI. In each group, we had six age-gender sub-groups: men aged 60-69; men aged 70-79; men aged 80+; women aged 60-69; women aged 70-79; and women aged 80+. After fitting the HNB2, we computed predicted probabilities for each count of each hypothetical group selected. In this exercise, we presented only the predicted probabilities at count 0-10, since estimates of those count events sufficiently showed the healthcare utilization trend among the selected hypothetical groups. As such, summing up the predicted probabilities of those counts may not be 1.

The results of the predicted probabilities are presented in Table 6 and visualized in Figure 4. Overall, it can be seen that the predicted probabilities of healthcare utilization decreased when number of outpatient visits increased. It is reasonable because individuals' health could be managed or under controlled after the first or second visit to doctors, except severe health conditions, so the probabilities of subsequent visits for them could be diminished. Readers may find that it would be easier to interpret the results for groups with and without SHI if number of outpatient visits are divided into two parts: count 1-4 and count 5-10. The reason is that each part showed the distinct trends of healthcare utilization among the two groups. As for group without SHI at count 1-4, women used more healthcare utilization than did men across age groups, and

people in younger age groups had higher predicted probabilities of using healthcare utilization than their counterparts in older age groups, regardless of their gender. By contrast, the results at count 5-10 show a totally reversed trend as compared to those of count 1-4.

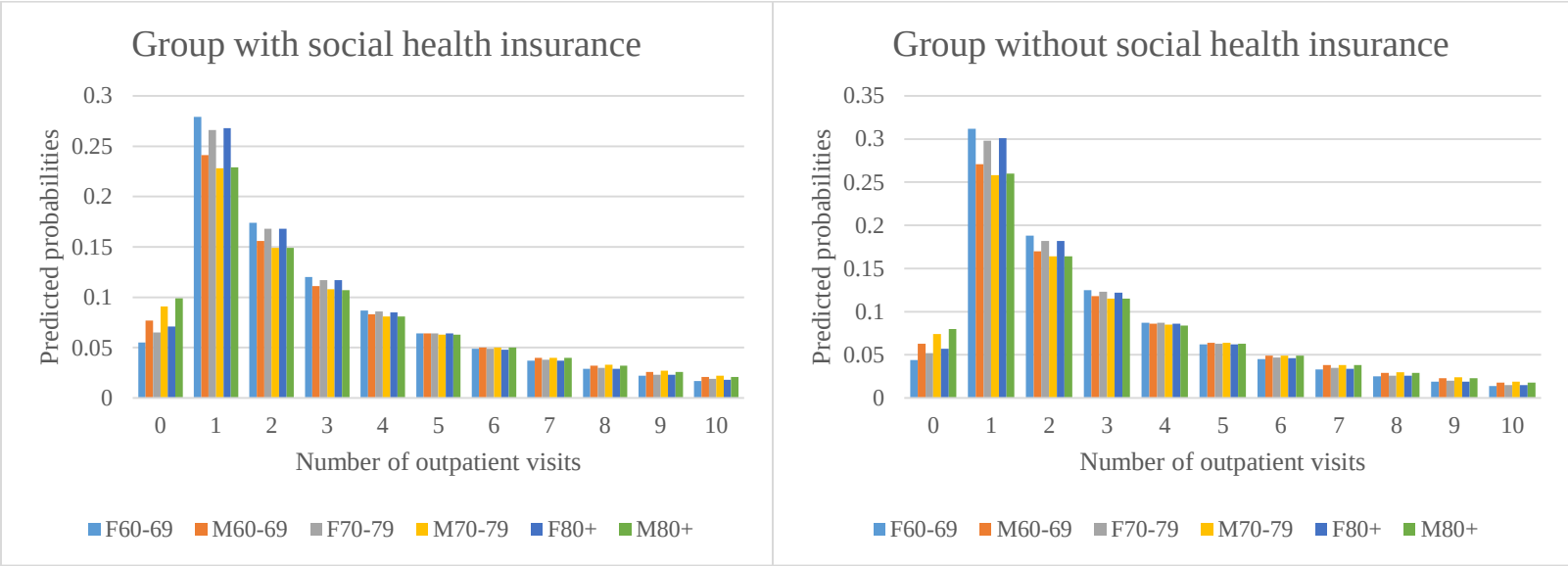
Table 6 Results of the predicted probabilities at specific count events among groups

Groups		Number of outpatient visits										
		0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
	F60-69	0.044	0.312	0.188	0.125	0.087	0.062	0.045	0.033	0.025	0.019	0.014
	M60-69	0.063	0.271	0.17	0.118	0.086	0.064	0.049	0.038	0.029	0.023	0.018
Without	F70-79	0.052	0.298	0.182	0.123	0.087	0.063	0.047	0.035	0.026	0.02	0.015
SHI	M70-79	0.074	0.258	0.164	0.115	0.085	0.064	0.049	0.038	0.03	0.024	0.019
	F80+	0.057	0.301	0.182	0.122	0.086	0.062	0.046	0.034	0.026	0.019	0.015
	M80+	0.08	0.26	0.164	0.115	0.084	0.063	0.049	0.038	0.029	0.023	0.018
	F60-69	0.055	0.279	0.174	0.12	0.087	0.064	0.049	0.037	0.029	0.022	0.017
	M60-69	0.077	0.241	0.156	0.111	0.083	0.064	0.05	0.04	0.032	0.026	0.021
With	F70-79	0.065	0.266	0.168	0.117	0.086	0.064	0.049	0.038	0.03	0.023	0.019
SHI	M70-79	0.091	0.228	0.149	0.108	0.081	0.063	0.05	0.04	0.033	0.027	0.022
	F80+	0.071	0.268	0.168	0.117	0.085	0.064	0.048	0.037	0.029	0.023	0.018
	M80+	0.099	0.229	0.149	0.107	0.081	0.063	0.05	0.04	0.032	0.026	0.021

Note: SHI means social health insurance; F and M denote female and male, respectively.

The results for group with SHI revealed the same pattern of using healthcare utilization as those of group without SHI. A possible explanation for such findings on gender is that although women, on average, live longer than do men, they tend to have poor health than men. By contrast, men tend to have chronic diseases that are associated with higher rates of mortality than do women. Thus, at low count events (e.g., count 1-4), women may use more healthcare utilization than do men, but men may need more healthcare at higher count events (e.g., count 5-10) due to the severity of their health conditions. As for comparison of groups with and without SHI, it can be seen that at count 1-4, people without SHI used more healthcare utilization than did those with SHI. However, the results show a reversed trend at count 5-10. This finding could be partly explained by medical costs for outpatient services, i.e. high medical costs could present a huge barrier for people without SHI to access to healthcare, especially at high count events.

Figure 4 Visualizations of the predicted probabilities at specific count events among groups



Note: F denotes females and M means males

DISCUSSION AND CONCLUSIONS

Results of this study show strong evidence of overdispersion, but no substantial excess of zeros in the outpatient visits data. Consequently, the PRM performed the worst among the count models considered. In addition, extensions of the PRM also showed poor fit. It has been shown that ignoring the overdispersion issue leads to deflated standard errors and inflated *z*-values, though estimates of parameters from the PRM are still consistent even when the equidispersion property is violated (5,9,10). Although the ZINB2 model had a convergence problem, its results showed a better fit than those of the PRM and its extensions. In this study, we used the same set of explanatory variables to model both structural and sampling zeros, which could have been a reason for problems in obtaining convergence in the ZINB2 model (21,22). The results of in-sample selection show that the NBRM fit the data better than did the ZIM. Among the NBRM, the NB2 model was preferred over its NB1 counterpart.

An assumption of the NBRM that the process of generating zeros is the same as that of positive observations, has been criticized as too restrictive in the modeling of healthcare utilization (1). Those critics argue that the decision to initiate the first contact to a doctor and the subsequent visits may be made in the context of two different processes. As such, the LCNB2 undoubtedly fit the data better than did the NBRM. However, the in-sample selection results show that the HNB beat all other count models considered. Among the HNB, the HNB2 fit the data better than did the HNB1.

The results of 10-fold cross-validation reconfirmed that, the HNB2, in most replications, was the best-fitted model. Data come in all shapes and sizes, and comparison among regression models for healthcare utilization and other fields has been widely conducted, though the results are mixed in the literature. Particularly, Deb and Norton (23) find that the HNB model is more

appropriate than the PRM and the NBRM for estimation of office-based visits, while the NBRM best fits the data for emergency department visits. Regarding comparisons of the HRM and LCM, Jiménez-Martín et al. (24) find that the LCM is preferred in the case of general practitioners, while the TPM fits the data better than the LCM when count outcome is number of visits to specialists. Cameron and Trivedi (5), using a recreational trips data set for comparison of the HNB and LCM, find that the HNB fits that data better than does the LCM. In line with Cameron and Trivedi's work, Winkelmann (25), who uses Poisson-log-normal in the second part of the HRM, shows that the HNB describes number of doctor visits better than does the LCM. Findings of this study are in agreement with the findings of those studies, but in contrast with those of studies by Deb and Trivedi (4, 26) and Sarma and Simpson (27), which find strong evidence in favor of the LCM relative to the HNB.

The results of marginal effects show that Kinh people had more outpatient visits than non-Kinh people. A possible explanation could be barriers of access to healthcare among minority people due to long commutes to health institutions (28). People from larger families had fewer outpatient visits than those from smaller families. This finding is inconsistent with that of Deb and Trivedi (4). There is no clear explanation for such finding, although we speculate that it may indicate unobserved financial stress (e.g., people from larger families may choose not to go to a hospital because of lack of finances, even though a certain level of healthcare is needed). The results for region of residence show a mix finding, with both positive and negative effect on number of outpatient visits. This result is consistent with previous studies, indicating that healthcare utilization varies by sub-region (4, 27). Having SHI had significant effects on number of outpatient visits, implying that having SHI leads to an increase in demand for outpatient visits. A possible explanation could be the *ex-ante* moral hazard in healthcare utilization for non-

hospitalized services, i.e. people with SHI tend to use more outpatient services because they know that insurance companies bear part of the cost for such services. Findings of this study are in agreement with previous studies, showing that having SHI or supplemental health insurance increases individuals' healthcare utilization (5,26,27).

Healthcare utilization is responsive to need factors, measured by NCDs and disability. Particularly, having at least one NCD or disability increased number of outpatient visits, which seems reasonable, since most NCDs or disabilities require care management, rather than hospitalization (with the exception of severe conditions). Similar findings have been found in the literature (1,4,26). As for lifestyle variable, smoking had a negative effect on number of outpatient visits and this finding contrasts with that of Sarma and Simpson (27).

The study has some limitations, primarily arising from the nature of the data. First, this is a cross-sectional study, so it cannot provide any causal analysis between the determinants and healthcare utilization. Second, the cross-sectional study, moreover, only captures whether a person used outpatient services at the time of the survey conducted, thus we don't know how individuals' healthcare utilization behaviors change over times. A longitudinal study is needed to observe such changes in healthcare utilization behaviors. Third, we acknowledge the possibility of recalled bias, since the count outcome used in this study was based on self-reported information. Final, this study used the 2006 VHLSS to demonstrate our empirical econometric strategies, thus the findings of this study may not reflect the current trend of healthcare utilization of Vietnamese older people.

Despite those limitations, this study's findings lay the groundwork for future research on the modeling of healthcare utilization in developing countries, and those findings could be used to forecast on healthcare demand and making provisions for healthcare costs. The factors

associated with number of outpatient visits and the trends of using healthcare utilization among specific groups could be served to inform policy-making and guide public health interventions to mitigate inequity in healthcare utilization for the rapidly aging population. Although older women, on average, tend to have poorer health than men, this study showed that the intensive margins were higher for men than it was for women, suggesting that policy should target not only women, but also men. Improvement in count data in the future is essential to provide an accurate understanding of the associated factor with healthcare utilization. Some other important variables are encouraged to be included in future surveys such as detailed types of SHI or complemental health insurance, waiting time, travel time, the number of the number of visits to a general health professional, the number of visits to a specialist, and the number of nights spent as a hospital patient. Such variables could provide interesting findings and essential parameters related to healthcare utilization.

Declarations

Ethics approval and consent to participate

Ethical approval was provided by the Vietnam General Statistics Office as the VHLSS is conducted every two years.

Consent for publication

Not applicable as all respondent names were anonymized

Availability of data and materials

The data are available from the Vietnam General Statistics Office on reasonable request.

Funding sources

This study does not receive any funding.

Competing interest

The authors declared no potential conflicts of interest in the authorship and/or publication of this manuscript.

Author's contribution

LDD and RLG conceptualized the study design. LDD and UMJ analyzed the data and interpreted the results. LDD, RLG, and UMJ oversaw the research. All authors read and approved the final manuscript.

Acknowledgements

We thank Professor Lawrence Morris Hunter – the editor of Center for Professional Communication, GRIPS for his excellent language editing.

References

1. Pohlmeier W, Ulrich V. An Econometric Model of the Two-Part Decisionmaking Process in the Demand for Health Care. J Hum Resour [Internet]. 1995;30(2):339. Available from: https://www.jstor.org/stable/146123?origin=crossref
2. Mullahy J. Heterogeneity, excess zeros, and the structure of count data models. J Appl Econom. 1997;12(3):337–50.
3. Lambert D. Zero-inflated poisson regression, with an application to defects in manufacturing. Technometrics. 1992;34(1):1–14.
4. Deb P, Trivedi PK. The structure of demand for health care: latent class versus two-part models. J Health Econ [Internet]. 2002 Jul;21(4):601–25. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0167629602000085
5. Cameron AC, Trivedi PK. Regression analysis of count data. 2nd ed. Cambridge: Cambridge University Press; 2013. 566 p.
6. United Nations economic and SC for A and the P. Ageing in the Asian and Pacific Region: An overview. Bangkok; 2017.
7. Vietnam Ministry of Health, Health Partnertship Group. Joint annual health review 2016: towards healthy aging in Vietnam. Ha Noi; 2018.
8. Mullahy J. Specification and testing of some modified count data models. J Econom [Internet]. 1986 Dec;33(3):341–65. Available from: https://linkinghub.elsevier.com/retrieve/pii/0304407686900023
9. Jones AM, Rice N, Bago d'Uva T, Balia S. Applied health economics. 2nd ed. London:

Routledge; 2013. 396 p.

10. Deb P, Norton EC, Manning WG. Health econometrics using Stata. College Station, TX: Stata Press; 2017.

11. Lindsay BG. Mixtue model: theory, geometry and application. NSF-CBMS Regional Conference Series in Probability and Statistics, Institute of Mathematical Statistics; 1995.

12. McLachlan GJ, Lee SX, Rathnayake SI. Finite Mixture Models. Annu Rev Stat Its Appl [Internet]. 2019 Mar 7;6(1):355–78. Available from: https://www.annualreviews.org/doi/10.1146/annurev-statistics-031017-100325

13. Ramsey JB. Tests for specification errors in classical linear-least squares regression analysis. J R Stat Soc Ser B. 1969;31(2):350–71.

14. Vuong QH. Likelihood ratio tests for model selection and non-nested hypotheses. Econometrica [Internet]. 1989 Mar;57(2):307. Available from: https://www.jstor.org/stable/1912557?origin=crossref

15. Wilson P. The misuse of the Vuong test for non-nested models to test for zero-inflation. Econ Lett. 2015;127(c):51–3.

16. Akaike H. Statistical predictor identification. Ann Inst Stat Math [Internet]. 1970 Dec;22(1):203–17. Available from: http://link.springer.com/10.1007/BF02506337

17. Schwarz G. Estimating the dimension of a model. Ann Stat. 1978;6(2):461–4.

18. Leroux BG. Consistent estimation of a mixing distribution. Ann Stat. 1992;20(3):1350–60.

19. Arlot S, Celisse A. A survey of cross-validation procedures for model selection. Stat Surv. 2010;4:40–79.

20. Picard RR, Cook RD. Cross-validation of regression models. J Am Stat Assoc [Internet]. 1984 Sep;79(387):575–83. Available from:

http://www.tandfonline.com/doi/abs/10.1080/01621459.1984.10478083

21. Gerdtham UG. Equity in health care utilization: further tests based on hurdle models and Swedish micro data. Health Econ [Internet]. 1997;6(3):303–19. Available from: http://www.ncbi.nlm.nih.gov/pubmed/9226147

22. Grootendorst P V. A comparison of alternative models of prescription drug utilization. Health Econ [Internet]. 1995 May;4(3):183–98. Available from: http://doi.wiley.com/10.1002/hec.4730040304

23. Deb P, Norton EC. Modeling health care expenditures and use. Annu Rev Public Health. 2018;39:489–505.

24. Jiménez-Martín S, Labeaga JM, Martínez-Granado M. Latent class versus two-part models in the demand for physician services across the European Union. Health Econ [Internet]. 2002 Jun;11(4):301–21. Available from: http://doi.wiley.com/10.1002/hec.672

25. Winkelmann R. Health care reform and the number of doctor visits—an econometric analysis. J Appl Econom [Internet]. 2004 Jul;19(4):455–72. Available from: http://doi.wiley.com/10.1002/jae.764

26. Deb P, Trivedi PK. Demand for medical care by the elderly: a finiture mixture approach. J Appl Econom [Internet]. 1997 May;12(3):313–36. Available from: http://doi.wiley.com/10.1002/%28SICI%291099-1255%28199705%2912%3A3%3C313%3A%3AAID-JAE440%3E3.0.CO%3B2-G

27. Sarma S, Simpson W. A microeconometric analysis of Canadian health care utilization. Health Econ [Internet]. 2006 Mar;15(3):219–39. Available from: http://doi.wiley.com/10.1002/hec.1057

28. Nguyen VC. Mapping the reform process in the public delivery of health services in Viet

Nam. Hanoi; 2010.

Table 1 Definition of the selected variables

Variable	Definition	Mean	S.D.
Outpatient visits	Number of outpatient visits in the last 12 months	4.336	6.437
Age	Age in years	71.465	7.977
Sex	Male = 0; female = 1	0.599	0.490
Ethnicity	Non-Kinh people =0; Kinh people = 1	0.894	0.307
Household size	Log of household size	1.251	0.567
Place of residence	Rural = 0; urban = 1	0.262	0.439
Red River Delta	=1, otherwise = 0	0.233	0.423
East Northern Mountainous areas	=1, otherwise = 0	0.098	0.298
West Northern Mountainous areas	=1, otherwise = 0	0.024	0.153
North Central Coast	=1, otherwise = 0	0.098	0.298
South Central Coast	=1, otherwise = 0	0.122	0.328
Central Highlands	=1, otherwise = 0	0.044	0.024
Southeast	=1, otherwise = 0	0.135	0.342
Mekong Delta	=1, otherwise = 0	0.245	0.430
Marital status	Married = 0; single = 1	0.405	0.491
Social health insurance	No health insurance = 0; has health insurance = 1	0.549	0.497

Health subsidy	No health subsidy = 0; received health subsidy = 1	0.587	0.493
Employment status	Not working = 0; working = 1	0.424	0.494
Education	Education of respondents in years	4.075	3.642
Household income	Log of household income	10.017	0.878
Smoking	Not smoking = 0; smoking = 1	0.323	0.468
Non-communicable diseases (NCDs)	No NCD = 0; having at least one NCD = 1	0.356	0.479
Disability	No disability = 0; having at least one disability = 1	0.285	0.452

Note: S.D. denotes standard deviation

Table 5 The results of marginal effects of the HNB2 as a whole

Variables	Coefficient (S.E.)	*P*-value
Age	0.154(0.20)	0.445
Age square	-0.001(0.00)	0.361
Sex (male-reference)		
Female	-0.685(0.42)	0.102
Ethnicity (non-Kinh people-reference)		
Kinh people	1.091(0.385)	0.005
Place of residence (rural areas-reference)		
Urban areas	0.480(0.381)	0.208
Region of Vietnam (Red River Delta-reference)		
East Northern Mountainous areas	-0.692(0.22)	0.002
West Northern Mountainous areas	0.003(0.68)	0.997
North Central Coast	-0.226(0.27)	0.409
South Central Coast	0.322(0.31)	0.303
Central Highlands	1.218(0.59)	0.042
Southeast	3.689(0.64)	0.000
Mekong Delta	4.136(0.49)	0.000
Marital status (married-reference)		
Single	0.182(0.24)	0.443
Log household size	-0.801(0.35)	0.024
Social health insurance (no-reference)		
Yes	0.581(0.28)	0.039

Employment status (no-reference)		
Yes	-0.172(0.26)	0.517
Education	0.004(0.05)	0.933
Log household income	-0.064(0.30)	0.831
Health subsidy (no-reference)		
Yes	0.591(0.36)	0.100
Smoking (no-reference)		
Yes	-0.941(0.37)	0.012
Non-communicable diseases (no-reference)		
Yes	2.167(0.24)	0.000
Disability (no-reference)		
Yes	1.158(0.33)	0.000

Note: S.E. denotes standard errors