REVSTAT – Statistical Journal
Volume 7, Number 1, April 2009, 55–66

AN INSURANCE TYPE MODEL FOR THE HEALTH COST OF COLD HOUSING: AN APPLICATION OF GAMLSS

Authors: Robert Gilchrist
– STORM, London Metropolitan University,
Holloway Road, London, N7 8DB, U.K.
r.gilchrist@londonmet.ac.uk

Alim Kamara
– STORM, London Metropolitan University,
Holloway Road, London, N7 8DB, U.K.
alimskay@yahoo.co.uk

Janet Rudge
– HRSJ, London Metropolitan University,
Holloway Road, London, N7 8DB, U.K.
janetrudge111@googlemail.com

Abstract:

- This paper introduces a substantive problem, namely the link between fuel poverty and excess winter morbidity amongst older people, and shows how the GAMLSS suite of programs (`www.gamlss.com`) can be used to provide a very flexible method of modelling both the number of hospital admissions and the corresponding lengths of stay in hospital. The approach is closely related to the models that have been used to model the number of insurance claims, and their cost (see Heller *et al.* (2007)). We fit the Beta Binomial distribution and a variety of continuous distributions.

Key-Words:

- *Beta Binomial; GAMLSS; insurance; morbidity; costs.*

AMS Subject Classification:

- 62J12, 62P12.

1. INTRODUCTION

Fuel poverty is defined as "the inability to afford adequate warmth in the home" and is related to poor energy efficiency of homes as well as householders' incomes. In the U.K. a household is defined as suffering from fuel poverty if more than 10% of their income is spent on fuel. In 2008, 4.5 millions people were defined as "fuel poor". It is projected that 5.8 million will be "fuel poor" in 2009. Older households are the group most vulnerable to fuel poverty, and are also particularly susceptible to cold-related health effects. The significant numbers recognised as fuel poor have as yet unrecognised implications for costs to public services.

Conventionally, research has referred to effects of cold homes in terms of excess winter deaths (e.g. Wilkinson *et al.* (2001)). These deaths are known to be associated with outdoor winter temperatures, but direct evidence of links to low indoor temperatures is limited. Mortality statistics disguise the full extent of potentially long-term chronic conditions exacerbated by cold. Hence we have concentrated on measuring excess winter morbidity (illness) in relation to fuel poverty, rather than mortality, because of the consequent implications for winter pressures on health services.

We have previously demonstrated links between fuel poverty risk and excess winter hospital episodes among older people in Newham, using this excess as a measure of associated morbidity (e.g. Rudge and Gilchrist (2007)). In this paper we refer to that work and describe the means by which this measure could be developed as a costing element for a health impact assessment tool. The results could contribute to the debate regarding the case for increased energy efficiency investment on public health grounds, in addition to the accepted environmental grounds. Our methodology is closely related to the models that have been used for insurance claims. The two stage aspect of our modelling, firstly estimating the probability of being an emergency respiratory admission, followed by estimating the probability of dying after admission, gives a model for the effect of fuel poverty on this form of mortality.

2. SUBSTANTIVE BACKGROUND TO FUEL POVERTY AND POOR HEALTH

The U.K. Department of Health (2001) recognises that fuel poverty affects health inequalities, particularly among older people. The potential benefits of energy efficiency investment for older fuel poor households involve improvements in comfort, health and well being. Identifying cost savings associated with such benefits is complicated by the many confounding factors involved in showing direct causal links between housing characteristics and health.

There are no current precise methods of calculating the cost to the health services of cold-related disease arising from poor housing. The newly prevailing emphasis on dealing with climate change and carbon emissions may deflect attention from the needs of the fuel poor, who cannot afford to use energy extravagantly. Energy-saving targets tend to skew energy-efficiency investment in favour of fuel-rich households. However, public health implications demand that such investment should also be health-driven.

3. DATA AND STATISTICAL METHODOLOGY

The main source of data here considered is our existing database for Newham hospital admissions over 1993-96. These data are anonymised with respect to individuals, having been provided at enumeration district (ED) level.

Our previous work examined the excess morbidity for different ages and genders in terms of a range of explanatory variables. We now propose extending this work by analysing daily episodes by length of stay and investigating the associated costs for such episodes. Our proposed methodology is based upon the modelling of the propensity for an individual to be an emergency respiratory hospital admission, together with the duration of stay in hospital for such admissions. This approach is similar to that used for insurance claims (see e.g. Heller, *et al.* (2007)) in which the probability of a claim and the size of a claim are both modelled. Having modelled the probability of being a hospital respiratory admission and the length of the consequential stays in hospital, we could use data on the average cost of such hospital admissions, adjusted for the duration of stay, to give a model for the cost of the Newham admissions. The effect of FPR will be determined by considering the excess cost in winter over that in summer.

Our methodology utilises the R-based GAMLSS package (see Rigby and Stasinopoulos (2005) and `www.gamlss.com`). GAMLSS is a suite of programs written in R (see `www.r-project.org`). We consider the probability of being admitted as following a Beta Binomial distribution, this being a more flexible extension of the more traditional Binomial distribution. Our "default" approach is to use logistic regression. We also noted that some of the elderly people die after admission to hospital. We model this probability of death also using the Beta Binomial. The corresponding length of episode is modelled from a selection of continuous distributions.

The GAMLSS package allow us easily to find the the maximum likelihood estimates of the several parameters of a wide range of distributions and to incorporate random effects and smoothing terms. We can make use of the many facilities of R, such as automatic model selection, and we can easily access the wide range of diagnostics available in R. Up to 4-distributional parameters can be modelled in terms of the risk factors. The potential risk factors are shown in the accompanying Table 1. We utilise nominal factors ED, gender and age to allow differing parameters to be fitted for the differing numbers of "at risk" males and females, of differing ages, in each enumeration district. The definition of the fuel poverty index FPR is discussed further below. Potential confounding factors are considered, using 1991 Census data, including pensioners with limiting long term illness and ethnic composition. Daily weather data were obtained for 1993–1997.

Table 1: Explanatory variables and factors. ** SAP35 is energy rating, or measure of energy efficiency, on a scale of 0–100, where 0 is poorest. # denotes component of FPR.

Variable	**Description**
hh1#	% households with one or more pensioner(s)
hh2	% small households (one or two persons households)
undoc#	% households under-occupied (1 person with >4 rooms; 2 persons with >5 rooms)
lowsap#	% dwellings with poor energy efficiency (below SAP35**)
ctb#	% households in receipt of Council Tax Benefit (indicator of low income)
tow	Townsend deprivation score
ch	% households with no central heating
pens	% lone pensioner households with no central heating
pre	% dwellings built before 1945
pensm	Total male pensioners as % of total population
pensf	Total female pensioners as % of total population
penswh	% of white pensioners in the ED
FPR	Fuel Poverty Risk Index = $(\text{hh1} * \text{undoc} * \text{lowsap} * \text{ctb}) * 10^{-3}$
pwh	White pensioners (% total pensioners)
mmeant	Monthly mean air temperature, °C
mmaxt	Monthly maximum air temperature, °C
mmint	Monthly minimum air temperature, °C
mrain	Monthly rainfall totals, mm
msun	Monthly sunshine hours
mmwd	Monthly mean wind speed
msol	Monthly solar radiation, W hr m-2
mtdif	Difference from previous month mean temperatures, °C
dwigs	Total number of dwellings
house	Total number of households
pop	% population 65 years old or more
age	(1) 65–74, (2) 75–84, (3) 85+
nage	Age with 2 levels only: (1) 65–84, (2) 85+
sex	(1) Male, (2) Female
q	Season factor with 3 levels: (1) "Summer", (2) November, January, February, (3) December
nq	Season factor with 2 levels: (1) Not December, (2) December
z	Factor with 48 levels denoting month
E	Factor with 450 levels specifying enumeration district (ED)

The lagged influence of weather is considered, together with maximum, minimum and mean monthly temperature and average monthly rainfall, wind speed, hours of sunshine, and solar radiation levels.

The chance of a repeat admission of an individual appears to be low, although this is not easy to determine precisely as the original data predated inclusion of patient identifier codes. Hence, although an assumption of independence of the observed admissions and of the observed lengths of episode is not too unreasonable, some correlation between occurrences might be expected, and perhaps some correlation between lengths of episode. GAMLSS makes it possible to incorporate a possible random effect in our linear predictor to allow for over-dispersion caused by the unknown correlation. Moreover, the Beta Binomial distribution is itself a form of over dispersed Binomial distribution; i.e. a Binomial distribution with a Beta distributed random effect.

4. DEFINITION OF THE FUEL POVERTY RISK INDEX

Our population-based study of the London Borough of Newham involved creating a Fuel Poverty Risk Index (FPR), derived from known risk factors, to compare with a cold-related health indicator, based on excess winter emergency respiratory hospital admissions (see Rudge and Gilchrist (2005)). Our data level was limited to small areas, rather than individuals, for patient anonymity reasons.

Datasets were collated for enumeration districts (EDs), which contain, on average, about 220 households: we collected data on household age, size and tenure from the 1991 Census; Council Tax Benefit (CTB) for 1998; estimated energy ratings for dwellings, based on classification by tenure (census data), size and type (from a drive round survey and census) and building age; numbers of emergency episodes for all respiratory diagnoses for patients aged above 64 years for August '93 to July '97 from Hospital Episode Statistics (HES). (Emergency admissions are more likely to reflect seasonal effects than elective admissions.)

The FPR was calculated for EDs as a product of the following (unweighted) factors, all as percentages of total households or total dwellings:

- households with one or more pensioners;
- households in receipt of CTB (indicating low income);
- dwellings with poor energy efficiency (i.e. below the 1991 national average energy rating);
- under occupancy (small households occupying relatively large homes).

5. A STATISTICAL MODEL FOR THE EXPECTED TOTAL DURATION OF EMERGENCY RESPIRATORY HOSPITAL ADMISSIONS

We here develop a model to explain the observed illness counts in each ED, in each month, in terms of the potential explanatory variables, and notably FPR. We model the counts for males and females, and for the three age categories. We consider data for each of 48 months. Our particular interest is in the difference between the counts observed in summer and winter, and whether we can explain this difference in terms of the explanatory variables. To examine this we model the probability p_{ijkl} of an individual of gender i, in age group j, in ED k, being ill in month l, $i=1,2,;\ j=1,..,3;\ k=1,..,450,\ l=1,..,48$.

Our count data consists of the number of people who are ill in a given month, as a proportion of the total number at risk. Perhaps the most natural model for such data is the Binomial distribution, with the observed counts restricted by a “Binomial Denominator”. We here use a logistic Beta-Binomial assumption which can allow for potential “over-dispersion” in our counts.

Thus we assume we have observed numbers Y_{ijkl} of emergency respiratory admissions of gender i, age j, in ED k, in month l, $i=1,2;\ j=1,2,3;\ k=1,..,450;\ l=1,..,48$. The number of people at risk in each “cell” is n_{ijkl}. We assume that Y_{ijkl} follows a Beta Binomial distribution, $\mathbf{BB}(n_{ijkl}, p_{ijkl}, \sigma_{ijkl})$. Our basic assumption is that we have a logit link, i.e. $p_{ijlk} = 1/(1+\exp(-\eta_{ijkl}))$, where η_{ijkl} is a linear predictor based upon the explanatory variables in Table 1.

5.1. Distributional assumption

In defining the probability function, we drop the suffices i,j,k,l for clarity of exposition. The probability function of a random variable, Y which follows the Beta Binomial distribution denoted here as $\mathbf{BB}(n,p,\sigma)$, is given by

$$p_Y(y|p,\sigma) \;=\; \frac{\Gamma(n+1)}{\Gamma(y+1)\,\Gamma(n-y+1)}\,\frac{\Gamma\left(\frac{1}{\sigma}\right)\Gamma\left(y+\frac{p}{\sigma}\right)\Gamma\left[n+\frac{(1-p)}{\sigma}-y\right]}{\Gamma\left(n+\frac{1}{\sigma}\right)\Gamma\left(\frac{p}{\sigma}\right)\Gamma\left(\frac{1-p}{\sigma}\right)}$$

for $y=0,1,2,...,n$, where $0<p<1$ and $\sigma>0$ (and n is a known positive integer). Note that $E(Y)=np=\mu$, say, and $\mathrm{Var}(Y) = np\,(1-p)\left[1+\frac{\sigma}{1+\sigma}\,(n-1)\right]$. Here σ may be viewed as a random effect parameter.

For our modelling we have a r.v. Y_{ijkl}, where we model p_{ijlk} and σ_{ijkl} in terms of our explanatory variables and factors. We assume that the duration d_{ijkl} of observed stays of patients for cell i, j, k, l are such that $d_{ijkl} \sim D(\psi_{ijkl}, \alpha_{ijkl}, \beta_{ijkl}, \gamma_{ijkl})$ where D is one of the many 4-parameter distributions available in GAMLSS. (We here restrict ourselves to distributions with a closed form for the mean and variance, as this is more convenient for derivation of the expectation and variance of the cost to the NHS of fuel poverty). Our default approach is to assume a log link, i.e. $E(d_{ijkl}) = \psi_{ijkl} = \exp(\zeta_{ijkl})$, where ζ_{ijkl} is a linear predictor based on the explanatory variates in Table 1.

5.2. A model for the probability of dying after admission

A proportion of the emergency respiratory admission die whilst in hospital. We model the probability of dying using the Beta Binomial with logit link, in a similar way to the modelling of the probability of being admitted. The full range of possible covariates and factors was considered.

5.3. Modelling the length of stay in hospital

The length of stay in hospital is different for those who survived and those who died. We model each distribution using a range of continuous densities available in GAMLSS, such as the Gamma, Generalised (3 parameter) Gamma, Inverse Gaussian and a Generalised (3 parameter) Inverse Gaussian.

5.4. Model selection strategy

We illustrate our selection strategy for the 2-parameter Beta Binomial. (We extended this naturally for distributions with more parameters). We initially used the step Akaike criterion to select a model for $\mu_{ijkl} = n_{ijkl} * p_{ijkl}$, keeping σ_{ijkl} constant. We then used a step Akaike approach to fit σ_{ijkl}, for the current "best" linear predictor for μ_{ijkl} (with any remaining parameters constant for the more general case). Using the current "best" linear predictors for σ_{ijkl}, the model for μ_{ijkl} was refitted, and so on. We finally removed the terms whose removal was not significant on a χ^2 scale. We combined levels of factors where this did not result in significant deterioration in scaled deviance.

6. RESULTS

6.1. The probability of being an emergency respiratory admission

From the 1991 Census, there were about 25,000 people in Newham over 64 years old. The total count of emergency respiratory episodes amongst this age group was 3378 (over 4 years), 16% of which ended in death. Respiratory episodes far outnumber those for other possible cold-related diagnoses.

We fitted Beta Binomial models to explain morbidity counts in terms of the wide range of explanatory variables, removing any that were not statistically significant. We attempted to avoid a so-called ecological fallacy by using a wide range of explanatory variables. Investigation of the monthly data for 450 EDs determined that "winter" was better defined as November–February, rather than the traditional UK use of December–March.

The accompanying Table 2 shows our "best" model, using a logit link, for the probability of being an emergency respiratory hospital admission, and a log link for the σ coefficient. The linear predictor for p_{ijkl} has an interaction between "season" and FPR, showing that morbidity counts rise with increasing fuel poverty risk index in "winter", with a notably large effect in December. This is over and above the underlying effect of winter itself, irrespective of FPR. Effects are evident for age, with higher counts for older people, and sex, with lower counts for women.

Table 2: Best fitting BB model. Logit link for p, using log link for σ. age2 represents age level 2, etc., mmaxtt[z] denotes mmaxtt indexed over months z, hh1[E] denotes hh1 indexed over enumeration districts E, etc.

Variable	Estimate	Std. Error	t value	Pr(> $\|t\|$)
(Intercept)	−6.802 e+00	1.724 e−01	−39.462	0.000 e+00
age2	6.971 e−01	3.966 e−02	17.574	4.958 e−69
age3	1.803 e+00	5.004 e−02	36.022	2.701 e−282
sex2	−7.115 e−01	3.582 e−02	−19.861	1.311 e−87
q2	1.878 e−01	5.790 e−02	3.244	1.179 e−03
q3	4.647 e−01	8.970 e−02	5.180	2.223 e−07
mmaxt[z]	−1.245 e−02	5.162 e−03	−2.413	1.584 e−02
hh2[E]	1.019 e−02	1.976 e−03	5.157	2.521 e−07
lowsap[E]	−1.996 e−03	9.576 e−04	−2.085	3.711 e−02
ctb[E]	8.214 e−03	1.163 e−03	7.062	1.655 e−12
ch[E]	2.602 e−02	2.883 e−03	9.026	1.807 e−19
pens[E]	−4.041 e−02	4.919 e−03	−8.215	2.148 e−16
pre[E]	−2.538 e−03	8.277 e−04	−3.067	2.164 e−03
fpr06[E]	−5.392 e−05	4.175 e−05	−1.291	1.966 e−01
fpr06[E]:nq2	1.692 e−04	7.009 e−05	2.414	1.578 e−02

There was a strong month effect. To understand this further, we considered monthly weather-related factors. Of all these, maximum temperature was most significant, with a higher maximum leading to lower morbidity counts. Having allowed for the maximum temperature effect, other weather related variables were not significant. The log link proved most convenient and as before, we considered all possible covariates and factors.

The σ coefficient (a random effect) depends only upon the age of the people and their gender; see Table 3. (The linear predictor of the σ coefficient is always negative; it is larger for the over 84 year olds than for the over 64 year olds, and is larger for men than women. As a log link is used, the actual value of σ is calculated from the exponential of the linear predictor).

Table 3: The linear predictor for the sigma parameter in the best fitting BB model for the probability of admission (log link). nage2 represents the 85+ years old.

Variable	Estimate	Std. Error	t value	$\Pr(>\|t\|)$
(Intercept)	-4.567	1.362 e−01	-33.517	6.690 e−245
nage2	3.315	1.777 e−01	18.654	1.600 e−77
sex2	-1.175	1.777 e−01	-6.612	3.800 e−11

6.2. Modelling the probability of a patient dying after admission

We fitted the Beta Binomial with logit link and found that only age was significant in explaining the probability π of death of admitted emergency respiratory patients. Older patients were more likely to die; the probabilities of dying were 0.12, 0.16 and 0.24, respectively, for ages 65–74, 75–84 and 85+. The random effect parameter, which we will call λ, did not depend upon the covariates (its value being 1.76).

Table 4: Best fitting Beta Binomial model for probability π of an individual's death, using a logit link for π and log link for λ. The best fit λ is constant. age2 represents age level 2, etc.

Variable	Estimate	Std. Error	t value	$\Pr(>\|t\|)$
(Intercept)	-1.944	9.710 e−02	-20.02	2.42 e−84
age2	0.263	1.154 e−01	2.278	2.28 e−02
age3	0.780	1.227 e−01	6.355	2.36 e−10

6.3. Modelling the length of stay (episode)

We found that, for both those who died and those who survived, the Inverse Gaussian distribution with log link gave the best fit (smallest AIC) amongst the distributions we considered. Our best model for the survivors had a linear predictor for the mean depending only upon age (older people staying longer), with the dispersion decreasing with age. For those who died, the length of stay depended upon gender, with women staying longer. The dispersion parameter was a constant.

Table 5: Linear predictor for length of stay of survivors (Inverse Gaussian, log link). age2 represents age level 2, etc.

Variable	Estimate	Std. Error	t value	Pr $(>\|t\|)$
(Intercept)	1.965	4.30 e−02	45.790	0.000 e+00
age2	0.233	6.10 e−02	3.830	1.32 e−04
age3	0.530	8.52 e−02	6.230	5.49 e−10

Table 6: Linear predictor for length of stay of those who died in hospital (Inverse Gaussian, log link). S2 represents female.

Variable	Estimate	Std. Error	t value	Pr $(>\|t\|)$
(Intercept)	2.547	1.31 e−01	19.46	0.000 e+00
S2	0.672	2.36 e−01	3.830	1.32 e−04

7. CONCLUSION

We model both the propensity to be ill and the probability of survival after hospital emergency respiratory admission by the Beta Binomial distribution. We model the subsequent probability of dying in hospital and the length of stay in hospital, thereby providing a potential model for the cost of excess winter morbidity attributable to fuel poverty. Our approach is similar to that used in modelling the probability and cost of insurance claims. The GAMLSS software enables us not only to use the Beta Binomial in modelling the probabilities of admission and survival, but also to use a wide range of continuous distributions to model the length of time that a patient stays in hospital. It may be noted that mortality due to fuel poverty is a topical issue in the UK. Our emphasis has been on morbidity. With 16% dying, the determining of a direct relationship between mortality and "fuel poverty" would require substantially more data than our 2835 admissions. However, we could give an estimate of the mortality (after emergency respiratory admission)attributable to FPR by combining the probability of being admitted and the subsequent probability of dying (the latter only depending on age).

REFERENCES

[1] DEPARTMENT OF HEALTH (2001). *National Service Framework for Older People.* At: `http://www.doh.gov.uk/nsf/olderpeople/pdfs/nsfolderpeople.pdf` [Accessed 12/3/03].

[2] HELLER, G.; STASINOPOULOS, D.M.; RIGBY, R.A. and DE JONG, P. (2007). Mean and dispersion modelling for policy claim costs, *Scandinavian Actuarial Journal*, 1–12.

[3] RIGBY, R. and STASINOPOULOS, D.M. (2005). Generalized additive models for location, scale and shape (with Discussion), *Applied Statistics*, **54**, 507–554.

[4] RUDGE, J. and GILCHRIST, R. (2005). Excess winter morbidity among older people at risk of cold homes, *Journal of Public Health*, **27**(4), 353–358.

[5] RUDGE, J. and GILCHRIST, R. (2007). Measuring the health impact of temperatures in dwellings: investigating excess winter morbidity and cold homes in the London Borough of Newham, *Energy and Buildings*, **39**, 847–858.

[6] WILKINSON, P.; LANDON, M.; ARMSTRONG, B. *et al.* (2001). *Cold Comfort: the social and environmental determinants of excess winter death in England, 1986–96*, The Policy Press, Bristol.