Original article

Use of Median Regression to Predict Hospitalization and Pharmaceutical Costs in a Children's Hospital

Simoncelli M, Lebel D, Bussières JF

Mariève Simoncelli, B. Pharm., M.Sc., Health Technology Assessment Unit (UETMIS), CHU Sainte-Justine, Montreal, Qc

Denis Lebel, B.Pharm., M.Sc., F.C.S.H.P., Pharmacy department and Pharmacy Practice Research Unit, CHU Sainte-Justine, Montreal, Qc

Jean-François Bussières, B.Pharm., M.Sc., F.C.S.H.P., Head, Pharmacy department and Pharmacy Practice Research Unit, CHU Sainte-Justine, Clinical professor, Faculty of Pharmacy, Université de Montréal, Montreal, Qc

Correspondence: Jean-François Bussières, CHU Sainte-Justine, Montreal, Qc, Canada, H3T1C5 – 514.345.4603 – jf.bussieres@ssss.gouv.qc.ca

Abstract

Aim: The aim of the present study was to explore the impact of the following variables on both hospitalization and pharmaceutical costs in a children's hospital: patient's weight, major diagnostic category (MDC), length of stay (LOS), an index reflecting the relative use of healthcare resources during hospitalization and clinical severity. **Methods**: The CHU Sainte-Justine is a 500-bed mother-child teaching hospital located in Montreal, Qc, Canada. We included in our analysis all inpatient pediatric episodes of care. All data collated describe the episodes of care for the fiscal year 2005-2006 (from April 1, 2005 to March 31, 2006). Two statistical models were developed to explain pharmaceutical cost and total hospitalization cost. The following independent variables were tested in both models: patient's weight on admission (continuous variable), MDC (discrete variable with a value between 1 and 25, except 12, 13, 14, 15, 20 and 24), LOS, NIRRU and clinical severity (discrete variable with a value of 1 (mild or minor), 2 (moderate), 3 (major or high) or 4 (extreme)) A total of 9 202 episodes met our inclusion criteria. **Results**: The prediction generic formula for both types of cost is as follows: Cost = Intercept + (b1*weight) + (b2*NIRRU) + (b3*LOS) + (clinical severity x) + (MDC x1) + (MDC x2*LOS) + (LOS*weight) + (MDC x3*weight) + (weight*LOS*MDC x3) where b1, b2, b3 are partial regression coefficients and where x indicates the level of the variable (1, 2, 3 or 4 for clinical severity). **Conclusion**: This exploratory study shows the feasibility of building a median regression model. In the model proposed, several factors such as diagnosis, disease severity and length of stay were shown to be significant predictors for both hospitalization and pharmaceutical costs. On the other hand, patient's weight influences only pharmaceutical cost.

Introduction

Every year, the Canadian Institute for Health Information (CIHI) reports a growth in healthcare expenditures. It is recognized that among all categories of healthcare expenditures, drugs is the one accounting for the most drastic raise, which even surpasses the inflation rate [1]. Despite awareness from decision makers, very little is known about the factors influencing the costs related to hospitalizations. Since children's hospitals have been shown to differ from adult hospitals with regards to their budget and spending patterns, our hypothesis is that cost predictors may also present distinctive characteristics for both of these settings.

While healthcare cost predictors were explored by several groups of authors, the main focus of researches published so far has been geared toward cost predictors that were specific to some pathologies or treatments [2]. Likewise, models were developed in order to gain a better understanding of factors explaining cost variability in the context of coronary artery bypass surgery [3], rheumatoid arthritis [4] or intensive care [5]. Another study based on data collected in Brazilian hospitals examined the predictive value of patient characteristics for global hospitalization cost [6]. More specifically, the authors attempted to explain hospitalization cost by using several variables such as socio-demographic data (age and gender) and other factors which are linked to the complexity of care (*e.g.* admission to the intensive care unit (ICU), death during the course of hospitalization).

Among other potential explanatory variables, diagnostic categories have been shown to be a significant cost predictor in most of the studies retrieved from our literature review. First, the association between a classification based on the Adjusted

Clinical Groups (ACG) and the use of healthcare resources was assessed in a Spanish longitudinal study comprising 46 general practitioners and 10 pediatricians [7]. The ACG are based on ICD (international classification of diseases) codes and define a classification of patients according to several categories with respect to their age, gender and all the diagnoses received during their hospital stay. The authors concluded that such clinical classification system is a major predictor of healthcare resources utilization. Moreover, age and gender explained 13% of the variance of medical visits, 24% of the number of prescriptions and 16% of the prescriptions related costs. For pediatric patients, the correlation with age was particularly pronounced [7]. A similar study was conducted in two primary healthcare centres in Sweden to assess the impact of ACG, age and gender on patient related costs [8]. It reached the conclusion that 38% of patient related costs were explained by a weighting factor assigned to the ACG over the two years of the analysis. Only a negligible fraction of these costs (less than 1%) could be explained by patients' age and gender [8]. In addition, ACG was assessed in the context of pediatric care. An analysis based on the electronic charts of 9 659 children from five primary healthcare centres was carried out to explain pharmacy costs variability using ACG [9]. This study reveals that approximately 22% of these costs were explained by ACG and therefore, the author concluded that such classification of hospitalizations represents a major predictor of pharmacy cost. Other variables explored in this model such as age, gender and number of diagnoses have only yielded a minor or no predictive value. The fact that more than half of pharmacy cost variance could not be explained in this study underlines a gap in knowledge with regards to cost predictors [9].

Then, Pirson et al. evaluated the impact of the following variables on healthcare resources utilization: disease severity (defined as minor (1), moderate (2), major (3) or extreme (4)), mortality risk, length of stay, social factors, age, gender, type of admission, destination at discharge, intensive care unit stay, nosocomial infections and rehospitalization in the year following the first admission. Variables that were shown to be significantly correlated with a high healthcare resources utilization in the multivariate analysis were length of stay (LOS), disease severity, the presence of a social factor and an intensive care unit stay [10].

Other authors have focused on the impact of the length of stay on hospitalization cost. In a Spanish study including over 21 800 hospitalizations, length of stay was shown to be a major predictor of direct hospitalization cost variability [11]. On the contrary, Taheri et al. found that when the length of stay was one day shorter, the total cost of medical care was only decreased by an estimated 3%. Indeed, it is known that the first days following an hospital admission are the most costly since the intensity of the resources consumed is higher [12].

In an analysis based on an healthcare database (Healthcare Cost and Utilization Project Kids' Inpatient Database), Merenstein et al. reported that the hospital charges for five common diagnoses (defined by ICD-9 codes) were superior for children's hospitals than for adult hospitals from different American states, even after adjusting for potential confounders [13]. However, no statistically significant difference between these two settings was detected with regards to LOS [13].

The aim of the present study was to explore the impact of the following variables on both hospitalization and pharmaceutical costs in a children's hospital: patient's weight, major diagnostic category (MDC), length of stay (LOS), an index reflecting the relative use of healthcare resources during hospitalization (NIRRU- stands for "niveau d'intensité relative des resources utilisées") and clinical severity.

Methods

Characteristics of the hospital

The CHU Sainte-Justine is a 500-bed mother-child teaching hospital located in Montreal, Qc, Canada. In Canada, the healthcare system is financed by the government. The fact that patients are not billed directly for services used during their hospitalization implies that any case costing initiative would not be straightforward.

Inclusion and exclusion criteria

We included in our analysis all inpatient pediatric episodes of care. Mothers' episodes of care, new babies' episodes of care and any episode of care attributed to patients older than 18 years old were excluded. The following major diagnostic categories (MDCs) were excluded because they were not reflecting the pediatric population: MDC 12 (Diseases and Disorders of the Male Reproductive System), MDC 13 (Diseases and Disorders of the Female Reproductive System) and MDC 14 (Pregnancy, Childbirth and the Puerperium). Since neonates represent a subpopulation with distinctive clinical characteristics, MDC 15 (Newborns and Other Neonates With Conditions Originating in the Perinatal Period) was excluded from the analysis. MDC 0 (undetermined), 20 and 24 were also excluded since they comprised less than 20 episodes of care.

The initial sample comprised 18 118 episodes of care for inpatients. From this number, 9 202 episodes met our inclusion criteria.

Data sources

All data collated describe the episodes of care for the fiscal year 2005-2006 (from April 1, 2005 to March 31, 2006). The episodes of care were selected using the healthcare administrative database Magic® (stands for "Module d'aide à la gestion de l'information clinique"). This database was developed by *Mediamed Technologies* and integrates clinico-administrative data providing from the key management information systems used in some hospitals in Quebec, a Canadian province. One of Magic®'s module was designed to compute all healthcare resources consumed by a patient during a given hospitalization in order to build a total cost which is subdivided by clinical sector [14].

Definition of the variables

Two statistical models were developed. The first model's aim was to explain pharmaceutical cost while the second model's aim was to explain total hospitalization cost. Pharmaceutical cost comprises all costs related to drugs and medical devices as well as human resources linked to the pharmacy department. Hospitalization cost includes nursing services, operating room, emergency room, imaging, laboratories, interventional hemodynamics, pharmacy, radio-oncology as well as all other clinical and administrative units.

The following independent variables were tested in both models: patient's weight on admission (continuous variable), MDC (discrete variable with a value between 1 and 25, except 12, 13, 14, 15, 20 and 24), LOS, NIRRU and clinical severity (discrete variable with a value of 1 (mild or minor), 2 (moderate), 3 (major or high) or 4 (extreme).

Statistical Analysis

Our outcome variables were pharmaceutical and total hospitalization costs. Descriptive statistics such as mean, standard deviation, median and percentiles were used initially. Both response variables were extremely skewed to the left and even after log transformed retained strong left skewness. Therefore, we decided to use median regression analysis as a multivariate model. Independent variables were patient's weight, MDC, LOS, NIRRU and clinical severity. Non-parametric univariate correlation analyses were performed to assess the association between both types of cost and the continuous

predictors (patient's weight, LOS and NIRRU) using Kendal's tau correlation coefficient. Correlations between both types of cost and categorical variables (MDC and clinical severity) were tested using Kruskal Wallis non-parametric test. Significant variables were then used as predictors in a median regression model with cost as a dependant variable (pharmaceutical cost in the first model and hospitalization cost in the second model). P-values less than 0.05 were considered significant. Analyses were performed using SPSS 15.0 and SAS 9.1. The SAS experimental procedure proc Quantreg was used to fit median regression models.

Results

Description of data

Both pharmaceutical and hospitalization costs are extremely skewed. As presented in **Table 1**, 75 % of patients generate costs that are well below the mean pharmaceutical cost of 640.40$ (75th percentile of 303.30$) and mean hospitalization cost of 6727.40$ (75th percentile of 5112.80$). This is further shown by the scatter plot of pharmaceutical cost vs. hospitalization cost (see **Figure 1)**. Since both cost distributions are skewed, there is no tendency and therefore, costly hospitalizations are not necessarily related to high pharmaceutical cost. This skewness required the use of median analysis and non-parametric methods.

Table 1 Distribution of Pharmaceutical and Hospitalization Costs per Patient*

Response variables	N	25th percentile	Median	75th percentile	Mean	SD**
Pharmaceutical Cost	9202	27.64	96.28	303.30	640.40	3634.20
Hospitalization Cost	9200	1308.80	2420.30	5112.80	6727.40	18906.70

*All costs are expressed in Canadian dollars
**Standard deviation

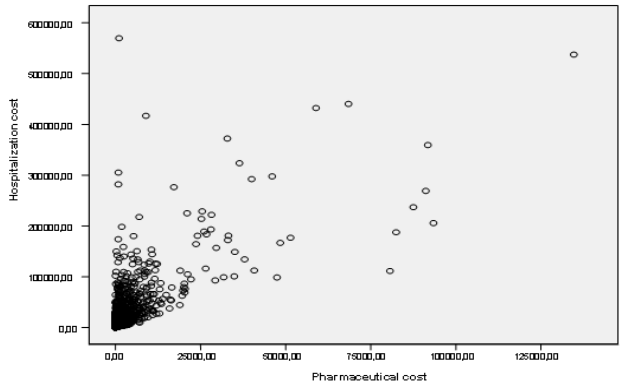

Figure 1 Scatter Plot of Pharmaceutical Cost vs. Hospitalization Cost per Patient

Table 2a provides the descriptive statistics for weight, NIRRU and LOS. **Table 2b** describes pharmaceutical costs and hospitalization costs per patient according to MDC. **Table 2c** provides descriptive statistics for clinical severity.

Non-parametric correlations between continuous predictors (LOS, NIRRU and patient's weight) and the response variables were significant (see **Table 3a**). Kruskal-Wallis tests for association between categorical predictors (clinical severity and MDC) were also significant (see **Table 3b**).

Median Regression Model

After various regression models being fitted, we finalized the model including main effects of weight, NIRRU, LOS, clinical severity, MDC as well as the following interactions terms: "MDC and LOS", "LOS and weight", "MDC and weight" and "weight, LOS and MDC". Every comparison refers to MDC 11. The interaction terms were hypothesized a priori and proved to be significant. This means for instance that the effect of LOS on the both types of cost depends significantly on the diagnosis (MDC).

All main effects were very significant predictors of the pharmaceutical cost. In a similar fashion, the same covariates were significant predictors for hospitalization cost. This is to be expected since pharmaceutical cost is part of hospitalization cost.

The prediction generic formula for both types of cost is as follows:

Cost = Intercept + (b1*weight) + (b2*NIRRU) + (b3*LOS) + (clinical severity$_x$) + (MDC $_{x1}$) + (MDC $_{x2}$*LOS) + (LOS*weight) + (MDC $_{x3}$*weight) + (weight*LOS*MDC $_{x3}$)

*b1, b2, b3: partial regression coefficients

*x: indicates the level of the variable (1, 2, 3 or 4 for clinical severity)

Discussion

There is a growing interest in identifying major cost predictors linked to hospital stays in Canada. Children's hospital stays represent a challenge as their cost depend on numerous explanatory variables. In fact, there are limited published data about the impact of different variables on both hospitalization and pharmaceutical costs in children's hospitals, as patient's weight, major diagnostic category (MDC), length of stay (LOS), an index reflecting the relative use of healthcare resources during hospitalization (NIRRU- stands for "niveau d'intensité relative des resources utilisées") and clinical severity. This exploratory study shows the feasibility of building a median regression approach that was applied by other authors. For instance, Nikolson et al. assessed the association between certain patients' characteristics and hospitalization cost related to premature birth for the state of

Table 2a Descriptive Statistics for Continous Variables

Variables	25th percentile	Median	75th percentile	Mean	SD
Weight (kg)	11.00	20.00	42.00	27.91	21.72
NIRRU*	0.44	0.68	1.19	1.18	1.80
LOS (days)	1.00	2.00	5.00	5.42	9.90

* Index of relative use of healthcare resources

Table 2c Descriptive Stats for Categorical Variable Clinical Severity

Clinical Severity	N	Percent	Pharmaceutical cost per patient			Hospitalization cost per patient		
			25th percentile	Median	75th percentile	25th percentile	Median	75th percentile
1	4565	49,6	0	0	52,32	568,64	964,93	1770,42
2	3024	32,9	0	51,97	117,73	864,63	1363,39	2396,41
3	1228	13,3	33,78	124,22	334,15	1617,95	2799,73	5780,11
4	385	4,2	127,54	448,87	1437,25	3043,27	9108,15	21917,63
Total	9202	100.0						

Table 2b Descriptive Statistics for Categorical Variable Major Diagnostic Category (MDC)

MDC	Description	N	Percent	Pharmaceutical cost per patient			Hospitalization cost per patient		
				25th percentile	median	75th percentile	25th percentile	median	75th percentile
1	Diseases and Disorders of the Nervous System	774	8,4	26.34	91.86	276.70	1178.08	2642.29	6840.69
2	Diseases and Disorders of the Eye	140	1,5	30.60	91.44	194.13	1772.16	2631.32	3646.72
3	Diseases and Disorders of the Ear, Nose, Mouth	950	10,3	26.02	52.67	110.80	1456.88	2238.98	3457.16
4	Diseases and Disorders of the Respiratory System	1380	15,0	26.35	59.28	135.91	815.68	1269.79	2443.59
5	Diseases and Disorders of the Circulatory System	400	4,3	52.22	145.16	567.04	961.97	2451.83	11601.26
6	Diseases and Disorders of the Digestive System	1297	14,1	0	65.79	171.54	1231.49	1885.03	3099.71
7	Diseases and Disorders of the Hepatobiliary System and Pancreas	115	1,2	52.41	246.54	586.05	1870.14	2949.14	8483.47
8	Diseases and Disorders of the Musculoskeletal System And Connective Tissues	823	8,9	26.13	64.45	204.35	2616.75	4734.39	9178.83
9	Diseases and Disorders of the Skin, Subcutaneous Tissue and Breast	230	2,5	33.16	118.95	299.59	1477.44	2276.91	4133.42
10	Endocrine, Nutritional and Metabolic Diseases and Disorders	567	6,2	31.09	125.65	388.83	1978.95	3738.99	10206.53
11	Diseases and Disorders of the Kidney and Urinary Tract	369	4,0	56.79	137.81	257.95	1548.64	2814.36	5083.09
16	Diseases and Disorders of the Blood and Blood-Forming Organs and Immunological Disorders	478	5,2	71.05	251.64	938.11	1620.15	2970.06	8026.63
17	Myeloproliferative Diseases and Disorders and Poorly Differentiated Neoplasms	523	5,7	564.78	1151.84	2817.57	2632.93	4329.35	16401.39
18	Infectious and Parasitic Diseases	289	3,1	61.48	147.28	345.69	1471.56	2192.30	4018.18
19	Mental Diseases and Disorders	187	2,0	26.10	113.81	278.96	2158.17	6849.23	15311.31
21	Injury, Poisoning and Toxic Effects of Drugs	157	1,7	0	36.77	131.29	669.07	1252.78	3476.44
22	Burns	44	,5	132.48	335.64	636.85	1920.23	4048.24	8166.68
23	Factors Infleuncing Health Status and Other Contacts With Health Services	448	4,9	0	0	78.47	490.27	872.06	1815.55
25	Human Immunodeficiency Virus Infections	31	,3	91.77	389.03	3070.38	4291.65	10864.85	60212.90
Total		9202	100,0						

Table 3a Kendall's tau Correlation Coefficients for Continuous Variables

Variables	Pharmaceutical cost	Hospitalization cost
LOS	,418**	,596**
NIRRU	,365**	,540**
Weight (kg)	,122**	,106**

** Correlation is significant at the 0.01 level (2-tailed)

Table 3b Kruskal Wallis Test for Categorical Variables

	Chi-Square	df	P-value
Clinical Severity			
Pharmaceutical cost	2001,448	3	,000
Hospitalization cost	1803,257	3	,000
MDC			
Pharmaceutical cost	1721,261	18	,000
Hospitalization cost	1729,315	18	,000

Maryland [15]. Such statistical analysis was selected because of the skewness of the distribution and to consider the outliers [15].

In the proposed model, several factors such as diagnosis, disease severity and length of stay were shown to be significant predictors for both hospitalization and pharmaceutical costs. On the other hand, our model demonstrates that patient's weight influences only pharmaceutical cost. A regression model (not presented in this article) that did not include the patient's weight showed a lower predictive value than the selected model with regards to pharmaceutical cost (Pearson's correlation coefficient = 0,651 vs. 0,733). However, both models were equivalent for predicting hospitalization cost.

While hospitals are publicly funded in Canada, very few data have been published about the cost structure and the detailed data required to run such models. We believe that a similar study should be replicated in other hospitals in order to identify a more straightforward model that would assist healthcare decision makers in predicting pharmaceutical cost.

The median regression was used because of the extreme skewness of the data. Therefore, the median is a much better estimate for central tendency than the mean. However, the median regression might not be so great at predicting costs for patients who are likely to incur extremely high medical costs (outlying values).

The regression model was validated using cost estimates representing 7 462 episodes of care for the fiscal year 2006-2007. It was populated with real data describing patients' characteristics for the fiscal year 2007-2008. Predicted costs were compared to real costs and correlation coefficients were computed for both hospitalization and pharmaceutical costs. The results demonstrate a good predictive value with a Pearson's correlation coefficient of 0,733 and 0,860 for pharmaceutical cost and hospitalization cost, respectively.

The median regression model proposed has several limitations. Amidst methodological issues, the main source of uncertainty provides from the sample data per se. Indeed, all data used in our analysis were collected using Magic®, an integrator that aggregates data from multiple management information systems across the hospital. Moreover, most of these data were not tested using a rigorous validation process, which brought up the question of their reliability. Another limitation is related to relying on MDC as an independent variable. This could be a source of bias considering that the MDC category to which an episode of care belongs is based solely on the main diagnosis received upon admission. Therefore, the sole use of the MDC does not reflect the complexity of care received during an hospitalization nor all the other secondary diagnoses. However, other indicators such as clinical severity and NIRRU were introduced in the model to give a more realistic overview of the financial burden associated to a given episode of care.

Conclusion

This exploratory study shows the feasibility of building a median regression approach. In the model proposed, several factors such as diagnosis, disease severity and length of stay were shown to be significant predictors for both hospitalization and pharmaceutical costs. On the other hand, patient's weight influences only pharmaceutical cost.

Reviewer: none.

Acknowledgements: The authors wish to thank Lubomir Alexandrov, M.Sc., Biostatistician.

Funding: none.

Conflict of interest: The authors have no conflict of interest regarding this manuscript.

References

1. Appleton KM, Rogers PJ, Ness AR. Is there a role for n-3 long-chain polyunsaturated fatty acids in the regulation of mood and behavious? a review of the evidence to date from epidemiological studies, clinical studies and intervention trials. Nutr Res Rev 2008;21(1):13-41.
2. Polverejan E et al. Estimating mean hospital cost as a function of length of stay and patient characteristics. Health Econ. 2003;12(11):935-47.
3. Ghali WA et al. Identifying pre- and postoperative predictors of cost and length of stay for coronary artery bypass surgery. Am J Med Qual 1999;14(6):248-54.
4. Clarke AE et al. Predicting the short term direct medical costs incurred by patients with rheumatoid arthritis. J Rheumatol 1999;26(5):1068-75.
5. Rapoport J et al. Explaining variability of cost using a severity-of-illness measure for ICU patients. Med Care, 1990;28(4):338-48.
6. Ramiarina R, Almeida RM, Pereira WC. Hospital costs estimation and prediction as a function of patient and admission characteristics. Int J Health Plann Manage 2008;23(4):345-55.
7. Orueta JF et al. Adjusted Clinical Groups (ACGs) explain the utilization of primary care in Spain based on information registered in the medical records: a cross-sectional study. Health Policy 2006; 76(1):38-48.
8. Engstrom SG et al. The importance of comorbidity in analysing patient costs in Swedish primary care. BMC Public Health 2006;6:36.
9. Aguado A. et al. Variability in prescription drug expenditures explained by adjusted clinical groups (ACG) case-mix: a cross-sectional study of patient electronic records in primary care. BMC Health Serv Res 2008;8:53.
10. Pirson M. et al. Analysis of cost outliers within APR-DRGs in a Belgian general hospital: two complementary approaches. Health Policy 2006;76(1):13-25.
11. Cots Reguant F. et al. [Relation of direct hospitalization costs with length of stay]. Gac Sanit 1997;11(6):287-95.
12. Taheri PA, Butz DA, Greenfield LJ. Length of stay has minimal impact on the cost of hospital admission. J Am Coll Surg 2000;191(2):123-30.
13. Merenstein D, Egleston B, Diener-West M. Lengths of stay and costs associated with children's hospitals. Pediatrics 2005;115(4):839-44.
14. MédiaMed Technologies, Products and services- MAGIC 5.0. http://www.mediamedtech.com/produitsEtServices/produit.php?produitID=3 (page accessed on June 8, 2008).
15. Nicholson WK, Frick KD, Powe NR, Economic burden of hospitalizations for preterm labor in the United States. Obstet Gynecol 2000;96(1):95-101.