The importance of age and other variables in predicting paediatric patient flows in New South Wales

JOHN FP BRIDGES AND RALPH M HANSON

John F P Bridges is a PhD Candidate at the City University of New York, Graduate School, New York and is Lecturer, Department of Economics, Case Western Reserve University, and Ralph M Hanson is Director, Information Services at the Children's Hospital, at Westmead.

Abstract

This research focuses upon the relationship between a child's age and the likelihood that the child was treated at a Specialist Children's Hospital rather than at a local hospital. While it is generally regarded that younger patients are more resource intensive, a study was required to determine whether the Specialist Children's Hospitals attracted younger patients. The analysis is based on 42,363 children treated in Greater Metropolitan Sydney in 1996/97, and on separations classified (role delineated) as non-tertiary DRGs (defined as Level 4 activity). However, this activity is of varying degrees of severity. A number of variables were used to explain why a child was treated at either a local hospital or at a specialist children's hospital. This study clearly demonstrates that Specialist Children's Hospitals do attract younger and more severe patients.

Background

The Paediatric Services Review (NSW Health, 1998) in New South Wales undertook a process of identifying the proportion of activity that is suitable for provision by level 4 hospitals (District Paediatrics Units) as opposed to level 6 hospitals (Specialist Teaching Hospitals). This exercise was undertaken to better guide decisions made on flow reversals. To do this, an estimation of the costs of level 4 activity is needed and, more importantly, any cost differentials that may exist across the various levels of hospitals and across different cohorts of patients. The results of the South Australian Paediatric Casemix Costing Study (1996) indicate that the cost of treating younger children is higher than that of treating older children. The purpose of this study was to determine whether Specialist Children's Hospitals did in fact treat a population level 4 patients that were younger than those treated locally. If so, it could be expected that their costs of treatment are higher.

Data

The data is based on 1996/97 inpatient data collection from New South Wales. It includes all children aged 1 month to 14 years 11 months whose separation was deemed within the defined group of level 4 AN-DRG-3. Activity was grouped into two cohorts for the purposes of analysis. Cohort A included areas nominated by the Paediatric Services Review Reference Group as appropriate indicators of level 4 activity. This was because it was believed that these areas were sufficiently distant from the Specialist Teaching Hospitals as to prevent high numbers of self-referrals by parents. Cohort A includes the Illawarra Area Health Service (IAHS) and the Central Coast Area Health Service (CCAHS), both of which lie on the boundary of Greater Metropolitan

Sydney. Cohort B included Areas of sufficient locality to a children's hospital to see high rates of self referral and who had less than 70% self sufficiency in level 4 activity. These include the Wentworth Area Health Service (WAHS), the South-Western Area Health Service (SWSAHS), the Central Sydney Area Health Service (CSAHS) and part of the Western Sydney Area Health Service (WSAHS), excluding that part immediately adjacent to the New Children's Hospital.

Residents treated in either their area of residence or at a Specialist Children's Hospital were included in the analysis. Patients treated elsewhere were excluded.

Preliminary Data and Graphical analysi

Figure 1 plots the percentage of separations treated at the children's hospital by age group for cohort A. From the diagram one can see the outlier in the 12 year old age group. The linear trend line is the OLS estimate for the cohort. Despite its negative slope it is not significant. This OLS equation has an R2 of 16%. If the 12 year old age group is removed the negative slope becomes significant and the OLS equation has an R2 of 50%.

Figure 1: Age and rate of presentation at a Children's Hospital - Cohort A

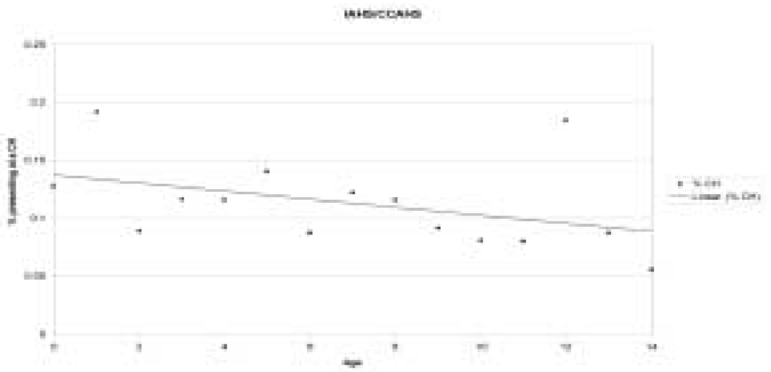
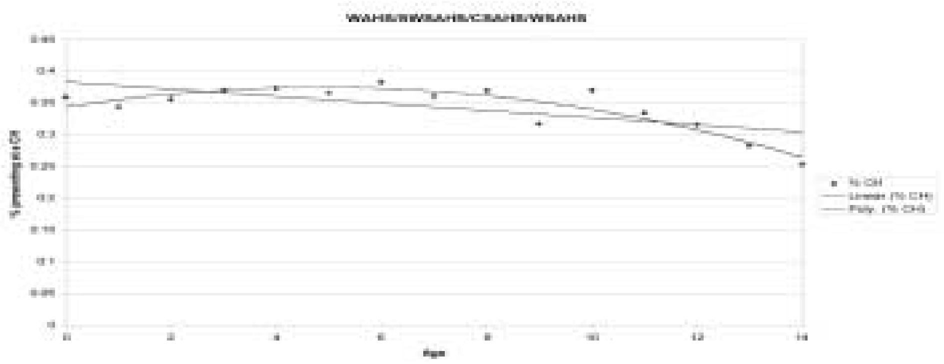


Figure 2 plots the percentage of separations treated at the children's hospital by age group for cohort B. From the diagram one can see that there are no substantial outliers for any age group. The linear trend line is the OLS estimate for the cohort. The negative slope illustrated in figure 4 is significant. This OLS equation has an R2 of 46%. Figure 4 also illustrates a quadratic polynomial trend curve. This quadratic was estimated by Excel as a best fit second order trend.

Figure 2: Age and rate of presentation at a Children's Hospital - Cohort B



Logistic Regression Analysis

Logistic regression is a form of regression analysis where the dependent variable is a binary (0 - 1) variable. For the purposes of the logit regressions presented here, patients treated at Specialist Children's Hospitals were coded as 1 and patients treated at a local hospital were coded as 0. The independent variables are outlined in appendix 1. A stepwise logistic analysis was estimated on the full sample of cohort A and B. The results for these simple regressions can be found in Table 1. These results are with all outliers included.

Table 1. Results of Logit analysis.

	Cohort A Sample = 10,624		Cohort B Sample = 31,739	
	Simple	Interaction	Simple	Interaction
Intercept	0.388	0.025	0.640	0.663
	(0.13)	(0.16)	(0.05)	(0.08)
Age	-0.115	-0.135	-0.044	-0.105
Age2	0.007	0.006		
	(0.00)	(0.00)		
LOS	0.032	0.190	0.057	0.111
	(0.02)	(0.04)	(0.01)	(0.02)
SD	-0.291	0.451	0.706	1.363
	(0.08)	(0.17)	(0.03)	(0.10)
Emerg	-2.845	-3.057	-1.243	-1.438
	(0.10)	(0.14)	(0.03)	(0.06)
Trans	-1.748	-1.532	1.057	1.179
	(0.15)	(0.17)	(0.08)	(0.13)
Surg	-0.615	-0.611		0.245
	(0.08)	(0.08)		(0.10)
CWT	0.261	0.290	0.156	0.186
	(0.10)	(0.10)	(0.04)	(0.04)
Pub	-0.896	-0.595	-0.830	-0.719
	(0.09)	(0.12)	(0.04)	(0.06)
AgeEmerg		0.103		0.051
		(0.02)		(0.01)
AgeLOS				0.007
				(0.00)
AgeDay				-0.023
				(0.01)
AgeTrans				0.042
				(0.02)
AgePub				0.028
				(0.01)
LOSEmerg		-0.182		-0.08
		(0.04)		(0.02)
LOSTrans		-0.162		-0.160
		(0.04)		(0.02)
DayPub		-0.655		-0.527
		(0.18)		(0.09)
SurgPub				-0.364
				(0.10)
Concordant	79.10%	80.20%	72.20%	72.70%

Note: Standard errors in parenthesis.

The simple models for cohort A and B presented in table 1 demonstrate that age is a significant predictor for the likelihood of presentation of level 4 activity at a Children's Hospital. To better understand the marginal effects of age in the models one needs to simulate them. However, it may be beneficial to consider other interaction terms.

Logistic regression with interaction terms

Under the theory of logistic regression there is a case for the inclusion of interaction terms. Interaction terms are to account for the relationship between variables. Using the full set of observations for Cohort A and B a logistic regression was estimated using a number of interaction terms. Stepwise regression was used in gaining an understanding of what variables were significant. Models were estimated including all the appropriate non-interactive terms and are reported in table 1.

Using logistic regression for simulation

With the inclusion of interaction terms into the models a clear understanding of the effect of individual explanatory variables is more difficult to achieve. As stated above, using simulation can better highlight the marginal effects of age. One other benefit of using simulation methods is that they clearly identify the concept of probability that underlies a logistic regression.

Before we display the results of the above regression in simulation form there is a note of caution. Whilst we have aimed at showing a clinically relevant example, the simulation is based on a regression using all data. For example, while Simulation One is based on DRG 124, it is also indicative of any DRG that has the same average characteristics of being mainly a same day procedure on a public patient. If the above methodology is to be used for predicting the flows of a specific DRG, then it is imperative that the regressions be based on that DRG alone.

Figure 3: Simulation One

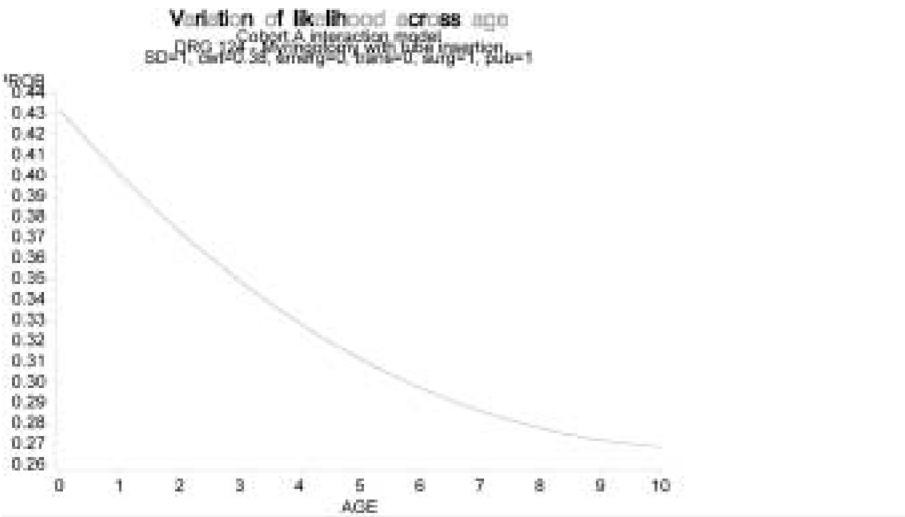


Figure 3 illustrates the simulation of DRG 124, Myringotomy with tube insertion. It is based on the interaction model for cohort A. The simulation has been done over the 0-10 year old range, the clinically relevant age. The simulation graphically demonstrates that Specialist Children's Hospitals are much more likely to treat patients from the lower age cohorts.

Conclusions and recommendations

This paper has shown that age is an important predictor of the likelihood of presentation of Level 4 activity at a Specialist Children's Hospital. Younger patients are more likely to present to a Specialist Children's Hospital, even when adjustments are made for complexity, length of stay and other characteristics of the separations. Thus one can expect that the average cost for treating Level 4 activity in a Specialist Children's Hospital is higher than in local hospitals because of the complicating factor of age (see Bridges & Hanson, 2000). Using a simple graphical and bivariate regression analysis it was necessary to trim outliers to demonstrate the effects of age, however, using logistic regression analysis the results were clearly demonstrated without the removal of outliers. Thus the logistic methods outlined in this paper are more robust to outliers.

There are three possible explanations for this. First, the logit analysis is a non-linear estimation method, and hence its co-efficients are interpreted differently from OLS. Second, the simple bivariate regressions had collapsed the observations into 15 cohort means for the various age groups. For a detailed discussion of this see Cramer (1964) who discusses the collapsing of observations into efficient groupings. Thirdly, the logit models presented here are multi-veriate. Of the three possible explanations it is the third that is most likely to have made the logit regressions more robust to outliers. The relatively low standard errors of the other covariates, besides age, indicate that they are important in predicting patient flows. Bridges & Hanson (2000) demonstrate, in the context of predicting patient cost, that taking a multi-variate approach to regression analysis can lead to significantly different results as compared to individual bivariate analyses. While this paper demonstrates that it is important to focus on individual factors such as age, such analyses need to be done in a multivariate framework.

References

Bridges J & Hanson R 2000, 'Casemix funding for a Specialist Paediatrics Hospital: A hedonic regression approach.' *Australian Health Review*, vol 23, no 3, pp171-5.

Cramer J 1964, 'Efficient grouping, regression and correlation in Engel curve analysis', *Journal of the American Statistical Association*, vol 59, pp233-50.

NSW Health 1998, *Report on Level 4 Paediatric Units, Statewide Services Development Branch,* October, NSW Health Department, Sydney

South Australian Paediatrics Costing Study 1996 *National Paediatric Nursing Study Phase 2,* Paediatric Nursing Study Consortium, Adelaide, pp1-19.

Appendix 1: Explanation of variables

Variable	Description
Age	Age in years
Age2	Age squared
LOS	Length of stay with same day coded as LOS=0
SD	Same day dummy, same day = 1, other =0
Emerg	Emergency department admission, admitted through ED = 1, other = 0
Tran	Transferred admission, admitted through transfer = 1, other = 0
Surg	Surgical patient, surgical AN-DRG-3 = 1, other = 0
CWT	NSW AN-DRG-3 un-discounted cost weight (Flowinfo V 0.2)
Pub	Public/Private patient dummy, private patient = 0, Public =1.
AgeLOS	Age multiplied by LOS
AgeSD	Age multiplied by SD
AgeTrans	Age multiplied by Trans
AgeEmerg	Age multiplied by Emerg
AgePub	Age multiplied by Pub
LOSEmerg	LOS multiplied by Emerg
LOStrans	LOS multiplied by trans
LOSpub	LOS multiplied by pub