

Socio-Economic Predictors of Health Insurance Claims: Evidence from Ghana

Samuel Antwi^{1,2}, Xicang Zhao¹, Eric Kofi Boadi² & Eugene Oware Koranteng²

¹ School of Finance and Economics, Jiangsu University, Jiangsu, China

² Accountancy Department, Koforidua Polytechnic, Koforidua, Ghana

Correspondence: Samuel Antwi, School of Finance and Economics, Jiangsu University, 301 Xuefu Road, Zhenjiang, Jiangsu, China. E-mail: samyantwi@yahoo.com

Received: December 23, 2013

Accepted: January 7, 2014

Online Published: February 25, 2014

doi:10.5539/ijef.v6n3p216

URL: <http://dx.doi.org/10.5539/ijef.v6n3p216>

Abstract

The objective of this study was to determine whether there is an association between socio-economic variables and health insurance claims and if other variables mediate between these relationships. In this work, data was collected from the insured of the Ghana National Health Insurance Scheme with the help of the National Health Insurance database and the patients' attendance register of the Koforidua Regional Hospital, from 1st January to 31st December 2011. The generalized linear regression (GLR) models and the SPSS version 17.0 were used for the analysis. The result shows that socio-economic variables have a statistically significant independent association with health insurance claims, demonstrating in each case that the more socially disadvantaged the greater the health insurance claims. Alcohol drinking and cigarette smoking are mediators of the relationship between each socio-economic measure and health insurance claims. Low income has a significant effect on health insurance claims independent of the other socio-economic variables, but this becomes non-significant after adjusting for alcohol and smoking. It is suggested that the government should consider building more health centers, clinics and cheap-compounds in at least every community, to help reduce the travel time in accessing health care. The ministry of health and the Ghana health service should engage older citizens by encouraging them to use hospitals when they are sick instead of other alternative care providers.

Keywords: national health insurance, claims, logistic regression, socio-economic factors, Ghana

1. Introduction

In most African and South American countries, economic problems have resulted in decreased government funding of the health sector and reduced access to health care for most of the population. For example in Nigeria, the public spending per capita for health is less than \$5 and can be less than \$2 in poorer parts of the country (WHO, 2010). Situations like this have prompted governments of low income countries to explore alternative forms and sources of health sector financing. In many European countries social health insurance is one of the principal methods of health financing and has a long history (Carrin et al., 2005).

"One programme which all the major political parties in Ghana have openly agreed upon, though with different approaches, is the implementation of the National Health Insurance Scheme" (Daily Graphic May 17, 2005). The financial performance of an Insurance Scheme is a function of the premium collected, the cost of health care services of the insured, the level of external subsidy, the size of the pool, and degree of economies of scale that is achieved (Aikins, 2004).

According to World Health Organization in 2010, about 1.3 billion people in the world lack access to effective and affordable health care due to financial limitations or governments inability to provide the necessary coverage. For instance, the United Nation Development Programme (UNDP) report for the year 2007 (UNDP, 2007), estimates that about 70 percent of the population of Ghana use alternative medicine which includes traditional health care while 30 percent use orthodox medical care. The report further posits that in terms of orthodox health care in Ghana, only 18.4 percent of the sick or injured consulted a health practitioner and a sizeable proportion of rural areas and northern Ghana generally are excluded due to the inability to pay. The Millennium Development Goals for poverty reduction and health will not be met without a concerted effort aimed at extending health interventions to the world's poor (Sachs et al., 2005). Against this background, the Ghanaian parliament in 2003

passed the National Health Insurance (NHI) Act, Act 650 promoting Mutual Health Insurance Schemes for the extension of social protection in health to the poor regardless of ability to pay at the point of accessing health services. This culminated in the official launch of the National Health Insurance Scheme (NHIS) in March 2005. The scheme gives prominence to community/district mutual health insurance schemes as a key strategy for the extension of social security in health to every Ghanaian in a bid to enhance access to health care especially for the rural poor and combat social exclusion. There have also been debates about the levels of premium being collected currently by the Scheme. One School of thought have it that if premiums are not increased from the current levels, the scheme may run into severe financial difficulties due to high level of claims.

In this study we looked into some of the problems enumerated above and critically examine some important characteristics of the insured of the national health insurance scheme. We addressed the issues of the views and attitudes of the National Health Insurance Authority (NHIA) regarding estimation of national health insurance claims as well as factors that the authority consider in premium rating. Again, the paper contributed to the existing literature on national health insurance studies in Ghana by providing empirical results pertaining to the significant effect of the insured's risk factors on claim made or otherwise. It is believed that results would have implications for the successful implementation of national health insurance policies in Ghana.

In order to achieve the aim of the study, the binary logit model was employed. The choice of this statistical technique is based on the dichotomous nature of the response variable (whether the insured has made claim or not). The data of the study was drawn from the NHIA data base and the hospital records of the Koforidua regional hospital. The remaining part of the paper is organized as follows: section 2 describes the concept of the methods employed in the research. The data, empirical analysis, and results are presented in section 3; Section 4 provides the concluding remarks as well as recommendations.

2. Methods

2.1 Study Area and Source of Data

The eastern region of Ghana has 21 administrative Municipalities and Districts with Koforidua as the regional capital. Eastern Region of Ghana has an estimated population of 2,194,508, with 3.1% growth rate. It is the sixth largest region with a land area of 19,323 sq. km., thus representing about 8% of the total land area of the country (Statistical Service, 2011). The region is bounded on the East by the Volta Region, South by Greater Accra region, West by Central Region and on the North by Ashanti Region. It has the largest number of health facilities in the country. The Koforidua Regional Hospital is a state run referral hospital. The hospital, with about 250 patient beds is the largest facility in the region which serves both as the first consultation point for patients within its catchment, and as a referral centre for about other 25 primary health centers. These facilities are managed by the Ministry of Health and the Ghana Health Service.

2.2 Data Collection and Data Management

Data used in this study were obtained as primary data from hospital attendance records at the Koforidua Regional Hospital, from 1st January 2011 to 31st December 2011. Data was collected from the policyholders of NHIS in Ghana using the simple random sampling technique with the help of the NHIA database. A total of 4549 policyholders were sampled from the NHIA data base. For this study, the hospital attendance register was used which has patients' age, sex, date of admission and discharge, insurance claim or otherwise, living condition, number of children, level of education, employment status, cigarette smoking, alcohol drinking, billed charges (i.e. for treatment), marital status, health status and whether the patient was an outpatient or inpatient. Income levels were obtained from the health insurance data base. From the data, the following variables were coded:

outcome (claim = 1, no claim = 0); length of stay (inpatient=1, outpatient = 0); marital status (married = 1, unmarried = 0); distance to the hospital ((distance > 5km) = 1, (distance \leqslant 5km) = 0).

The distance of 5 km was chosen to reflect travel time of 1 hour on foot. Age, sex, health status, income level, number of children, level of education, employment, cigarette smoking, alcohol drinking, living condition, marital status, length of stay and billed charges were employed as deciding factors.

2.3 Model Specification, Estimation and Tests

The response variable in logistic regression is usually dichotomous, that is, the response variable can take the value 1 with a probability of success p , or the value 0 with probability of failure, $(1-p)$. To explain the logistic regression, we show here the logistic function $f(z)$, which describes the mathematical form on which the logistic model is based

$$f(z) = \frac{1}{1 + e^{-z}} \quad (1)$$

Where z denotes the values of this function, such that, $-\infty \leq z \leq +\infty$. The relationship between the predictor and response variables is not a linear function in logistic regression; instead, the logistic regression function is used, which is the logit transformation of p. To obtain the logistic model from the logistic function, we write z as the linear sum.

$$z = \alpha + \sum_{i=1}^k \beta_i x_i \quad (2)$$

Where are independent variables of interest and α and are constant terms representing unknown parameters and k is the last term. Combining (1) and (2) gives:

$$f(z) = \frac{1}{1 + e^{-\left(\alpha + \sum_{i=1}^k \beta_i x_i\right)}} \quad (3)$$

For notational convenience, we will denote the probability statement as simply $p(x)$ where x is a notation for the collection of variables x_1 through x_k . Thus, the logistic model may be written as

$$f(X) = \frac{1}{1 + e^{-\left(\alpha + \sum_{i=1}^k \beta_i x_i\right)}} \quad (4)$$

However, since the above logistic model is non-linear function, the logit transformation would be used to make it linear.

$$\text{Logit}(X) = \ln_e \left(\frac{P(x)}{1 - P(x)} \right) \quad (5)$$

Where,

$$P(x) = \frac{1}{1 + e^{-\left(\alpha + \sum_{i=1}^k \beta_i x_i\right)}} \quad (6)$$

This transformation allows us to compute a number, logit $p(x)$, for an individual with independent variables given by x.

$$\text{Logit } P(x) = \alpha + \sum_{i=1}^k \beta_i x_i \quad (7)$$

Thus, the logit of $p(x)$ simplifies to the linear sum. The quantity $p(x)$ divided by $1-p(x)$, whose log value gives the logit, describes the odds for a policyholder not making a claim, with independent variables specified by x.

$$\frac{P(x)}{1 - P(x)} = \text{Odds for individual } X \quad (8)$$

The goal of logistic regression is to correctly predict the category of outcome for individual cases using the most parsimonious model. To this end, a model is created that includes all predictor variables that are useful in predicting the response variable (Kleinbaum & Klein, 1994). For this study, the risk of making an insurance claim are influenced by predictors such as age, distance, billed charges, sex, marital status, length of stay, number of children, level of education, employment, cigarette smoking, alcohol drinking, living condition, health status and income level. The following logistic regression model was fitted to the data.

$$\text{Logit}(P(y = 1)) = \beta + \varepsilon + \sum_{i=1}^{k=14} \beta_i x_i \quad (9)$$

Where P is the probability of claim made, the x's are independent variables of interest, α and the β_i are constant term and coefficients respectively representing unknown parameters and ε is the residual term. The coefficients of the model predictors are tested via the hypothesis as follows:

$$H_0: \beta_j = 0$$

$$H_1: \beta_j \neq 0 \quad j = 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14$$

Once a logistic regression model has been fit to a given set of data, the adequacy of the model is examined by overall goodness-of-fit tests and examination of influential observations. One concludes a model fits if the differences between the observed and fitted values are small and if there is no systematic contribution of the differences to the error structure of the model. A goodness-of-fit test that is commonly used to assess the fit of

logistic regression models is the Hosmer–Lemeshow test (Hosmer & Lemeshow, 1980). Although appropriate estimation methods which take into account the sampling design in estimating logistic regression model parameters are available in various statistical packages, there is a corresponding absence of design-based goodness-of-fit testing procedures. Due to this noted absence, it has been suggested that goodness-of-fit be examined by first fitting the design-based model, then estimating the probabilities, and subsequently using iid-based tests for goodness-of-fit and applying any findings to the design-based model (Hosmer & Lemeshow, 2000). The hypothesis for model fitness can be measured by the Hosmer and Lemeshow test as follows

H_0 : The model fits the data;

H_1 : The model does not fit the data.

3. Empirical Results

3.1 Descriptive Analysis

Table 1, shows the number of observations for the study. The total number of observations for this study was 4549. More than 50% of the people who were sampled were females (56.0%) and the rest (44.0%) were males. The table indicates the frequency of the respondents who made or did not make a claim. Majority of the respondents have made an insurance claim (92.0%), the rest have not made claims in the year (8.0%). The results indicates that majority of the respondents (46.3%) are in the age group of 18-39 years, followed by the age group 40-60 years (19.4%), the rest are in the age groups of 0-17 and 61-100 (19.2% and 15.1%) respectively. The results indicate that majority of the insured who made claims are among the working group aged between 18-39 years (45.2%). The results show that majority of the respondents were unmarried (58.7%) and the rest are married (41.3%). Table 1 shows that majority of the insured sampled have very good or good health status (65.3%), (24.0%) of the insured had fair health status and the rest had poor health status (10.7%). Again, majority of the patients that attended hospital are charged bills between GHS1-400 (58.6%), (26.2%) of the insured that attended hospital were billed between GHS 401-800, the rest were billed GHS 801 or more (9.7%). Majority of the insured who were sampled earned incomes between GHS1-1000 (54.0%), (31.4%) of the insured earned no income and the rest earned GHS1001 or more (14.6%). Majority of the insured travel less or equal to 5km to the hospital (51.2%), and the rest travel more than 5km to the hospital. Majority of the sampled insured persons used outpatient services at the hospital (57.2%), and (36.8%) used inpatient services and the rest (6.0%) had not used the hospital services in the year. Table 1 indicates that majority of the insured persons who were sampled had no child (48.7%), (18.8%) had one child, (17.1%) had two children and (15.4%) had three children or more. Table 1 also shows that more than half of the insured persons had no education or only basic education (62.7%), (15.4%) had education up to senior high school, (6.7%) of the insured persons had professional qualification and (15.2%) had education up to the degree level. (29.9%) of the insured persons were full time workers, (23.9%) were working part-time, (31.9%) were unemployed and (14.3%) were economically inactive. Majority of the insured that were sampled had never smoked (66.4%), (19.7%) were smokers, and (13.9%) were ex-smokers. Majority of the respondents (42.7%) were non drinkers of alcohol, (22.9%) were ex-drinkers, (18.9%) were regular drinkers and (15.5%) were occasional drinkers of alcohol. Majority of the insured persons (58.2%) rent single rooms or double rooms as their living apartment, (22.1%) rent flats or self-contained apartments and (19.7%) of the insured persons are living in their own houses.

Table 1. Composition of the survey population

Age(Years)	Numbers	%
0-17	873	19.2
18-39	2105	46.3
40-60	884	19.4
61-100	687	15.1
Sex	Numbers	%
Male	2002	44.0
Female	2547	56.0
Marital Status	Numbers	%
Married	1879	41.3
Unmarried	2670	58.7
Health Status	Numbers	%
Very Good	1165	25.6
Good	1805	39.7
Fair	1092	24.0
Poor	487	10.7
Billed Charges	Numbers	%
No	250	5.5

GHS 1-400	2666	58.6
GHS 401-800	1191	26.2
GHS 801-1200	246	5.4
GHS 1201-1600	123	2.7
>GHS 1600	73	1.6
Income Level	Numbers	%
No income	1429	31.4
GHS 1-1000	2456	54.0
GHS 1001-2000	600	13.2
GHS 2001-3000	59	1.3
GHS>3000	5	0.1
Distance	Numbers	%
> 5km	2220	48.8
≤ 5km	2329	51.2
Length of Stay	Numbers	%
Non	273	6.0
Outpatient	2602	57.2
Inpatient	1674	36.8
Number of Children	Numbers	%
No Child	2215	48.7
1 Child	854	18.8
2 Children	781	17.1
3+ Children	699	15.4
Level of Education	Numbers	%
No Education	972	21.4
Basic Education	1875	41.4
Senior High School	707	15.5
Professional	307	6.7
Degree	693	15.2
Employment	Numbers	%
Working full-time	1364	29.9
Working part-time	1089	23.9
Unemployment	1454	31.9
Economically Inactive	642	14.3
Cigarette Smoking	Numbers	%
Smoker	895	19.7
Ex-Smoker	634	13.9
Never Smoked	3020	66.4
Alcohol Drinking	Numbers	%
Regular Drinker	864	18.9
Occasional Drinker	702	15.5
Ex-Drinker	1041	22.9
Non-Drinker	1942	42.7
Living Condition	Numbers	%
Own House	894	19.7
Rented Flat/Self Contained	1004	22.1
Rented Single Room/Double Room	2651	58.2
Claim	Numbers	%
Yes	4185	92.0
No	364	8.0

3.2 Odds Ratio Analysis of Risk Factors

The computation of the crude odds ratio for risk factors, X, is given by the estimate Exp (B). The crude odds ratio of risk factor determines the influence it has on the claim outcome. The Wald's and log likelihood ratio tests are also performed to ascertain the significant effect of the risk factors. A probability value of less than or equal to 0.05 was considered to be statistically significant. Hence the inclusion of that risk factor is important in determining the claims outcome Y= 0 or 1.

Table 2. Logistic regression predicting likelihood of health insurance claim

Variable	Estimates	S.E.	Wald	Df	P-Values	e^β
Sex	0.591	0.140	17.688	1	0.000	1.805
Age	-0.290	0.069	17.787	1	0.000	0.748
Marital Status (MS)	0.404	0.193	4.376	1	0.036	1.498
Health Status (HS)	0.168	0.131	1.650	1	0.199	1.183
Billed Charges (BC)	0.744	0.389	3.660	1	0.056	2.104
Income Level (IL)	0.364	0.243	2.242	1	0.136	1.439
Length of Stay (LS)	-5.620	0.521	116.294	1	0.000	0.004
Distance (D)	0.617	0.260	5.630	1	0.018	1.853
Number of Children (NC)	0.094	0.239	0.156	1	0.693	1.099
Level of Education (LE)	-0.395	0.235	2.795	1	0.692	1.098
Employment Status (ES)	0.360	0.690	0.272	1	0.602	0.698
Cigarette Smoking (CS)	-2.927	1.104	7.033	1	0.008	0.054
Alcohol Drinking (AD)	2.793	0.263	111.112	1	0.000	16.329
Living Condition (LC)	0.153	0.077	3.907	1	0.048	1.165
Constant	3.036	0.330	84.459	1	0.000	20.828

The parameters of the model were estimated using maximum likelihood approach. The estimates for each independent variable are interpreted relative to the referenced category. The estimated odds ratio for all parameters is presented in Table 2. The working group (18–40 years) is 0.748 more likely to make a claim compared to children (1–17 years) with 95% confidence interval (p-value = 0.000) is statistically significant. The odds ratio of 1.805 and a confidence interval of 95%, indicates that females are 1.805 as likely to make a claim compared to their male counterparts, giving a similar statistically significant results. Unmarried policyholders are 1.748 as likely as their married counterparts to make insurance claim at 95% confidence interval (p-value = 0.036).

Similarly the results indicates that the odds of making an insurance claim increases by a factor of 1.853 with a confidence interval of 95% when the insured persons travels less than or equal to 5km to attend to the hospital (p-value = 0.018). Table 2 shows an odds ratio of 0.004 indicating that, inpatients make 0.004 insurance claims as likely as their outpatients' counterparts with 95% confidence interval (p-value = 0.000) controlling for other factors in the model. Cigarette smokers are 0.054 more likely to make a claim compared to non-smokers with 95% confidence interval (p-value = 0.008) is statistically significant. The odds ratio of 16.329 and a confidence interval of 95% (p-value = 0.000), indicates that alcohol drinkers are 16.329 more likely to make a claim compared to non-alcohol drinkers giving similar statistically significant results. Those living in rented single/double room apartment are 1.165 more likely to make a claim at 95% confidence interval (p-value = 0.048). The results suggest a non-negligible effect (p-value = 0.056) of the billed charges to influence health insurance claims.

Health status and income level had probability values of more than 0.05, which means that the health status and income level predictor variables are not significant with 95% confidence interval. The logistic regression obtained was as follows (see Table 2 above):

$$\begin{aligned} \text{Logit}(P(y=1)) = & 3.036 + 0.591\text{Sex} - 0.290\text{Age} + 0.404\text{MS} + 0.168\text{HS} + 0.744\text{BC} + 0.364\text{IL} - 5.620\text{LS} \\ & + 0.617\text{D} + 0.094\text{NC} - 0.395\text{LE} + 0.360\text{ES} - 2.927\text{CS} + 2.793\text{AD} + 0.153\text{LC} \end{aligned}$$

It is noted that, the risk factors; sex, age, marital status, length of stay, cigarette smoking, alcohol drinking, living condition and distance are significant at $\alpha = 0.05$ with their respective significance values equal to 0.000, 0.000, 0.036, 0.000, 0.008, 0.000, 0.048 and 0.018. Therefore, these risk factors are relevant in predicting national health insurance claims in Ghana. From Table 2, it is revealing to note that, the risk factors- health status, billed charges, income level, number of children, level of education and employment status are not statistically significant when the other factors are held constant.

Table 3. Odds ratio analysis for risk factors in logistic regression model for health insurance claims

	Model 1 Age + Sex		Model 2 Demographic		Model 3 Socio-Economic		Model 4 smoking/Alcohol Drinking	
	OR	95% C.I.	OR	95% C.I.	OR	95% C.I.	OR	95% C.I.
Sex								
Male	1.00		1.00		1.00		1.00	
Female	1.80**	1.33-1.68	1.42**	1.26-1.60	1.23**	1.08-1.40	1.17*	1.11-1.46
Age								
0-17	1.00		1.00		1.00		1.00	
18-39	0.75**	1.26-2.04	1.39*	1.05-1.83	1.56**	1.17-2.08	1.65**	1.23-2.22
40-60	1.06	0.83-1.36	1.02	0.79-1.33	1.24	0.94-1.62	1.21	0.91-1.60
61-100	1.51**	1.18-1.94	1.28	0.95-1.71	1.06	0.78-1.44	1.36	0.98-1.86
Marital Status								
Married		1.00		1.00		1.00		
Unmarried		1.49		1.20-1.69	1.35**	1.13-1.61	1.27**	1.05-1.53
Number of Children								
No Child		1.00		1.00		1.00		
1 Child		1.03		0.85-1.25	0.92	0.75-1.12	0.94	0.76-1.16
2 Children		0.92		0.75-1.13	0.77*	0.62-0.96	0.80	0.64-1.00
3+ Children		1.09		0.66-1.20	0.58**	0.42-0.79	0.59**	0.43-0.82
Income Level								
No Income					1.43**	1.11-1.95	1.32	0.99-1.77
GHS 1-1000					1.28	0.99-1.66	1.19	0.91-1.56
GHS 1001-2000					1.26	0.98-1.61	1.24	0.96-1.60
GHS 2001-3000					1.25	0.96-1.63	1.24	0.94-1.62
GHS>3000					1.00		1.00	
Level Education								
No Educ.					1.098	0.88-1.40	1.16	0.92-1.47
Basic Educ.					1.12	0.91-1.38	1.19	0.94-1.50
S.H.S.					1.24	1.00-1.54	1.33*	1.06-1.66
Professional					1.36*	1.07-1.71	1.46**	1.15-1.86
Degree					1.00		1.00	
Employment Status								
Working Full-Time					1.00		1.00	
Working Part-Time					1.18	0.97-1.43	1.18	0.97-1.44
Unemployed					0.69**	1.18-2.32	1.58	1.10-2.24
Econ. Inactive					2.14**	1.79-2.56	2.06**	1.71-2.48
Living Condition								
Own House					1.00		1.00	
Rented Flats/Self Contained					1.35**	1.10-1.67	1.22	0.98-1.52
Rented Single/Double					1.65**	1.22-1.68	1.32*	1.11-1.56
Room								
Cigarette Smoking								
Smoker							0.05*	1.04-1.46
Ex-Smoker							0.98	0.83-1.14
Never Smoked							1.00	
Alcohol Drinking								
Regular Drinker							10.09**	1.77-2.47
Occasional Drinker							6.24**	4.05-5.39
Ex-Drinker							1.76**	1.45-2.13
Non-Drinker							1.00	
-2log likelihood	7481.59		7415.25		7186.16		6711.49	
Δ df	6		6		6		6	
Nagelkerke R ²	0.017		0.030		0.075		0.164	

Note: *P<0.05, **P<0.01.

We now look at how other variables affect the relationship between socio-economic variables and health insurance claims using logistic regression models ($n = 4549$). The order of variables entered into the models reflects an a priori judgment of the primary causal ordering between variables. Age and gender are prior variables (model 1), followed by inclusion of marital status and number of children in model 2. The primary direction of association assumes that socio economic factors have an effect on smoking and alcohol drinking. Smoking and alcohol drinking are assumed to be causally prior to making health insurance claims and included in the model. The independent effects of each of the socio-economic characteristics on national health insurance claims are shown in model 3 after adjusting for age, marital status, number of children, and the other variables.

Being unemployed or economically inactive is associated with making national health insurance claims (model 3). The odds ratio for the unemployed reduces slightly following adjustment for alcohol drinking and cigarette smoking. The unemployed and economically inactive retain significantly elevated odds of national health

insurance claims. Educational qualifications are linked to national health insurance claims with significantly higher odds ratios for respondents with no qualifications. However, these associations with national health insurance claims become greater in model 4 after adjusting for alcohol drinking and cigarette smoking. The findings suggest that health insurance claims is linked to a portion of the insured persons who are more disadvantaged in socio-economic circumstances.

Table 4. Odds ratios and coefficients for interaction effects in logistic regression models of GNHI claims

Variables	O.R.	β	Standard Error
Sex*Income	0.2	-1.79	(0.48)***
Sex*Length of Stay	4.9	1.60	(0.31)***
Sex*Distance	2.1	0.73	(0.27)**
Sex*Cigarette Smoking	7.7	2.04	(0.33)**
Sex*Alcohol Drinking	5.1	1.63	(0.29)***
Sex*Living Condition	0.3	-1.32	(0.46)**
Age*Income Level	2.0	0.68	(0.21)***
Age*Length of Stay	0.4	-0.98	(0.49)*
Age*Distance	0.1	-2.44	(0.54)***
Age*Cigarette Smoking	1.7	0.56	(0.34)
Age*Alcohol Drinking	0.3	-1.24	(0.47)**
Age*Living Condition	1.4	0.33	(0.32)
Age*Level of education	0.7	-0.37	(0.53)
Cigarette Smoking*Alcohol Drinking	20.0	2.99	(0.37)***
Living Condition*Income Level	7.9	2.07	(0.34)***

Note: +p<0.10,*p<0.05,**p<0.01,***p<0.001, adjusted for design effects. Notes: Coefficients are shown for the intercept and variables included in interaction terms. Estimates are based on logistic regression using normalized sampling weights and taking the sample design into account. Sample sizes are final, weighted samples.

Table 4 shows estimated effects of sex, age, income level, length of stay, distance, cigarette smoking, alcohol drinking, living condition and level of education and terms representing the interaction between sex, age, and those variables on the odds of making national health insurance claims. Sex*Income interaction shows that there is a positive association with national health insurance claims and this does not vary significantly by sex. Females who are in-patients are 4.9 times likely to make an insurance claim compared with males who are out-patients. Females who live 5km or more from the hospital are 2.1 times more likely to make an insurance claim compared with men who live less than 5km from the hospital. Table 4 shows that women who are cigarette smokers are 7.7 times more likely to make an insurance claim compared with men who are cigarette smokers at 95% confidence interval. The table shows that females who take alcohol have a positive association with national health insurance claims. Females who take alcohol are 5.1 times more likely to make an insurance claim compared to men who drink alcohol at 95% confidence interval. Sex had a significant effect with living condition at 95% confidence interval. It shows that females who live in rented single/double rooms are 0.3 times likely to make an insurance claim compared with men rent single/double rooms. Age and income interaction shows that there is a positive association with national health insurance claims at 95% confidence interval. Those who are 40 years and below, and have no income are 2.0 times more likely to make health insurance claims compared with the aged who earn high income at 95% confidence interval.

The table shows that age and length of stay interaction has a positive association with national health insurance claims. Those who are 40 years and below and in-patients are 0.4 more likely to use the hospital services and therefore make an insurance claim compared to the aged who use out-patient services of the hospital. Age and distance interaction shows there is a positive association with national health insurance claims at 95% confidence interval. Those who are 40 years and below and stay less than 5km to the hospital are 0.1 more likely to make an insurance claim compared to the aged who stay more than 5km to the hospital at 95% confidence interval. The interaction between age and cigarette smoking shows that there is no association with national health insurance claims. The interaction between age and level of education shows no association with national health insurance claims and the interaction is statistically insignificant at 95% confidence interval. Cigarette smoking and alcohol drinking interaction shows that there is a positive association with national health insurance claims. Those who are regular smokers and regular alcohol drinkers are 20 times more likely to make health insurance claims

compared to non-smokers and non-alcohol drinkers at a confidence interval of 95%.

4. Conclusion

In this study, almost all of the socio-economic and demographic factors were found to have significant association with claims made in the national health insurance scheme. Associations with socio-demographic variables were explored using multiple logistic regressions. The first model looked at associations of each major variable (odds ratios) adjusting for cigarette smoking and alcohol drinking which emerged as a strong determinant in univariate analyses. The second model was fully adjusted to disentangle interactions. The dependent variable is the national health insurance claims. Independent variables were self-explanatory. The pattern was similar among men and women. When adjusted for cigarette smoking and alcohol drinking, those living in rented single rooms/double rooms were twice as likely to make national health insurance claims as those in own house. With all variables entered, cigarette smoking and alcohol drinking remained strong determinants. Women reporting no or low income status tended to make more insurance claims. Together both confirm that women have the propensity to make national health insurance claims more than men according to socio-economic characteristics. The regression model was fully adjusted, with variables entered simultaneously. Less well educated women were more likely to make an insurance claim. Disadvantaged people (younger people, with lower education and people with no income, and among women) were more likely to make an insurance claim.

In a previous study carried out in Malaysia (Aizuddin et al., 2011) on willingness to join and contribute to national healthcare financing scheme, it focused on the farmers in Selangor in which age, education level and per capita income of respondents influenced willingness to pay for National Healthcare Financing scheme.

This study provides evidence that health insurance claims, in the eastern region of Ghana could be described as high. One major shortcoming of using this data is that they only represent those patients who visited the Koforidua regional hospital. Meanwhile, some insured persons may have sought care in other health facilities in the region.

The paper systematically analyzed the main factors that affect the pricing of national health insurance premiums in Ghana and proposed that sex, age, marital status, cigarette smoking, alcohol drinking, living condition, distance and length of stay at the hospital should be used in the determination of the Ghana National health insurance premiums since they considerably influence claims.

Based on the statistically significant influential factors which should be used to determine claims, the study with the help of logistic regression has developed a model to predict the likelihood of the Ghana national health insurance claims. The paper has designed a mechanism for evaluating prospective insured persons and determining their risk factors which will help the insurance company to determine the appropriate premium for prospective clients.

The paper established that insured persons who stay 5km away from the hospital make less National health insurance claims. The paper has established a scientific model for insurance premium that include the cost of transportation. This will enable the poor to have access to means of transport easily especially in times of emergency. This measure will relieve majority of the insured who may be denied access to health care at the time when they cannot afford the cost of transportation to the health facility.

The Ghana National Health Insurance Scheme is on course. Much more needs to be done beyond what currently pertains, for the appropriate determination of premiums for Health Insurance in Ghana; since the sustainability and credibility of the health insurance industry is dependent on getting it right so far as the determination of premiums is concerned. This can be facilitated by institutionalizing the collection of appropriate data on a continuous basis, to assist this important area of the industry's administration, and all effort must be made in this direction if success is envisaged. The government should consider building more health centers, clinics and cheap-compounds in at least every community, to help reduce the travel time in accessing health care. The ministry of health and the Ghana health service should engage older citizens by encouraging them to use hospitals when they are sick instead of other alternative care providers.

Acknowledgements

We are grateful to Badu Kyere Kwame, national health insurance claims manager at the Koforidua Regional Hospital, for his efforts in securing the data used for this study.

References

Abdul-Aziz, A. R., Harris, E., & Munyakaza, L. (2012). Risk factors in malaria mortality among children in

- northern Ghana. *International Journal of Business and Social Research*, 2(5), 35–45.
- Aikins, J. W., & Prinstein, M. J. (2004). Cognitive moderators of the longitudinal association between rejection and adolescent depressive symptoms. *Journal of Abnormal Child Psychology*, 32(2), 147–158. <http://dx.doi.org/10.1023/B:JACP.0000019767.55592.63>
- Aizuddin, A. N., Hod, R., Rizal, A. M., Yon, R., & Al Junid, S. M. (2011). Ability and Willingness to pay for Healthcare and contribute for National Health Financing Scheme among Farmers in Selangor. *Malaysian Journal of Community Health*, 17(1), 56–63. Retrieved from <http://www.communityhealthjournal.org/2428>
- Atim, C. (1999). Social movements and health insurance: A critical evaluation of voluntary, non-profit insurance schemes with case studies from Ghana and Cameroon. *Social Science & Medicine*, 48, 881–896. [http://dx.doi.org/10.1016/S0277-9536\(98\)00390-6](http://dx.doi.org/10.1016/S0277-9536(98)00390-6)
- Azzalini, A., Bowman, A. W., & Härdle, W. (1989). On the use of nonparametric regression for model checking. *Biometrika*, 76(1), 1–11. <http://dx.doi.org/10.1093/biomet/76.1.1>
- Boland, A. L. (2007). The identification problems and the validity of economic models. *South African Journal of Economics*, 36(3), 236–240. <http://dx.doi.org/10.1111/j.1813-6982.1968.tb02298.x>
- Brown, C. C. (1982). On a goodness-of-fit test for the logistic model based on score statistics. *Communication in Statistics. Theory and Methods*, 11(10), 1087–1105. <http://dx.doi.org/10.1080/03610928208828295>
- Carrin, G. (2002). Social health insurance in developing countries: A continuing challenge. *International Social Security Review*, 55(2), 57–69. <http://dx.doi.org/10.1111/1468-246X.00124>
- Carrin, G., & James, C. (2005). Social health insurance: Key factors affecting the transition towards universal coverage. *International Social Security Review*, 58(1), 45–64. <http://dx.doi.org/10.1111/j.1468-246X.2005.00209.x>
- De Allegri, M., Sanon, M., & Sauerborn, R. (2006). To enrol or not to enrol? A qualitative investigation of demand for health insurance in rural West Africa. *Soc. Sc. Med.*, 62(6), 1520–1527. <http://dx.doi.org/10.1016/j.socscimed.2005.07.036>
- Ghana Statistical Service-Annual Report. (2012).
- GNA. (2005, May). *Collection of NHIS premiums Underway. Daily Graphic*.
- Government of Ghana, MoH. (2003). National health insurance bill. Retrieved from <http://www.ghana.gov.gh/pbcopin/nhib.pdf>
- Hosmer, D. W., & Lemeshow, S. (1980). Goodness-of-fit tests for the multiple logistic regression models. *Communication in Statistics. Theory and Methods*, 9(10), 1043–1069. <http://dx.doi.org/10.1080/03610928008827941>
- Hosmer, D. W., & Lemeshow, S. (2000). *Applied Logistic Regression*. New York: Wiley. <http://dx.doi.org/10.1002/0471722146>
- ILO. (2003). *Global health trust. Health insurance in Ghana*. Retrieved from http://www.ilo.org/global/wwcms_075517/index.htm
- Kleinbaum, D. G., & Klein, M. (1994). *Logistic regression: A self-learning text* (pp. 15–26) (2nd ed.). New York: John Wiley and Sons Publishers. <http://dx.doi.org/10.1007/978-1-4757-4108-7>
- Kleinschmidt, I., Sharp, B. L., Clarke, G. P. Y., Curtis, B., & Fraser, C. (2001). Use of Generalized Linear Mixed Models in the Spatial Analysis of Small-Area Malaria Incidence Rates in KwaZulu Natal, South Africa. *American Journal of Epidemiology*, 153(12), 1213–1221. <http://dx.doi.org/10.1093/aje/153.12.1213>
- Le Cessie, S., & Van Houwelingen, H. C. (1995). Testing the fit of a regression model via score tests in random effects models. *Biometrics*, 51, 600–614. <http://dx.doi.org/10.2307/2532948>
- Le Cessie, S., & Van Houwelingen, J. C. (1991). A goodness-of-fit test for binary regression models, based on smoothing methods. *Biometrics*, 47, 1267–1282. <http://dx.doi.org/10.2307/2532385>
- Ministry of Health Ghana. (1996a). *Review of health insurance in Ghana*. Unpublished monograph.
- Ministry of Health Ghana. (1996b). *Cash and carry programme: Report on programme review*. Unpublished report.
- National Health Insurance Act. (2003). Ghana.

- National Health Insurance Regulation. (2004). Ghana.
- Sachs et al. (2005b). *Investing in development: A practical plan to achieve the millennium development goals, united nations millennium project*. London: Earthscan.
- Sachs, J. D. (2005a). *The end of poverty: Economic possibilities of our time*. New York: Penguin.
- Samuel, A., & Zhao, X. (2012), A logistic Regression model for Ghana national health insurance claims. *International Journal of Business and Social Research*, 2(7), 139–147.
- Su, J. Q., & Wei, L. J. (1991). A lack-of-fit test for the mean function in a generalized linear model. *Journal of the American Statistical Association*, 86(414), 420–426.
<http://dx.doi.org/10.1080/01621459.1991.10475059>
- WHO. (2010a). *The world health report 2010. Working Together for Health*. World Health Organisation, Geneva.
- WHO. (2010b). *The world health report 2010. Health systems: Improving performance*. World Health Organisation, Geneva.

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).