

A Service of

ZBW

Leibniz-Informationszentrum Wirtschaft Leibniz Information Centre for Economics

Momanyi, Kevin

Working Paper An Econometric Analysis of The Impact of Telecare on The Length of Stay in Hospital

Suggested Citation: Momanyi, Kevin (2017) : An Econometric Analysis of The Impact of Telecare on The Length of Stay in Hospital, ZBW - Leibniz Information Centre for Economics, Kiel, Hamburg

This Version is available at: https://hdl.handle.net/10419/279407

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



WWW.ECONSTOR.EU

An Econometric Analysis of The Impact of Telecare on The Length of Stay in Hospital

Kevin Momanyi

September, 2017

Abstract: This paper presents some preliminary results of a study investigating the effect of telecare on the length of stay in hospital using linked administrative health and social care data in Scotland. We make various assumptions about the probability distribution of the outcome measure and formulate three Negative Binomial Models to that effect i.e. a basic Negative Binomial Model, a zero-inflated Negative Binomial Model and a zero-truncated Negative Binomial Model. We then bring the models to data and estimate them using a strategy that controls for the effects of confounding variables and unobservable factors. These models provide an alternative to the Propensity Score Matching technique used by the previous studies. The empirical results show that telecare users are expected to spend a shorter time in hospital than non-users, holding other factors constant. The results also show that older individuals, males, rural residents and individuals with comorbidities have a longer length of stay in hospital, on average, than their counterparts, all things equal. Future research will involve conducting a sub-group analysis, investigating the effectiveness of various telecare devices and determining the impact of telecare on admission to hospital.

Keywords: Telecare, Negative Binomial Models, length of stay in hospital.

JEL classifications: C32, C36, C55, D13, I12.

1. Introduction

Telecare refers to the use of devices to monitor individual's health and safety at home, and may complement or substitute for social care and unpaid care. It covers a wide range of devices from the basic community alarm, which allows the users to call for help by simply pressing a button to more sophisticated devices that allow for remote exchange of clinical data between the users and their care providers, and virtual consultation using audio and video technology (The Scottish Government, 2012, Brownsell, 2008). The use of telecare is beneficial to both the users and their carers. For instance, telecare reduces the need for residential care, mainly through delayed admissions and also offers increased choice and independence for the users. Telecare can also reduce pressures on carers by freeing up some of their time thus giving them more personal freedom (Clark et al., 2007; Giordano, Clark and Goodwin, 2011).

In this paper, we investigate whether telecare use influences the length of stay in hospital. We argue that the use of telecare could substitute for some health care services (especially nursing care services) that would have otherwise been provided in hospital, thereby resulting in a shorter length of stay in hospital. Our investigation is also motivated by several reasons: first, The Scottish Government, through its 2020 vision, aims at developing policies that reduce the length of stay in hospital and support selfmanagement of chronic illnesses (Scottish Government, 2015). It could therefore be the case that the promotion of telecare is one way of achieving these goals. Second, a shorter length of stay in hospital is associated with a reduction in the costs of care, and, as such, it makes economic sense to provide a way forward on how the length of stay in hospital can be reduced. Third, the length of stay in hospital has been widely used in the literature as one of the measures of health systems' performance (see for example, Almashrafi et al., 2016, Casteli et al., 2015, Steel, 2012) and it would thus be interesting to find out the relative contribution of telecare.

The literature on the effects of telecare is scanty, with only a handful of studies investigating the effect of telecare on the length of stay in hospital (see for example, Akematsu, 2012 and Akematsu 2013 which were conducted in Japan). The studies estimate the treatment effect using the Propensity Score Matching Technique (in which case the treatment effect is obtained after matching telecare users with non-users according to several characteristics) and note that telecare users have a comparatively short length of stay in hospital. We contribute to the existing literature by conducting an empirical study using Scottish data; specifically, linked administrative health and social care data. We

also formulate three models that can be used to derive the causal impact of telecare in the absence of randomized data. These models provide an alternative to the Propensity Score Matching Technique employed by the previous studies.

The specific objectives of the paper are to determine the effect of telecare on the length of stay in hospital, controlling for confounding factors, and to identify the factors that are associated with the length of stay in hospital. The rest of this paper progresses as follows: Section 2 describes the data; Section 3 discusses the methodology; Section 4 presents the results, and Section 5 concludes.

2. The data

We use the Scottish Linked Health and Social care data set which is a merger of four different information sources: the Scottish Morbidity Records (SMRs), prescribing data, Self-Directed Support data and the March 2011 Homecare Census data. The majority of our covariates of interest are obtained from the SMRs and the Homecare Census.

The SMRs contain episode level data on acute hospital admissions (SMR01) and psychiatric care admissions (SMR04). Some of the variables in these data sets are marital status, sex, date of admission to hospital, reason for admission, date of discharge from hospital and significant facility. We construct our outcome measure of interest as the number of days that a particular individual spends in hospital while receiving treatment.

The Homecare Census contains information on telecare use as well as a number of demographic characteristics. Examples of variables in this data set are age, ethnicity, living arrangement (whether or not a particular individual lived alone), telecare use and client group (which is a categorical variable with the following categories: Dementia and Mental Health, Learning disability, Frail elderly and Physical disability). The variable for telecare use comprises linked pill dispensers, linked smoke detectors, bogus caller butt– ons, property exit sensors and automated motion sensors.

3. Methodology

Modelling framework

The conceptual model adopted by this paper follows closely the Andersen's Behavioural Model of Health Services Use (Andersen 1995). According to the model, the length of stay in hospital is related to 'predisposing', 'enabling' and 'need' factors.

Predisposing factors are those factors that increase the likelihood of health service



Figure 1: A conceptual framework for analyzing the effect of telecare on the length of stay in hospital

use. They include demographic factors e.g. age and sex, people's beliefs and their social structure. Enabling factors are those factors that enable or impede service use e.g. income levels and access to health care services. The need factors are those factors that reflect the individuals' care needs e.g. multimorbidity and polypharmacy (see Andersen, 1995 for a more comprehensive discussion on these factors). In this paper, we conceptualize tele–care as an enabling factor. We assume that individuals use telecare to substitute for some of the health care services provided in hospital, thereby resulting in a shorter length of stay in hospital.

The conceptual model also shows that the observed variation in the length of stay in hospital may be brought about by factors operating at the health and social care system. We can further observe from the model that a particular individual's length of stay in ho-

spital is influenced by the individual's preferences regarding the use of health care services and his/her level of endowment, although indirectly. This study relates the length of stay in hospital to age, sex, rurality and client group (which are 'predisposing' factors), telecare use (which is an 'enabling' factor) and comorbidity (which is a 'need' factor). This is illustrated in Figure 1 on page 3.

In order to operationalize the model, we develop an analytical framework that takes the form of a constrained utility maximization problem. We assume that individuals maximize utility and, by extension, their health status by using telecare devices to sub– stitute for some health care services provided in hospital. The basic utility function for a particular individual is expressed as follows:

$$U_i = u(d, T, H) \tag{1}$$

where U is the individual's utility, $d = \{u_{i1}, u_{i2}, \dots, u_{in}\}$ is a set of 'predisposing' and 'need' factors, T is an indicator for telecare use and H is the individual's health product-ion function.

The health production function is, in turn, related to some observable health enhancing inputs, telecare use and biological endowments, μ_i . This is as shown below:

$$H_i = H(b_i, T_i, \mu_i) \tag{2}$$

where $b = \{b_{i1}, b_{i2}, \dots, b_{in}\}$ is a set of health enhancing inputs.

Following Equations (1) and (2), we can formulate the estimated version of Equation 1 as follows:

$$Y_i = \alpha + \beta U_i + \varepsilon_i \tag{3}$$

Notice from Equation 3 that an endogeneity problem could arise since the variable for telecare use, T, is contained in both Equations (1) and (2). In this study, we suspect that telecare use is potentially endogenous since individuals typically choose whether or not to use telecare via Self-Directed Support (The Scottish Government, 2010) and hence the unobservable factors guiding their choices might be correlated with the length of stay in hospital, Y. Notice also from Equations (1) and (2) that since a particular individual's utility is directly related to his/her level of endowment and that T is a choice variable, a problem of unobserved heterogeneity could arise if the differences in endowments and preferences within the study population cause the effect of telecare on the length of stay

in hospital to differ among various individuals. We therefore use the Two-Stage Residual Inclusion approach (Terza, 2008; Awiti, 2014; Wooldridge, 2002; Garren 1984) to control for potential endogeneity and potential unobserved heterogeneity¹. This technique involves estimating a reduced form model of telecare using a probit model, obtaining the residuals of the model and then including, in Equation 3, appropriate control terms based on these residuals. The model is said to suffer from endogeneity and unobserved hetero–geneity if the control terms are found to be statistically significant.

The study could also suffer from selectivity bias since we only have information on the length of stay in hospital for those individuals who were admitted to hospital or psychiatric care during the 2010/2011 financial year. Furthermore, as is expected of linked data sets, a selection issue could be brought about by missing data. We therefore control for potential selectivity bias using the approach proposed by Lee (1980). This method is implemented in two steps. In the first step, we estimate a probit model relating the probability of inclusion into the sample to various exogenous variables and instruments, and then obtain the residuals of the model. In the second step, we estimate Equation 3 but with the residuals obtained from the probit model as an additional regressor. Statistical significance of this additional variable is indicative of selectivity bias.

Empirical model

We estimate the treatment effect using a count regression model since the dependent variable, Y, is a count variable (Long and Freese, 2006; Atkins et al., 2013).

$$Y = \{0, 1, 2, \dots, n\}$$
(4)

We assume that Y is generated by an unobservable Poisson process that is given by:

$$Pr(Y = v) = \frac{\exp(-\lambda)^{v}}{v!}$$
(5)

where λ is the Poisson parameter being estimated and v is the observable count.

Since *Y* is a count variable such that $Y \in \mathbb{N}^0$ and given that $X \in U$, where $X = \{x_1, x_2, x_3, x_4, x_5\}$ and T = (0,1), the basic count regression model can be written as

¹ Endogeneity is defined as the correlation between a particular independent variable and the stochastic random error term, whereas unobserved heterogeneity is said to occur if the unobservable factors that bring about endogeneity cause the treatment effect to vary within the study population (Greene, 2002; Cameron and Trivedi, 2005). The two estimation issues are common place in health econometrics, and failure to control for them could lead to inconsistent estimates of the treatment effect (Cameron and Trivedi, 2005).

$$\lambda = \exp(\beta_{\circ} + \sum_{i=1}^{5} \beta_{i} X_{i} + \omega T)$$
(6)

or equivalently

$$\log \ \lambda = \beta_{\circ} + \sum_{i=1}^{5} \beta_{i} X_{i} + \omega T$$
⁽⁷⁾

where β and ω are the coefficients being estimated.

Notice from Equation 5 that the exponential function restricts Y to be positive. A special characteristic of the model is that it assumes equidispersion (Karazsia and Van–Dulmen, 2008; Zou, 2004; Grogger and Carson, 1991; Atkins et al., 2013; Greene, 2012) i.e. $E(Y|X) = \lambda = var(Y|X)$. To put it simply, the expected value of Y is equal to its variance.

Following Lloyd (2007) and McCullagh and Nelder (1984), if we let λ be a gamma random variable $\lambda \sim \Gamma(s, \theta)$ with a shape parameter, *s*, and a scale parameter , θ , then the Poisson Model in Equations (6) and (7) becomes the Negative Binomial Model with the following properties: (i) $var(Y|X) = \lambda(1 + \alpha\lambda)$ and (ii) $E(Y|X) = \lambda$. Notice that, unlike the Poisson Model which assumes equidispersion, the Negative Binomial Model is less restrictive as it allows for the variance of *Y* to exceed its expected value. Notice also that the Negative Binomial Model nests the Poisson Model as a special case. Specifically when $\alpha = 0$. The Negative Binomial Model is given by:

$$\log r = \beta_{\circ} + \sum_{i=1}^{5} \beta_i X_i + \omega T$$
(8)

where r is the parameter being estimated and follows a gamma-poisson mixed distribution.

Various versions of the Negative Binomial Model may be formulated depending on the nature of overdispersion. For example, one may simply use a basic Negative Binom– ial Model and rely on its statistical properties to handle the underlying overdispersion. Alternatively, one may use a zero-inflated Negative Binomial Model when *Y* contains more zeros than would normally be expected of a gamma-poisson mixed distribution (Chipeta et al., 2014; Atkins and Gallop, 2007; Cameron and Trivedi, 2005) or a zerotruncated Negative Binomial Model when the outcome measure only considers positive integers (Greene, 2002; Cameron and Trivedi, 2005; Groger and Carson, 1991). In this paper, we present the empirical results of the three formulations. The first model, Model 1, is a basic Negative Binomial Model; the second model, Model 2, is a zero-inflated Negative Binomial Model, whereas, the third model, Model 3, is a zero truncated Negative Binomial Model.

Model 1: We let the unobservable process generating the observable count , *r*, be a gamma-poisson mixed distribution, and (ii) $var(Y|X) = \lambda(1 + \alpha\lambda) > \lambda$. This model is shown in Equation 8. It is important to note that formulating the model in this way is similar to specifying a generalized linear model with a log-link function to transform the expectation of *Y* to $X\beta' + \omega T$ (Bates, 2010; Turner, 2008; McCullagh and Nelder, 1984).

$$g(Y) = \eta = \beta_{\circ} + \sum_{i=1}^{5} \beta_i X_i + \omega T + \varepsilon$$
⁽⁹⁾

where g is the link function, η is the linear predictor function and ε is a stochastic random error term.

Model 2: We let *r* be composed of two data generating processes: (i) a Bernoulli process generating the structural zeros such that for all values of *Y*, the probability that Y = 0 is constant i.e. Pr(Y = 0) = p and (ii) a gamma-poisson process that genera-tes *Y*, some of which may be zero. In our case, *Y* is such that there exists $Y_1, Y_2 \in Y$, where $Y_1 = \{0\}$ and $Y_2 = \{0\}$ with Y_1 representing day case charges and Y_2 representing the observations for the individuals who were never hospitalized during the period of analysis. We can therefore conceptualize $Y_3 = \{0, 1, 2, ..., n\}$ such that $Y_1 \subset Y_3$ and $Y_2 \notin Y_3$. Given that *r* is defined by two data generating processes, the corresponding empirical model is a two-part model. The first part is a Negative Binomial Model that predicts Y_3 and has its usual properties i.e. $\lambda \sim \Gamma(s, \theta)$ and $var(Y|X) = \lambda(1 + \alpha\lambda) > E(Y|X)$. The second part, on the other hand, is a logit model that predicts Y_1 . The model can be summarized as follows:

log
$$r = \beta_{\circ} + \sum_{i=1}^{5} \beta_{i} X_{i} + \omega T$$
, where $r \Rightarrow Y_{3}$ is observed (10)

$$Pr(Y = 0|X) = \Lambda(\eta)$$
, where Λ is the logistic link function. (11)

Model 3: We let Y be truncated at zero such that there exists $Y_4 \in Y$, where $Y_4 = \{1, 2, ..., n\}$ and $Y_4 \subset Y$. We further assume that unlike in Models (1) and (2) where r is

simply a gamma-poisson process, the unobservable process generating Y_4 follows a truncated gamma-poisson distribution since $Y_4 \in \mathbb{Z}^+$. Following Grogger and Carson, (1991), it can be shown that $E(Y_4|X) = \lambda R$, where R is a truncation factor given by: $[1 - Pr(0)]^{-1}$. The model is written as

log
$$r = \beta_{\circ} + \sum_{i=1}^{5} \beta_{i} X_{i} + \omega T$$
, where $r \Rightarrow Y_{4}$ is observed. (12)

The estimation strategy for Models (1), (2) and (3) is implemented in two stages:

1. We estimate a reduced form model of telecare use and a sample selection model as follows:

$$\gamma = \beta_{\circ} + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 G + \varepsilon_1$$
(13)

where $T, S \in \gamma$ and $S = \begin{cases} 1 & \text{if } Y, Y_3 \text{ or } Y_4 \text{ is observed,} \\ 0 & \text{otherwise.} \end{cases}$

 X_1, X_2, X_3, X_4 and X_5 are the exogenous variables of the model,

G is a vector of instrumental variables,

 ε_1 is a stochastic random error term.

2. We then obtain the residuals from the first stage models and include them in our econometric models of interest as controls for potential endogeneity, and potential selectivity bias. In order to control for unobserved heterogeneity, we also include, in the substantive models, an interaction term of the variable for telecare use with its residuals.

Identification strategy

Since the variable for telecare use is potentially endogenous and there could also be a selectivity issue in our study, we need two instrumental variables in the first stage models. The variables that serve as instruments, however, should be highly correlated with the variable for telecare use but not determined in Models (1), (2) and (3).

In this paper, we use the 'proportion of telecare users in each local council area in Scotland' and the 'Scottish Index of Multiple Deprivation (SIMD)' as instrumental variables. We expect the proportion of telecare users to be related to telecare use in that the higher the proportion of telecare users in a particular local council area, the higher the likelihood of telecare use in that local council area. Similarly, we expect an inverse relationship between telecare use and SIMD since the higher the area level deprivation, the lower the likelihood of accessing health and social care services including telecare. We however do not expect our chosen instrumental variables to be determined in the substantive models since they are aggregated measures at the population level.

4. Results

Table 1: Variable definitions

Table 2: Descriptive Statistics

Table 3: The first stage models

Table 4: The second stage models

The results of the study in this paper are presented in two parts. The first part contains a description of the study covariates and the descriptive statistics. Table 1 on page 10 gives the variable definitions while Table 2 on page 11 presents the descriptive statistics. The second part, on the other hand, contains the results of the first stage and second stage models.

Looking at Table 2, we can observe that the table has three columns. The first column contains the variable names; the second column contains the number of observations for each variable; while the third column contains the mean values (or medians and proporti– ons where applicable).

The results in Column 3 of the table show that the mean age of the homecare clients included in the sample is approximately 75 years. The descriptive statistics also show th– at about half of the total observations belonged to the frail elderly; approximately 18% of the observations were for the physically disabled homecare clients; about 7% were for the clients with a diagnosis of dementia or other mental health illnesses, and approximately 24% of the observations were for the homecare clients with learning disabilities. The results further show that approximately 51% of the total observations belonged to the in– dividuals with three or more comorbid conditions; implying that a substantial proportion of the homecare clients included in the sample were living with multiple conditions.

Variable	Definition
Length of stay in hospital, Y	The number of treatment days.
Age, X_1	Age of an individual in years.
Square of age	The square of age.
Comorbidity status, X_2	Coded 1 if individual has at least three comorbid conditions, 0 otherwise.
Client group, X_3	Coded 1 if 'Dementia and Mental Health', Coded 2 if 'Learning disability', Coded 3 if 'Physical disability', Coded 4 if 'Frail elderly'.
Sex, X_4	Coded 1 if individual is female, 0 otherwise.
Area of residence, X_5	Coded 1 if the area of residence is rural, 0 otherwise.
Telecare use, T	Coded 1 if individual used telecare, 0 otherwise.
Inclusion into the sample	Coded 1 if an observation is included in the estimation sample, 0 otherwise.
Project-ID	A unique reference number identifying each individual in the data set.
SIMD-decile	10 categories of the Scottish Index of Multiple Deprivation in ascending order; a measure of area level deprivation.
Trend	A time trend variable generated by STATA software; unit of time=1 week and the time period is the 2010/2011 financial year.
Proportion of telecare users in each local council area	The proportion of telecare users in each local council area in Scotland.

 Table 1: Variable definitions

Table 3 on page 12 contains the estimated average marginal effects of the reduced form model of telecare and the selection model. The average marginal effect of a particular covariate is interpreted as the change in the probability of observing the outcome measure due to a unit change in the covariate (Long and Freese, 2006). According to the results in the table, we can observe that the two instrumental variables have their expected signs. Specifically, the results show that, holding other factors constant, the variable for the proportion of telecare users in each local council area is directly related to telecare use, whereas the higher the area level deprivation, the lower the likelihood of using telecare, controlling for the other covariates in the model.

The results also show that the variables for age and client group are statistically significant at 5% level of significance. Specifically, the results indicate that a one year increase in age is associated with an increase in the probability of telecare use by 0.006,

Variable	Study sample	
	Ν	М
Length of stay in hospital	15,204	2.00
Age	15,204	75.22
Area of residence	15,158	0.09
Male	15,204	0.36
Female	15,204	0.64
Telecare use	15,204	0.11
Dementia and Mental Health	15,204	0.07
Learning disability	15,204	0.24
Physical disability	15,204	0.18
Frail elderly	15,204	0.52
Comorbidity status	15,204	0.51
Proportion of telecare users	15,204	0.15
SIMD-decile	15,050	5.00
Trend	15,204	24.0

Table 2: Descriptive statistics

Notes: N=number of observations; M=mean/median; s.d. =standard deviation;

other factors held constant. We can also observe from the table that the individuals with dementia or other mental health illnesses and those with learning or physical disabilities are more likely to use telecare than their counterparts, although the effect of 'Physical disability' on telecare use is not statistically significant at 5% level of significance.

Further, the results show that, other factors held constant, about 0.01% of the observed probability of telecare use is unexplained by the model since the average marginal effect of the trend variable is 0.0001 and statistically significant at 5% level of significance.

Looking at column 2 of the table, we note that the probability of inclusion into the sample for individuals with dementia or other mental illnesses is approximately 18% lower than that for the frail elderly, holding other factors constant. We also note that the average marginal effect of comorbidity status is -0.066; implying that a particular individual with three or more comorbid conditions has a lower probability of being included in the sample than another individual with fewer comorbid conditions by about 0.7%, hold-ing other factors constant. The results further show that there is a general decrease in the probability of sample selection over time by 0.8% since the average marginal effect of the trend variable is -0.008 and statistically significant at 5% level of significance; the

Variable	Reduced form model	Sample selection model	
	(Telecare = 1)	(Inclusion into the sample $= 1$)	
	(1)	(2)	
Age	0.006	0.002	
-	(5.24)	(1.58)	
Square of age	-0.0001	-0.00003	
	(5.58)	(3.36)	
Sex	-0.008	0.009	
	(1.24)	(1.54)	
Area of residence	0.011	-0.009	
	(0.97)	(0.89)	
Dementia and Mental Health	0.042	-0.183	
	(5.81)	(25.76)	
Learning disability	0.032	-0.004	
	(2.78)	(0.38)	
Physical disability	0.001	-0.043	
	(0.10)	(5.01)	
Comorbidity status	-0.005	-0.066	
	(0.83)	(11.99)	
SIMD-decile	-2.033	-0.636	
	(9.77)	(2.89)	
Proportion of telecare users	1.219	-0.681	
	(27.00)	(17.98)	
Trend	0.0001	-0.008	
	(0.72)	(51.25)	
Number of observations	48,571	48,571	
Number of homecare clients	25,598	25,598	

Table 3: The first stage models, Z statistics in parenthesis

The table contains the estimated average marginal effects and the Z-statistics. The Z-statistics have been clustered by Project-ID. Z-statistics greater than or equal to 1.96 imply statistical significance at 5% level of significance. The reference category for client group is 'Frail elderly'.

higher the area level deprivation, the lower the probability of sample selection, other factors held constant, and the higher the proportion of telecare users in a particular local council area, the lower the probability that an individual residing in that local council area was included in the sample, all things equal.

Table 4 on page 13 contains the exponentiated coefficients i.e. the Incidence Rate Ratios of the three variants of the Negative Binomial Models discussed in the previous

Variable			
	(1)	(2)	(3)
Age	1.042	1.048	1.054
	(5.76)	(6.01)	(4.93)
Square of age	0.999	0.999	0.999
	(7.33)	(7.73)	(6.39)
Comorbidity status	2.639	2.437	2.090
	(10.48)	(10.45)	(6.65)
Dementia and Mental Health	0.510	0.483	0.420
	(11.17)	(11.65)	(10.63)
Learning disability	1.699	1.484	2.091
	(6.88)	(4.63)	(6.65)
Physical disability	0.964	0.933	0.970
	(0.78)	(1.37)	(0.45)
Telecare use	0.568	0.620	0.488
	(4.27)	(3.18)	(3.87)
Sex	0.887	0.915	0.863
	(3.64)	(2.53)	(3.09)
Area of residence	1.245	1.244	1.375
	(4.31)	(4.04)	(4.29)
Trend	1.035	1.036	1.038
	(8.09)	(8.03)	(7.26)
Residuals of the telecare use model	0.013	0.011	0.04
	(11.13)	(10.22)	(10.22)
Residuals of the sample selection model	0.003	0.002	0.001
	(11.13)	(11.16)	(11.15)
Telecare interacted with residuals	38.362	19.560	88.104
	(3.76)	(2.71)	(3.41)
	024.41		
wald test for weak instruments: $y^2(n - ualua)$	824.41		
instruments, χ (p – barre)	(0.00)		
Likelihood ratio test for $\alpha = 0$ or $\log_{10} \alpha = -\infty (n - value)$	$\alpha = 1.142$	$\alpha = 1.512$	$\log \alpha = 22.910$
$u = 0$ of log $u = -\infty$ ($p = vanue$)	(0.00)	(0.00)	(0.00)
Likelihood ratio test for	2.75		
Instrument validity; $\chi^2(p - value)$	(0.2534)		
Wald Chi–square test; $\chi^2(p - value)$	411.78 (0.00)	391.38 (0.00)	47.03 (0.00)
Number of observations	15,157	21,511	15,157
Number of homecare clients	8,590	10,586	8,590

Table 4: The second stage models, Z statistics in parenthesis

The table contains the Incidence Rate Ratios for the three Negative Binomial Models. The Z statistics have been clustered by Project-ID. Z statistics greater than or equal to 1.96 imply that the Incidence Rate Ratio is statistically significant at 5% level of significance. The reference category for client group is 'Frail elderly'.

section. The first column of the table, labelled Column 1, contains the Incidence Rate Ratios for a basic Negative Binomial Model. The second column, labelled Column 2, contains the Incidence Rate Ratios for a zero-inflated Negative Binomial Model. The third column, labelled Column 3, contains the Incidence Rate Ratios for a zero-truncated Ne-gative Binomial Model².

In order to validate the models, we conduct four diagnostic tests. First, we test for instrument validity using a likelihood ratio test proposed by King (1980)³. The test compares two versions of the basic Negative Binomial Model: the model presented in Column 1 of Table 4 and the same model but with one of the instrumental variables as an additional regressor. Looking at the results of the test ($\chi^2 = 2.75$, p-value=0.2534), we note that the instrumental variables are valid since the null hypothesis is not rejected at 5% level of significance. Second, we test for weak instruments in the first stage models by conducting a joint significance test on their estimated coefficients. The results of this test show that the instrumental variables are not weak since the estimated coefficients are significantly different from zero ($\chi^2 = 824.41$, p-value=0.0000). Third, we conduct a likelihood ratio test on the dispersion parameter, α , to determine whether the Negative Binomial Model fits better than the Poisson model (Lloyd, 2007). The results of this test confirm the appropriateness of the Negative Binomial Model since the dispersion parameter is greater than one and statistically significant across all versions of the model. Fourth, we conduct a joint significance test on the estimated coefficients of the second stage models using the Wald Chi-square test (Greene, 2002; Cameron and Trivedi, 2005; Wooldridge, 2010). The results of this test show that the explanatory variables included in Table 4 help in explaining the observed variation in the length of stay in hospital.

Looking at the results presented in the Table, we can observe that the Incidence Rate Ratios are robust across the three models in the sense that the effects of the covariates remain unchanged. These results are interpreted as follows:

1. A one year increase in age is associated with an increase in the expected length of stay in hospital by about (4 to 6)%, holding other covariates constant.

² The Incidence Rate Ratios are computed as $e^{\beta} = d$, where *d* is the Incidence Rate Ratio and β is the estimated coefficient. *d* is interpreted as the change in the outcome measure by $(e^{\beta} - 1)\%$ for every unit change in a particular covariate (Winklemann, 2008).

³ The logic behind the test is that if a particular instrumental variable is indeed valid, then the model that includes it as an additional regressor should have a comparatively poor fit. Consequently, a failure to reject the null hypothesis implies that the instrumental variable is valid.

- 2. The expected length of stay in hospital for a female homecare client is approximately (9 to 14)% shorter than that for her male counterparts, controlling for the other covariates in the model.
- 3. Rural residents are more likely to spend a longer time in hospital than their urban counterparts, all things equal.
- 4. Controlling for other independent variables, the expected length of stay in hospital for telecare users is shorter than that for non-users by about (38 to 51)%.
- 5. Individuals with dementia or other mental illnesses are expected to spend a shorter time in hospital than the frail elderly, controlling for other regressors in the model.
- 6. Individuals with learning disabilities have a longer length of stay in hospital, on average, than their counterparts without those disabilities, other variables held constant.
- 7. The expected length of stay in hospital is directly related to comorbidity.
- 8. The models presented in Table 4 suffer from endogeneity, selectivity bias and unobserved heterogeneity since their control terms are statistically significant at 5% level of significance.
- 9. About 4% of the observed variation in the length of stay in hospital is un– explained by the study covariates since the Incidence Rate Ratio of the trend variable ranges from 1.035 to 1.038.

5. Discussion

This paper has presented some preliminary results of a study investigating the effect of telecare on the length of stay in hospital using linked administrative health and social care data in Scotland. We make various assumptions about the probability distribution of the outcome measure and estimate three variants of the Negative Binomial Model to that effect. The results show that, controlling for confounding variables and unobservable factors, telecare users are expected to spend a shorter time in hospital than non-users. This finding is similar to the findings by the previous studies looking into the same issue (see for example, Akematsu, 2012, Akematsu 2013), albeit using a different estimation strategy. Other significant predictors of the length of stay in hospital are age, female gender, rural residence and comorbidity. The results also show that the physically disabled homecare clients have a shorter length of stay in hospital, on average, than the frail elderly, although the estimated coefficient is not statistically significant at 5% level of significance. Future research will involve conducting a sub-group analysis, determining the impact of telecare on admission to hospital and investigating the effectiveness of various telecare devices.

Acknowledgements

I would like to acknowledge NHS National Services Scotland for providing the linked administrative health and social care data for this study, and eDRIS (electronic-Data Research and Innovation Service) for providing technical support while conducting the data analysis in the National Safe Haven.

References

- 1. Akematsu Y., Nitta S., Morita K., Tsuji M. (2013). Empirical analysis of the longterm effects of telecare use in Nishi-aizu town, Fukushima, Prefecture, Japan. *Technology and Health Care*, 21(2), 173-182.
- 2. Akematsu Y., Tsuji M. (2012). Does telecare reduce the number of treatment days? *Studies in Health Technology and Informatics*, 180, 507-511.
- 3. Almashrafi, A., Alsabti, H., Mukaddirov, M., Balan, B., Aylin, P., et al (2016). Factors associated with prolonged length of stay following cardiac surgery in a major referral hospital in Oman: a retrospective observational study. *BMJ Open*, 1-7.
- 4. Andersen, R. (1995). Revisiting the Behavioural Model and Access to Medical care: Does it Matter? *Journal of Health and Social Behaviour*, 36(1), 1-10.
- 5. Atkins, D., Baldwin, S., Zheng, C., Gallop, R., Neighbors, C., et al (2013). A tutorial on count regression and zero-altered count models for longitudinal substance use data. *Psychology of Addictive Behaviors*, 27(1), 166-177.
- 6. Atkins, D. and Gallop, R. (2007). Rethinking how family researchers model infrequent outcomes: a tutorial on count regression and zero-inflated models. *Journal of Family Psychology*, 21, 726-735.
- 7. Awiti, J. (2014). A multilevel analysis of prenatal care and birthweight in Kenya. *Health Economics Review*, 4(33), 1-26.
- 8. Bates, D. (2010). Generalised Linear Models. Available at <u>www.math.ust.</u> [Accessed 18th February 2016].
- 9. Cameron, A., Trivedi, P. (2005). *Microeconometrics: Methods and applications*. Cambridge University Press, 573-639.
- 10. Casteli, A., Daidone, S., Jacobs, R., Kasteridis, P., Street, A., et al. (2015). The determinants of costs and length of stay for hip fracture patients. *Plos One*, 1-4.

- 11. Chipeta, M., Ngwira, B., Simonga, C., Kazembe, L., et al (2014). Zero adjusted models with applications to analysing helminthes count data. *Biomed Central*, 7(856), 1-11.
- 12. Clark, R., Inglis, S., McAlister, F., Cleland, J., Stewart, S., et al. (2007). Telemonitoring or structured telephone support programmes for patients with chronic heart failure: systematic review and meta-analysis. *BMJ*, 334, 942-945.
- 13. Duane, P., Lori, E. (2011). Measuring skewness: A forgotten statistic? *Journal of Statistics Education*, 19(2), 1-18.
- 14. Garen, J.(1984). The Returns to Schooling: A Selectivity Bias Approach with a Continuous Choice Variable. *Econometrica*, 52(5), 1199-1218.
- 15. Giordano, R., Clark, M. and Goodwin, N. (2011). *Perspectives on telehealth and telecare.* WSD Action Network. The Kings Fund.
- 16. Greene, W. (2002). *Econometric analysis*. Fifth edition. Prentice Hall Publishers.
- 17. Greene, W. (2012). *Econometric analysis*. Seventh edition. Pearson Education, pp. 810-822.
- 18. Grogger, J., Carson, R. (1991). Models for truncated counts. *Journal of Applied Econometrics*, 6, 225-238.
- 19. Karazsia, B. and VanDulmen, M. (2008). Regression models for count data: illustrations using longitudinal predictors of childhood injury. *Journal of Pediatric Psychology*, 33(10), 1076-1084.
- 20. King, M.(1980). An Econometric Model of Tenure Choice and Demand for Housing as a Joint Decision. Discussion paper series A. No. 248. Department of Economics, University of Birmingham.
- 21. Lee, L. (1980). Generalized Eonometric Models with Selectivity. *Center for Econometrics and Decision Sciences, University of Florida*, 81, 1-22.
- 22. Long, J. and Freese, J. (2006). *Regression models for categorical dependent variables using Stata.* Second edition. College Station, TX: Stata press.
- 23. Long, J. (1997). *Regression models for Categorical and Limited Dependent Variables: Advanced Quantitative Techniques in the Social Sciences.* SAGE Publications.
- 24. Lyod, S. (2007). Maximum Likelihood Estimation of the Negative Binomial Dispersion Parameter for Highly overdispersed data with applications to infectious diseases. *Plos One*, 2(2), 1-8.
- 25. McCullagh, P. and Nelder, J. (1984). Generalized Linear Models. *The Annals of Statistics*, 12(4), 1589-1596.
- 26. Steel., D. Cylus., J. (2012) United Kingdom (Scotland): Health system review. *Health Systems in Transition*, 14(9): 1-150.
- 27. Steventon, A., Bardsley, M., Billings, J., Dixon, J., Doll, H., Hirani, S., et al. (2013). Effects of telecare on use of health and social services; findings from the

whole systems demonstrator cluster randomized trial. Age and Ageing, 0, 1-8.

- 28. Terza, J., Basu, A., Rahouz, P. (2008). Two-stage residual inclusion estimation: addressing endogeneity in Health Econometric Modelling. *Journal of Health Economics*, 27(3),531-543.
- 29. The Scottish Government. (2010). *The Self-Directed support: A National Strategy for Scotland*. The Scottish Government.
- 30. The Scottish Government. (2015). *Policy and legislation:* 2020 *vision*. Available at http://www.gov.scot/Topics/health/2020-vision [Accessed 4th February 2016].
- 31. Turner, H. (2008). *Introduction to Generalized Linear Models*. National Centre for Research Methods, UK and department of Statistics: University of Warwick.
- 32. Winklemann, R. (2008). *Econometric Analysis of Count Data*. Fifth edition. Springer.
- 33. Wooldridge, J. (2002). *Econometric analysis of cross-section and panel data*. MIT Press.
- 34. Wooldridge, J. (2010). *Econometric analysis of cross-section and panel data*. MIT Press.
- 35. Zou, G. (2004). A Modified Poisson Regression Approach to Prospective Studies with Binary Data. *American Journal of Epidemiology*, 159(7), 702-706.

Highlights

- 1. The results of our econometric models of interest show that telecare users are expected to spend a shorter time in hospital than their counterparts who are non-users, other factors held constant.
- 2. Individuals with learning disabilities are expected to spend a longer time in hospital, on average, than their counterparts who are not learning disabled, holding other factors constant.
- 3. The estimated treatment effect is similar to the findings of the previous studies that used the Propensity Score Matching Technique.
- 4. The estimated coefficients are robust across the three model formulations.