Paper submission for the European Population Conference 2008

Alan Marshall (University of Manchester, UK)

The use of relational models to estimate age specific disability prevalence schedules

Abstract (225 words)
The relational methodology involves transforming a standard schedule of rates for a particular demographic charactersitic in order to derive a schedule in another population for which direct data is unavailable or unreliable. This technique, originally developed by Brass (1971), to estimate mortality schedules has also been used to derive fertility and migration schedules (Coale and Trussell, 1996. Congdon, 1993). This paper investigates whether relational methodologies are also a valuable tool for the estimation of disability prevalence, and particularly the generation of age-specific disability profiles for a number of disability types at sub-national levels. Such estimates are required by planners of service provision.

This paper assesses the accuracy of disability rates predicted by relational models and by more conventional individual-level synthetic regression models for regions in England using data from the Health Survey for England and Census 2001. The results of this analysis indicate that relationals model are very comparable in terms of accuracy over all ages compared with the other models. The evidence suggests that relational models are most accurate for the most common disability types such as locomotor and personal care. For less common disability types there is evidence of a need for a more flexible relational rule to alter the shape of the base profile (Limiting Long Term Illness) to a greater extent. Smoothing of LLTI schedules using an appropriate function and a more complex relational rule are explored as possible solutions.

Extended summary (1239 words)
This paper demonstrates the potential of the relational methodology for estimation of disability schedules for subnational areas in the UK. It compares this approach with two more conventional synthetic estimation techniques to demonstrate the accuracy of the relational methodology and to identify and explore some refinements to improve accuracy.

In the UK, data on disability by type/severity is very limited for sub-national areas and is not available for districts (average population 130,000). However, estimates of disability by type, severity and age group are much in demand because of their importance for planning purposes to inform the provision of specialist services, equipment and support (Siegel, 2002).

The Health Survey for England (HSE) provides the most recent detailed source of data on disability with almost complete age coverage. A disability module is included in this survey every five years measuring the prevalence and severity of five disability types based upon questions that assess abilities to perform everyday activities. Data is released for the nine English regions.

The models here generate disability schedules for each region which are then compared with the actual survey results. Although these direct estimates are imprecise they are unbiased allowing an assessment of model accuracy to be obtained. The findings from this comparison are important because they provide vital information on model accuracy that will inform choices on models for sub-regional areas.

The relational approach was first suggested by Brass (1971) to model mortality schedules, and is based on a logit transformation of q(x) the probability of dying before age x (Congdon, 1993).

$$\text{Logit}[q(x)] = \frac{1}{2}\left[\frac{q(x)}{1-q(x)}\right] \qquad [1]$$

As q(x) varies between 0 and 1 then the logit of q(x) takes all values between -∞ and +∞. Any predicted value of the logit of q(x) between -∞ to +∞ will map into a value of q(x) between 0 and 1. If we have the predicted value of the logit of q(x) denoted $\hat{Y}(x)$ then this can be converted to a probability of dying at age x:

$$\hat{q}_x = \frac{\exp(2\hat{Y}(x))}{1+\exp(2\hat{Y}(x))} \qquad [2]$$

Brass proposed a simple relational formula to predict Y(x) from the logit of q(x) in the standard population $Y^s(x)$:

$$\hat{Y}(x) = \beta_0 + \beta_1 * Y^s(x) \qquad [3]$$

This approach is modified in two ways in this research. First, q(x) refers to disability prevalence rates rather than the probability of dying. Second, the Limiting Long Term Illness (LLTI) schedule from the 2001 Census is taken as the standard and this is mapped to schedules for specific disability types at the national level. Relational models are fitted at the national level for males and females separately. The national relational parameters are used to generate regional disability schedules using the regional LLTI schedules as the standard.

The two other more conventional models to which the relational model is compared are forms of individual synthetic regression estimators as described by Skinner (1993). Both involve fitting an individual regression model at national level with HSE data then using these parameters with regional census data to derive regional disability schedules.

The first model uses an ordinary least squares regression model to predict the probability that an individual has a disability. The explanatory variables that are included must satisfy two conditions, they must be measured in the census (with age/sex detail) and apply to the whole age range. On this basis age, sex, ethnic group and LLTI are selected for inclusion in the model. The model parameters are multiplied by age/sex (population) means in each region (census) to derive regional disability profiles. A weakness to note here is that ordinary regression is not recommended to predict a proportion especially where proportions are small because predicted proportions may be negative.

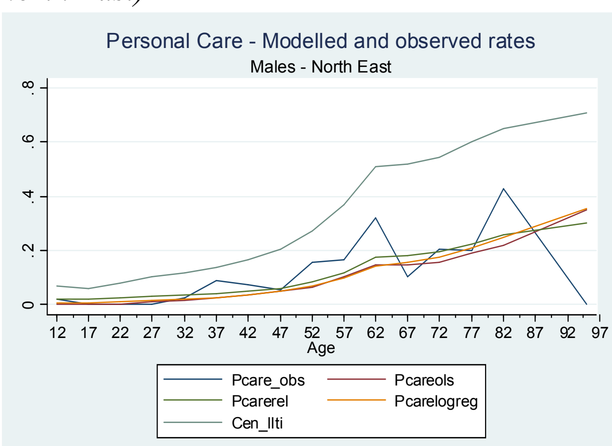
The second model uses a logistic regression model to predict the log odds that an individual has a disability. The explanatory variables are age, age squared, age cubed, sex and LLTI. The model is used to calculate age/sex specific probabilities for the groups with and without an LLTI. These probabilities are combined with counts from the census to derive disability profiles.

Preliminary results comparing root mean square errors across all ages suggest that the relatonal model is more accurate for the most common disabilities, locomotor (mobility), personal care and overall disability (having at least one of the five disability types). For the less common sight and hearing disabilities there is some evidence to suggest that individual synthetic regression estimators may be slightly more accurate than relational estimators.

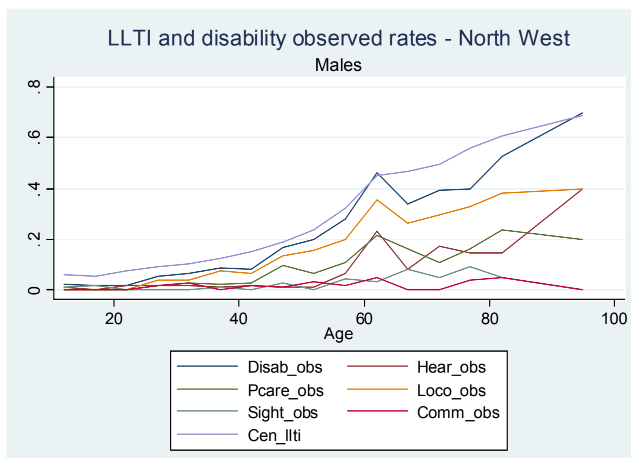
Comparing errors at particular ages reveals a key difference between the models. Relational models preserve unusual features of the LLTI schedule in the disability type schedules to a greater extent than the other models. An example of this is the peak in LLTI at pre-retirement ages and the subsequent slower increase in LLTI once into the retirement ages. This feature is more common in men than women and in industrial and deprived areas.

Graph 1 shows the modelled and observed rates for personal care and the LLTI rates for the North East region. The retirement peak can clearly be seen in the LLTI and relational modelled personal care rates (green line) whilst the schedules generated by the other two models smooth this peak (orange and red lines).

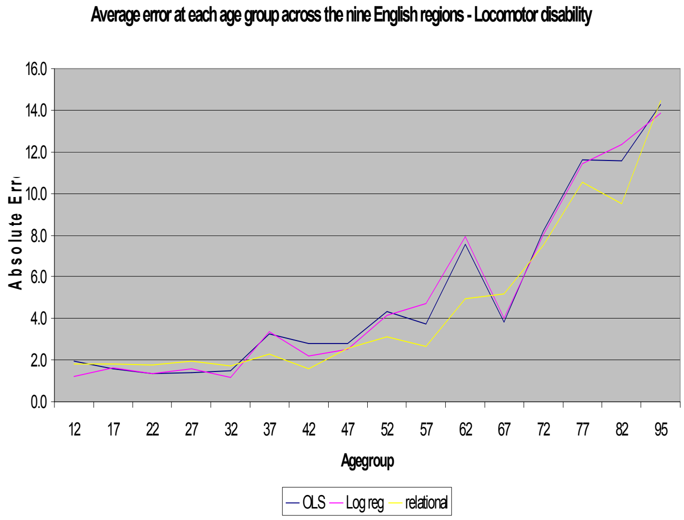
Graph 1 – Modelled and observed prevalence rate for Personal Care Disabilities (Males – North East)



Graph 2: LLTI and disbility type schedules for males in the North West



Where there is evidence that the retirement peak exists in a disability schedule it is desirable to keep this feature. Graph 2 shows LLTI and disability type observed rates for the North West. It clearly shows evidence of a retirement peak in the overall disability, locomotor and personal care and hearing disability schedules. The more rare communication and sight disabilities do not share this feature. The greater accuracy of the relational model, as indicated by the average absolute error at each age group across the nine regions of England, at retirement (as well as other) ages is comfirmed in Graph 3 below.



However, for some disability types and some ages, such as sight disability at the oldest ages, the relational model is less accurate, in this case consistently underestimating prevalence rates. This paper explores two strategies to improve the relational model for more rare disabilities (hearing, sight and communication). First, it is suggested that in these situations the LLTI schedule itself shall be parameterised using a suitable function. This removes unusual features unsuitable to certain disability schedules leaving the general shape and level of LLTI. Second, a more complicated relational rule as proposed by Zaba (1979) is fitted, allowing more deviation from the standard schedule at the oldest (and youngest) ages.

This paper shows that relational methodologies provide a promising approach to the estimation of age specific disability prevalence rates for sub-national areas; and the utility of the relational methodology can be extended beyond its traditional uses of mortality, fertility and migration estimation to other demographic characteristics that are strongly linked to age. These findings are of importance to those estimating schedules for small areas and to policy makers and planners.

References

BRASS, W. (1971) On the scale of mortality in Brass, W. (ed) *Biological Aspects of Demography* New York: Barnes and Noble

CONGDON, P. P. (1993) Statistical graduation in local demographic analysis and projection. *Journal of the Royal Statistical Society. Series A. Statistics in society,* 156, 237-70.

COALE, A. A. & TRUSSELL, J. J. (1996) The development and use of demographic models. *Population Studies,* 50, 469-84.

SIEGEL, J. (2002) *Applied Demography,* London, Academic Press.

ZABA, B. (1979) The Four-Parameter Logit Life Table System. *Population Studies,* 33, 79-100