Archives of Community Medicine & Public Health

Piergentili P[1]*, Simon G[1], Paccagnella O[2], Grassetti L[3], Rizzi L[3] and Samani F[4]

[1]*Azienda Assistenza Sanitaria n. 5, Friuli occidentale, Pordenone, Italy*
[2]*Dipartimento di Scienze Statistiche, Università di Padova, Italy*
[3]*Dipartimento di Scienze Economiche Statistiche, Università di Udine, Italy*
[4]*General Practitioner, Centro di Formazione Regionale per la Medicina Generale, Monfalcone, Italy*

Dates: **Received:** 27 August, 2015; **Accepted:** 22 September, 2015; **Published:** 24 September, 2015

***Corresponding author:** Paolo Piergentili, UOC Accreditamento, contratti, qualità ed esiti, Ulss 13 Mirano, via Mariutto 76, 30035 Mirano (Venezia–Italy), Tel:+39.3204643495; E-mail: piergentili@casemix.org

www.peertechz.com

Keywords: Risk adjustment; Casemix; Primary Health Care; Administrative databases; Health care costs

Research Article

Risk-Adjusted Models of Costs Referable to General Practitioners Based on Administrative Databases in the Friuli Venezia Giulia Region in Northern Italy

Abstract

Objective: To develop risk adjustment models for cost evaluation in primary health care in Italy based on administrative databases.

Setting: The 2007 administrative databases from the National Health Service of the Friuli Venezia Giulia Region were the data source. Data referred to the general population and included information on the use of health services (inpatient, outpatient, medication, home care) as well as on the major chronic health problems. Data included persons who, for their health condition, must not pay the contribution usually required for using health services (ticket exemption).

Design: Multilevel (multivariate) statistical analysis, where the tariff of services or the price of drugs (both summed up and separated) were the dependent variables, and the health conditions and other variables related to the citizens were the predictive variables.

Results: The analysis included 1,067,239 citizens registered with a General Practitioner (GP) and 1,129 GPs. The number of people with at least one ticket exemption was 461,532. A number of predictive models were developed, which considered the sum of all tariffs and prices, tariffs and prices of inpatient and outpatient services, and medications. The models had very robust results. The models explained a considerable share of the variation using the (R^2_1) parameter: the proportional reduction of error for predicting the level-1 dependent variable with respect to the model without any predictors. The R^2_1 was 44.6% for the sum of tariffs and expenditures, 30.9% for hospital tariffs, 27.1% for inpatient services and 49.3% for medications. The intra class correlation coefficient (ICC), which measures the proportion of residual variability due to the second level of analysis (GP), shows that, controlling for the Casemix, the amount of the residual variability driven by the GPs is very low or even negligible: 0.89% for total individual health care tariffs, <0.1% for inpatient services, 1.49% for outpatients services, and 2.0% for drug prescriptions.

Discussion: The health status information provided by the ticket exemption database proved to be valid beyond the Researchers expectations, despite known data quality problems. The large study sample size may have played a role in these results. Primary health care risk adjustment can be performed using administrative databases instead of GP clinical records, making it much easier and less costly. The overall variation explained by the models was in line with the findings in the literature. The residual variation attributable to GPs was unexpectedly low. In the inpatient setting, this may be partly due to a systemic effort to reduce inappropriate hospitalisation. In the other settings this result came as a surprise, and may lead to a reconsideration of GP behaviour variation. Outcome and output measures in PHC without proper risk adjustment may lead to inaccurate findings.

Introduction

A classification scheme of patients is a necessary tool for the evaluation of health services in primary health care (PHC). A classification of the persons who need assistance, rather than their *illness*, provides an instrument that accounts for all a patient's problems, the need for health services she or he expressed, and their evolution of the severity of the condition over time. In Europe, patient classification systems (PCS) are being used primarily to categorise hospital data, and applications in other settings are rare. Yet, it would be of equal or even of greater importance to focus on PHC in order to evaluate quality of care and cost containment on a systemic scale. In fact, hospital care can be viewed as part of the care process, which often begins in and returns to the PHC setting. That is, hospital care is just a part of the health care process. There are several conditions for which hospitalisation is just an alternative to other approaches (home care, clinics, outpatient care, etc.). Thus, only an analysis centred on PHC can correctly evaluate the quality and the cost of care of a given provider, as compared to the care assured by another provider for similar patient's condition. In Italy, as well as in many other countries, especially where a national health system (NHS) exists, General Practitioners (GPs) play a key role in health

care management. The concept of *episode of care* has been elaborated in order to deal with output and outcome comparison, and it has led to the development of some classification schemes and commercial software [1-3]. However, these systems require clinical data of a quality not common in Europe. Additionally, data collection is often complex and expensive, and creates privacy problems.

The Italian NHS provides comprehensive health care coverage to the entire population through geographically-based organizations called local health organizations (ASL in Italian). ASLs offer primary, hospital, and outpatient care (all diagnostic procedures and tests and referrals to specialists) and medications, either directly or through contracted providers. A huge information system has been developed over time to support administration. Information on treatments received by any Italian citizen under the umbrella of the NHS is available to ASLs, regions, and the Ministry of Health. The information is contained in large administrative databases, which include also clinical information. In addition, any citizen receives a unique health identifier, which is always included in any health record. This allows, through a record linkage process, to describe all treatments she / he has received from the NHS.

Given the information available, it is easy to calculate output/outcome indicators. However, there is no risk adjustment model that properly compares GPs, ASL, or health regions, which often assist population with a different Casemix. In the Italian NHS there is a convenient solution to this problem. Even if drugs and outpatient services are free of charge, the patient makes a small contribution (called a *ticket*). However, persons affected by severe conditions and chronic diseases, or low income, are exempted from the ticket. The list of the diseases and conditions that allow ticket exemption, based on ICD9CM classification, is set forth by a national law, and is reported in Table 1. Information on ticket exemptions is collected in administrative databases, including clinical information at the population level. Thus, ticket exemptions can be evaluated as a source of information on which to develop risk adjustment models.

The Region Friuli Venezia Giulia, in north-eastern Italy, has 1.2 million inhabitants, six ASLs, and eighteen hospitals. Complete and accurate records of the whole population and health services have been collected in networked administrative databases since 1970. This paper presents the results of a study that exploits these administrative databases to develop a risk adjustment model for PHC, using ticket exemption as a source of information on the health status of the population.

Material and Methods

Administrative databases concerning hospitalizations, medications, outpatient services, ticket exemptions, and the list of citizens associated with a GP in 2007 were made available by the Regional Health Authority of Friuli Venezia Giulia. All records included the unique health identifier. Children under fourteen were excluded, since they are usually (but not exclusively) registered with a family paediatrician. Also, data on drug costs of nursing home residents (about 6,500 people) was not included, since they receive medications through a different channel.

Table 1: List of conditions giving right to ticket exemption.

Code		Definition
001	253.0	Acromegaly and gigantism
002	394	Diseases of mitral valve
	395	Diseases of aortic valve
	396	Diseases of mitral and aortic valves
	397	Diseases of other endocardial structures
	414	Other chronic ischemic heart disease
	416	Chronic pulmonary heart disease
	417	Other diseases of pulmonary circulation
	424	Other diseases of endocardium
	426	Conduction disorders
	427	Cardiac dysrhythmias
	429.4	Functional disturbances of. cardiac surgery
	433	Occlusion/stenosis of precerebral arteries
	434	Occlusion of cerebral arteries
	437	Other/ill-defined cerebrovascular disease
	440	Atherosclerosis
	441.2	Thoracic aneurysm without rupture
	441.4	Abdominal aneurysm without rupture
	441.7	Thoracoabdominal aneurysm w/o rupture
	441.9	Aortic aneurysm unsp. site w/o rupture
	442	Other aneurysm
	444	Arterial embolism and thrombosis
	447.0	Arteriovenous fistula acquired
	447.1	Stricture of artery
	447.6	Arteritis unspecified
	452	Portal vein thrombosis
	453	Other venous embolism/thrombosis
	459.1	Postphlebetic syndrome
	557.1	Chronic vasc. insufficiency of intestine
	745	Bulbus cordis/ cardiac struct. Anomalies
	746	Other congenital anomalies of heart
	747	Other congenital anomalies circulatory s.
	V42.2	Heart valve replaced by transplant
	V43.3	Heart valve replaced by other means
	V43.4	Blood vessel replaced by other means
	V45.0	Post surgical cardiac device in situ
003	283.0	Autoimmune haemolytic anaemias
004	282	Hereditary haemolytic anaemias
005	307.1	Anorexia nervosa
	307.51	Bulimia
006	714.0	Rheumatoid arthritis
	714.1	Felty's syndrome
	714.2	Other rheum. arthritis, visceral/syst. invol.
	714.30	Polyart. juvenile rheum. arthritis, chronic/unsp.
	714.32	Pauciarticular juvenile rheumatoid arthritis
	714.33	Monoarticular juvenile rheumatoid arthritis
007	493	Asthma
008	571.2	Alcoholic cirrhosis of liver
	571.5	Cirrhosis of liver without alcohol
	571.6	Biliary cirrhosis
009	555	Regional enteritis
	556	Ulcerative enterocolitis
010	710.9	Unspecified diffuse connective tissue d.
011	290.0	Senile dementia uncomplicated
	290.1	Presenile dementia
	290.2	Senile dem. w. delusional/depress. feat.

Citation: *Piergentili P, Simon G, Paccagnella O, Grassetti L, Rizzi L, et al. (2015) Risk-Adjusted Models of Costs Referable to General Practitioners Based on Administrative Databases in the Friuli Venezia Giulia Region in Northern Italy. Arch Community Med Public Health 1(1): 012-021.*

	291.1	Alcohol amnestic syndrome
	290.4	Arteriosclerotic dementia
	294.0	Amnestic syndrome
012	253.5	Diabetes insipidus
013	250	Diabetes mellitus
014	303	Alcohol dependence syndrome
	304	Drug dependence
015	279.0	Deficiency of humoral immunity
	279.1	Deficiency of cell-mediated immunity
	279.3	Unspecified immunity deficiency
	279.4	Autoimmune disease not elsewhere class.
	279.8	Other spec. disorders immune mech.
016	070.9	Unspec. viral hepatitis w/o hepatic coma
	070.32	HVB w/o hepatic coma w/o hepatitis delta
	070.33	HVB w/o hepatic coma w. hepatitis delta
	070.54	Chronic hepatitis c w/o hepatic coma
	571.4	Chronic hepatitis
017	345	Epilepsy
018	277.0	Cystic fibrosis
019	365.1	Open-angle glaucoma
	365.3	Corticosteroid-induced glaucoma
	365.4	Glaucoma assoc. w. congenital anomalies, dystrophies, and systemic syndromes
	365.5	Glaucoma associated w. lens disorders
	365.6	Glaucoma assoc. w. other. ocular disorders
	365.8	Other specified forms of glaucoma
020	V08	Asymptomatic HIV infection status
	042	HIV disease
021	428	Heart failure
022	255.4	Corticoadrenal insufficiency
023	585	Chronic renal failure
024	518.81	Acute respiratory failure
025	272.0	Pure hypercholesterolemia
	272.2	Mixed hyperlipidaemia
	272.4	Other and unspecified hyperlipidaemia
026	252.0	Hyperparathyroidism
	252.1	Hypoparathyroidism
027	243	Congenital hypothyroidism
	244	Acquired hypothyroidism
028	710.0	Systemic lupus erythematosus
029	331.0	Alzheimer's disease
030	710.2	Sicca syndrome
A31		Hypertension without organ damage
031		Hypertension with organ damage
032	255.0	Cushing's syndrome
033	286	Coagulation defects
034	358.0	Myasthenia gravis
035	242.0	Toxic diffuse goitre
	242.1	Toxic uninodular goitre
	242.2	Toxic multinodular goitre
	242.3	Toxic nodular goitre, unspecified type
036	443.1	Thromboangiitis obliterans
037	731.0	Osteitis deformans w/o bone tumour
038	332	Parkinson's disease
	333.0	Other degenerative d. of the basal ganglia
	333.1	Essential and other spec. forms of tremor
	333.5	Other choreas
039	253.3	Pituitary dwarfism
041	341.0	Neuromyelitis optica
042	577.1	Chronic pancreatitis
044	295.0	Simple type schizophrenia
	295.1	Disorganized type schizophrenia
	296.0	Manic disorder, single episode
	295.2	Catatonic type schizophrenia
	296.1	Manic disorder, recurrent episode
	297.0	Paranoid state simple
	295.3	Paranoid type schizophrenia
	296.2	Major depressive disorder, single episode
	297.1	Paranoia
	298.0	Depressive type psychosis
	296.3	Major depressive d., recurrent episode
	297.2	Paraphrenia
	298.1	Excitative type psychosis
	299.0	Infantile autism
	295.5	Latent schizophrenia
	296.4	Bipolar affective disorder, manic
	297.3	Shared paranoid disorder
	298.2	Reactive confusion
	299.1	Disintegrative psychosis
	295.6	Residual schizophrenia
	296.5	Bipolar affective disorder, depressed
	295.7	Schizo-affective type schizophrenia
	296.6	Bipolar affective disorder, mixed
	298.4	Psychogenic paranoid psychosis
	295.8	Other specified types of schizophrenia
	296.7	Bipolar affective disorder unspecified
	296.8	Maniac-depressive psychosis, other or unspecified disorders
	297.8	Other specified paranoid states
	298.8	Other and unspecified reactive psychosis
	299.8	Other specified early childhood psychoses
045	696.0	Psoriatic arthropathy
	696.1	Other psoriasis and similar disorders
046	340	Multiple sclerosis
047	710.1	Systemic sclerosis
052	V42.0	Kidney replaced by transplant
	V42.1	Heart replaced by transplant
	V42.6	Lung replaced by transplant
	V42.7	Liver replaced by transplant
	V42.8	Other transplanted specified organ/tissue
	V42.9	Unspecified organ or tissue replaced by transplant
053	V42.5	Cornea replaced by transplant
054	720.0	Ankylosing spondylitis
055	010	Primary tuberculous infection
	011	Pulmonary tuberculosis
	012	Other respiratory tuberculosis
	013	Meninges and central nervous system TBC
	014	Intest., peritoneum, mesenteric glands TBC
	015	Tuberculosis of bones and joints
	016	Tuberculosis of genitourinary system
	017	Tuberculosis of other organs
	018	Miliary tuberculosis
056	245.2	Chronic lymphocytic thyroiditis
079	079.53	Human immunodeficiency virus type 2
	079.53	Human immunodeficiency virus

Citation: *Piergentili P, Simon G, Paccagnella O, Grassetti L, Rizzi L, et al. (2015) Risk-Adjusted Models of Costs Referable to General Practitioners Based on Administrative Databases in the Friuli Venezia Giulia Region in Northern Italy. Arch Community Med Public Health 1(1): 012-021.*

All referrals from the GP to a specialist, as well as self-referred contacts by patients, are recorded in the outpatient database. The database includes all services obtained through the NHS, including visits to a medical specialist, lab tests, radiology, imagining, ambulatory surgery, and so on. All services provided in a hospitalization episode of any kind are excluded. Services received in the emergency room are included, if they were not followed by hospitalization.

An anonymous code was substituted for the personal identifier before data was released to the research team to protect patient privacy. Databases were then linked by this code to fully reconstruct goods and services received from the NHS.

The total individual daily tariff was selected as the dependent variable. It was obtained by summing up the prices of prescribed drugs and tariffs of outpatient and inpatient health care services received. Tariffs do not always correspond to real costs, but they allow an indirect measure of the burden involved with different types of services. They are the only means to reasonably compare different services (hospitalization episodes, lab tests, etc.).

The time unit was the number of days during which each citizen was registered with the same GP (usually 365 days; a shorter period in case of GP change or patient death during the year). The dependent variable was transformed by means of the inverse hyperbolic sine function, which, unlike the logarithmic transformation, allows management of all real numbers (in particular 0). Data analysed was structured hierarchically, as patients were grouped by their GP.

Data structure is characterized by information measured at two levels: patient level (level1) and GP level. The set of level1 variables included the main available patient characteristics:

1. Gender;
2. Age;
3. Ticket exemptions (in the final models only the most frequent exemptions were included, while the variable "other" groups the remaining exemptions);
4. Information on the death of a patient in the reference year;
5. Information on childbirths in the reference year;
6. Change of GP in the reference year;
7. Home care services received, if any.

When the dependent variable is roughly continuous over strictly positive values but is zero for a large proportion of individuals, the Tobit approach is a more useful solution than standard regression modelling [4,5], and is characterized by latent variable modelling. The level 1 Tobit model can be described by the following equations:

$$y_i^* = x_i'\beta + \varepsilon_i \quad y_i = \begin{cases} y_i^* & if \quad y_i^* > 0 \\ 0 & if \quad y_i^* \le 0 \end{cases} \quad \varepsilon_i \mid x_i \sim N(0,\sigma^2) \qquad (1)$$

where $i = 1, 2, ..., N$ index patients, y^* is the latent variable measuring the real expenditure of the i-th patient, y measures the observed costs of the i-th patient, x_i is the vector of patient characteristics, β is the vector of parameters to be estimated, and ε_i is an independent and identically distributed error term.

The hierarchical structure of the data could be exploited in order to specify a multilevel solution for the analysis. Therefore, multilevel Tobit models can be estimated [6-8]. The multilevel approach allows separating the total variability into two components, variability within and variability between GPs, controlling for the differences in the patient characteristics. Let $j = 1, 2, \ldots, J$ index GPs (level 2 units) and $i = 1, 2, ..., n_j$ index patients (level 1 units) assisted by the j – th GP. Let y^* be a latent variable as before. The multilevel Tobit model is then defined as a latent variable model:

$$y_{ij}^* = x_{ij}^t\beta + u_j + \varepsilon_{ij} \quad y_{ij} = \begin{cases} y_{ij}^* & if \quad y_{ij}^* > 0 \\ 0 & if \quad y_{ij}^* \le 0 \end{cases} \quad \varepsilon_{ij} \mid x_{ij} \sim N(0,\sigma^2) \quad u_j \mid x_{ij} \sim N(0,\sigma_u^2) \qquad (2)$$

Where x_{ij} is the vector of characteristics of the i-th patient assisted by the j-th GP, β is the vector of parameters to be estimated, u_j is an independent and identically distributed level-2 error term and ε_{ij} is an independent and identically distributed level 1 error term, assuming these error terms not correlated.

The intra class correlation coefficient (ICC) is a measure of the proportion of the total variability explained by the variability between groups [6]. In standard linear models the R^2 coefficient, which measures the explained proportion of variance, is the basic estimate of model goodness of fit. However, defining R^2 in this way for hierarchical linear models is rather problematic. An alternative indicator is obtained by defining R^2 (at the patient level) as the proportional reduction of error for predicting the level 1 dependent variable with respect to the model without any predictors [9]. For this reason, this indicator will be called R^2_1 in the rest of the paper.

Models (1) and (2) were estimated using STATA software [10], the former by means of the *tobit* command, the latter through the *gllamm* (Generalized Linear Latent and Mixed Models) procedure [11]. Model (2) could have been estimated also by the *xttobit* command of the STATA software. However, the adaptive quadrature implemented in *gllamm* is superior in situations involving large cluster sizes [12,13]. A slow convergence was the price paid for this accuracy.

Results

The database analysed included 1,067,239 citizens registered with a GP in 2007 in the region. Considering that 37,029 (3.5%) persons changed their GP at least once during the year, leading to more than one record, the real number of patients (level 1 units) analysed was 1,105,759. Data came from 1,129 GPs, each of whom had an average of 1,109 citizens registered (standard deviation 392). The number of people with at least one ticket exemption was 461,532. Their mean age was 67.8 (standard deviation 17.3). Those having two or more exemptions were 35,964 (7.8%), leading to 505,932 records in the relative database.

Table 2 shows the most frequent exemptions, which were

Table 2: Distribution of the main ticket exemptions, Region Friuli Venezia Giulia, 2007.

Code	Definition	N.	%
002	Cardiovascular diseases	39,385	3.69
013	Diabetes	47,589	4.46
019	Glaucoma	14,356	1.35
025	Cholesterol	17,251	1.62
031	Hypertension with organ damage	69,869	6.55
A31	Hypertension without organ damage	42,732	4.00
048	Cancer	55,888	5.24
S51	Civil invalid, > 66%, < 100%	22,605	2.12
S52	Civil invalid, not self-sufficient	21,097	1.98
S57	Civil invalid, 100% self sufficient	18,590	1.74
	Others	100,865	9.45

Table 3: Content of some administrative databases, Region Friuli Venezia Giulia, 2007.

Services	Patients with at least one record	N. of records	All				Patients with at least 1 record			
			Mean tariff (€)	St. Dev.	Per-day mean tariff (€)	St. Dev.	Mean tariff (€)	St. Dev.	Per-day mean tariff (€)	St. Dev.
Inpatient	119,934	199,026	663.73	3,219.93	3.75	88.19	5,906.22	7,829.22	33.36	261.19
Outpatient	679,361	13,323,367	170.68	504.40	0.61	7.34	268.12	611.19	0.96	9.19
Prescriptions	697,505	9,297,450	218.16	474.37	0.77	10.35	333.81	552.90	1.17	12.78
Total	841,242	-	1,052.57	3,502.06	5.13	93.47	1,335.33	3,896.37	6.51	105.24

included in the models. Table 3 reports the number of records in the administrative databases regarding hospital and outpatient services and medications.

Three multilevel Tobit models were estimated: one model without any predictors, one model including only gender and age covariates, and the full model with variables accounting for Casemix patient characteristics added to age and gender in the linear predictor. Table 4 reports the reduction of prediction errors at the patient level when covariates are included into the multilevel Tobit model for total expenditures. Estimation results of the multilevel Tobit model for total expenditure are reported in Table 5.

The ICC shows in the multilevel model estimation that, controlling for the differences in the patient Casemix, the amount of GP-driven residual variability of the total individual health care tariffs is negligible (0.89%). The multilevel model produces similar results for the individual costs for inpatient services (ICC<0.1%), outpatient services (ICC=1.49%), and drug prescriptions (ICC=2.0%).

Based on this result, only the standard Tobit model (i.e., not accounting for the hierarchical structure of the data) was considered. This choice was also supported by the closeness of the estimated coefficients between the standard and the multilevel Tobit models. All patient variables resulted significantly and positively (except GP change) associated with the dependent variable in all models. The standard Tobit model, where each specific type of health service is considered a dependent variable, leads to similar conclusions (Table 6).

None of the characteristics associated with GPs, such as age, gender, seniority, and working in groups or alone, showed any association with the dependent variable. All of these features were not statistically significant in predicting different tariff costs. These variables were therefore not included in the model.

Conclusion

Risk adjustment is a necessary step to properly evaluate outcomes and costs of medical care. By now, casemix adjustment is a nearly universal routine in the hospital setting. This is not the case with PHC. Although PCSs are available for PHC, as mentioned in the introduction, their use is not widespread, especially outside the US. The Adjusted Clinical Groups (ACG®), which were developed at Johns Hopkins, are an example of the most common PCS in PHC. In many studies they were used to categorise PHC data [14,15]. The Clinical Risk Groups (CRG™) are another well-known PCS developed by the same group that maintains DRGs for the Health Care Financing Administration of the US government [16,17].

However, ACG®, CRG™, and other episode softwares require a level of data detail that can be found only in the patient record. Thus, they are built from providers' information systems, whether the providers are practices, single GPs, or more complex organizations [18]. It would be hard, and certainly impossible in Italy, to use ACG® in studies covering all populations, which is what is needed to manage a NHS. Thus, evaluation should be based both on GPs and population data, and the analyses performed should take into account the contribution of people who had no contact with the NHS whatsoever. Other methods must be developed to avoid delaying PHC risk

Citation: *Piergentili P, Simon G, Paccagnella O, Grassetti L, Rizzi L, et al. (2015) Risk-Adjusted Models of Costs Referable to General Practitioners Based on Administrative Databases in the Friuli Venezia Giulia Region in Northern Italy. Arch Community Med Public Health 1(1): 012-021.*

Table 4: Percentage of error reduction for predicting the dependent variable at the patient level when estimating the multilevel Tobit model, all tariffs Region Friuli Venezia Giulia, 2007.

Multilevel Tobit Model	R^2_1 (patient level)
Model 1: no predictors	–
Model 2: gender and age of patients	24.7%
Model 3: gender, age and Casemix characteristics of patients	44.6%

Table 5: Estimates of standard and multilevel tobit model on total expenditures Region Friuli Venezia Giulia, 2007.

Predictive variable		Tobit model		Multilevel Tobit model	
		Coeff.	St. error	Coeff.	St. error
Sex	Female	0.152	0.002	0.145	0.002
Age	35-44	0.146	0.003	0.141	0.003
	45-54	0.291	0.004	0.284	0.003
	55-64	0.533	0.004	0.525	0.003
	65-74	0.856	0.004	0.844	0.004
	75 or older	1.015	0.004	0.999	0.004
Ticket exemption	Cardiovascular diseases (002)	0.586	0.006	0.586	0.005
	Diabetes (013)	0.609	0.005	0.605	0.005
	Glaucoma (019)	0.389	0.009	0.384	0.008
	Cholesterol (025)	0.471	0.008	0.459	0.008
	Hypertension with organ damage (031)	0.498	0.004	0.497	0.004
	Hypertension without organ damage (A031)	0.426	0.005	0.432	0.005
	Cancer (048)	0.857	0.005	0.850	0.004
	Civil invalid > 2/3 (S51)	0.464	0.007	0.458	0.007
	Civil invalid with total inability (S52)	0.156	0.008	0.149	0.007
	Civil invalid, 100% self sufficient (S57)	0.453	0.008	0.447	0.008
	Other	0.644	0.004	0.633	0.004
GP change		– 0.118	0.004	– 0.100	0.004
Home care		0.651	0.005	0.672	0.005
Patient death		1.931	0.009	1.926	0.009
Patient pregnancy		1.285	0.011	1.292	0.010
Constant		– 0.077	0.003	– 0.063	0.004
Patient variance		1.124		1.015	
MMG variance		–		0.009	
ICC		–		0.89 %	
Log likelihood		– 1459576.1		– 1459450.9	

adjustment in Italy. The easiest strategy is to consider administrative data bases.

Administrative databases are a convenient source of information for the management and evaluation of health care systems. They are already available and often provide low-cost data on large number of patients or citizens, or even the whole population of a given area. However, the quality of this data may not be excellent. Since Lisa Iezzoni proposed the use of administrative databases to evaluate the quality of health services [19], they have been increasingly used in many research areas [20]: quality of care assessment, estimating adherence to best practice, cost evaluation and epidemiology of selected diseases, potential benefits and harms of specific health policies, and disease and intervention registries. The number of published works on the topic is large, particularly those using hospital data, but a literature review is beyond the scope of this article.

In Italy, in addition to studies based on hospital discharge abstracts, attention is paid to databases of medications and outpatient contacts. For instance, a large epidemiologic multicentre study estimated incidence and prevalence of a number of chronic diseases [21]. The Lombardia Regional Health Authority has developed a data warehouse that combines a number of health administrative databases containing data on a population of eight million people

Citation: Piergentili P, Simon G, Paccagnella O, Grassetti L, Rizzi L, et al. (2015) Risk-Adjusted Models of Costs Referable to General Practitioners Based on Administrative Databases in the Friuli Venezia Giulia Region in Northern Italy. Arch Community Med Public Health 1(1): 012-021.

Table 6: Estimates of standard Tobit models on different types of expenditures Region Friuli Venezia Giulia, 2007.

Coefficient		Hospital		Outpatient		Drug prescriptions	
		Coeff.	St. error	Coeff.	St. error	Coeff.	St. error
Sex	Female	0.147	0.016	0.135	0.001	0.059	0.001
Age	35-44	0.371	0.027	0.074	0.002	0.107	0.002
	45-54	0.471	0.029	0.141	0.002	0.234	0.002
	55-64	0.940	0.028	0.224	0.002	0.447	0.002
	65-74	1.584	0.028	0.386	0.002	0.689	0.002
	75 or older	2.187	0.029	0.335	0.002	0.807	0.002
Ticket exemption	Cardiovascular diseases (002)	1.628	0.034	0.226	0.003	0.379	0.003
	Diabetes (013)	0.926	0.032	0.275	0.003	0.524	0.003
	Glaucoma (019)	0.164	0.057	0.213	0.005	0.384	0.005
	Cholesterol (025)	0.147	0.055	0.205	0.005	0.483	0.005
	Hypertension with organ damage (031)	0.434	0.029	0.150	0.003	0.514	0.003
	Hypertension w/o organ damage (A031)	0.206	0.036	0.097	0.003	0.462	0.003
	Cancer (048)	2.460	0.028	0.699	0.003	0.263	0.003
	Civil invalid > 2/3 (S51)	1.110	0.044	0.250	0.004	0.324	0.004
	Civil invalid with total inability (S52)	0.964	0.044	-0.114	0.005	0.083	0.004
	Civil invalid, 100% self sufficient (S57)	1.286	0.045	0.184	0.005	0.275	0.005
	Other	1.404	0.025	0.391	0.002	0.384	0.002
GP change		-0.470	0.032	-0.078	0.003	-0.104	0.003
Home care		2.520	0.030	0.186	0.003	0.271	0.003
Patient death		5.709	0.046	0.093	0.006	0.087	0.006
Patient pregnancy		6.235	0.056	0.627	0.006	-0.264	0.007
Constant		-7.812	0.030	-0.222	0.002	-0.301	0.002
Log likelihood		-574396.7		-933202.5		-916031.8	
R^2_1		30.9%		27.1%		49.3%	

[22]. Other studies have used administrative databases to evaluate appropriateness of drug prescriptions [23,24]. In Friuli Venezia Giulia, the Regional Health Authority has developed a data warehouse system that contains several years of data from administrative databases and allows for fast data mining. This system has been widely used for health service evaluation and epidemiological studies [25-30].

In our study the issue was whether the clinical information provided by the ticket exemption file would enable us to build a predictive model. Given the purely administrative purpose goal of the exemption, in fact, epidemiologists are sceptical about its real value as source of information for risk adjustment. The authors know of only one other large study that uses ticket exemptions [22].

Models that use ticket exemption information had robust results. Since the exemption is granted after a specialty physician diagnosis based on standard criteria, it can be regarded as very reliable. However, there are quality issues in the data. For instance, the prevalence of diabetes exemptions is much lower than is reported in the literature [31]. Nevertheless, the available information allow to derive good estimates of the trend of tariffs due to the large number of individuals. Estimated coefficients have a narrow confidence interval, and the percentage of error reduction R^1_2 (the equivalent of the *variance explained* in Tobit statistical models) is large.

In our model age has an important predictive value. The age coefficient increases steadily in models considering the sum of all expenditures, and the oldest patients use ten times more resources than patients aged 35 – 44 when all other variables are equal. This is even more evident if we look at the inpatient tariffs. This means that outpatient care (remember that for all outpatient contacts, radiology and lab tests are included) is a resource commonly used by all age groups, and this may conceal a degree of inappropriateness in the use of diagnostic resources.

Women tend to spend more in all sectors, even after controlling for pregnancy. It is worth noting that being pregnant has a protective value in medication expenditures. Considering total expenditures, cancer and diabetes are the health conditions that consume the most resources. In the other models, the costliest illnesses are cancer, diabetes, and hypertension in the outpatient model; cancer and cardiovascular disease in the hospitalization model; and hypertension and diabetes in the medication cost model (this is probably because cancer medications are given in hospitals and do not fully appear in this data).

Even though there are substantial differences in the statistical methodology and underlying organization of the health services, the variance explained by our models is largely comparable to the values found in the development of the ACG® at Yale [14]. However, our model maintained its power when analysing the total charges, while the ACG® system sees a significant drop in the variance explained. The comparison with CRG™ is more difficult, since several models were developed from different data sources using a number of different approaches. Nonetheless, the R^2 reported ranged from 0.12 to 0.14, significantly lower than our findings [17].

In Italy there are grounds for adopting the use of administrative databases––ticket exemption databases in particular––for more complex tasks as well, such as risk adjustment in PHC. The inclusion of casemix patient characteristics into the models has a strong predictive power. Their inclusion significantly reduced residual cost heterogeneity in the model predicting the total expenditure at the patient level compared to the model without any predictor or to the model including gender and age only, and is therefore highly recommended.

The most surprising finding, however, was that, controlling for Casemix, the variability of the total expenditures at the GP level is very low. Variation in medical behaviour is a well-known phenomenon widely described in the literature, and many hypotheses have been developed to explain it [32-37]. Only a few studies have so far investigated differences among practices after any sort of risk adjustment (apart from the usual sex and age standardization). For instance a study carried out in an area close to the Friuli Venezia Giulia Region found almost no variance ascribed to GPs in the distribution of hospitalization tariffs after adjustment for PCS [38]. However, when considering medication cost [39], a residual variance remained, even though a marked decrease in the overall variance was observed after applying risk adjustment.

Comparison with other studies is not straightforward, given the differences in the study design and in the dependent variables. A study on the number of home visits reported a statistical analysis similar to the one of the present study, showing an ICC of 1.6% after adjustment [40]. This result is very similar to the one found in the outpatient services model, even though the services and the dependent variable (tariff instead of service frequency) considered in the analysis are quite different. A number of studies adopted a multilevel approach as a statistical solution, assuming the frequency of services as a dependent variable (encounters, referrals, or prescriptions). In one study the variance unexplained by the models at the practice level was between 12.9% and 27.2% according to the type of service [41]. Two more studies demonstrated a much lower residual unexplained variance at the practice level, comparable to the one found in the present study (0.1% for prescription number [42] and 3.6% for referrals [43]). A study on prescription cost showed a low variance explained by physicians after risk adjustment (1.8%) [44]. A generalization derived from this picture is that application of risk adjustment models explains a large quota of variation, and decreases the variation attributable to a single GP or practice. However, in the present study this effect appears much stronger and more evident, considering that the average number of citizens registered with a GP (around 1,100) is much lower than in the UK.

This study has some particular features that may help to explain homogeneous behaviour of GPs. First, it is population-based (the entire population of Friuli Venezia Giulia); that is, it examines data of all patients without any socio-economic bias. Second, the population study is very large, much larger than the study population of other research. Third, it considered all major services provided to the population. Fourth, it was conducted in a region among those with the lowest per capita overall expenditure: the lowest hospitalization rate and the lowest per capita medication expenditure in Italy [45]. Fifth, it used the sum of service tariffs or medication prizes as a dependent variable, which is here assumed as a proxy of the burden on the health system of caring for individuals or groups of patients with common conditions. Could some of these features be associated with the low variance attributable to GPs?

The size of the population gives power to the statistical results. The comprehensiveness of the services involved in the analysis demonstrates that physician behaviour is not related to a single part of the caring process, but can be regarded as a generalised professional attitude. The use of tariffs can change the results, especially when considering non-homogenous services. This is particularly true for hospital care. Usually, indicators include hospitalization rates, but it is quite obvious that there is a difference among hospitalizations for pathologies of different burden. The use of tariffs, which are derived from DRG weights, take this burden into account. Our hypothesis is that in the Friuli Venezia Giulia Region, inpatient services are devoted to patients with homogenous severity of conditions. The hospital use appears highly appropriate when comparing regional rates of hospitalization with national rates (Friuli Venezia Giulia has the second lowest age and sex standardised hospitalization rate among Italian regions [46], and the highest rate of utilization of home care [45]). This is no longer true for other types of services, such as medications, where a non-optimal adherence to clinical guidelines is documented [24]. However, a larger variance due to GPs is observed for medication use.

Risk adjustment is then necessary for PHC indicators. Outcome and output measures without proper risk adjustment may lead to biased findings. Age and sex adjustment appears to be insufficient to adequately account for differences among GP patient conditions.

The major weaknesses of this study appear to be the following:

1. The short period of observation (one year). This is largely compensated for by the size of the study population.
2. The lack of drug cost data on people in nursing homes. These are high consumers of medication, and also usually have several ticket exemptions. Thus this may affect the model coefficients.
3. Ticket exemptions are issued for chronic conditions only. Thus acute conditions and accidents are not considered in the models.
4. Data quality problems in the ticket exemption database. These are of two kinds: first, as already stated, the completeness of the information (many people who have a given condition do not have the corresponding exemption); second, the

diagnosis specificity (for instance, cancer is considered only one category, the same for diabetes, and so forth). Given the power of the models, however, these weaknesses may become an asset; if introduced for relevant purposes, ticket exemption information would certainly increase in quality, and the categories may be modified accordingly. This would further increase the power of the models and their capacity to explain variation.

Conflict of Interest

All authors have completed the ICMJE uniform disclosure form (available on request from the corresponding author) and declare: the research project was financed by a grant from the Italian Ministry of health, issued after a competitive call and granted through public organizations of the Italian National Health System. None of the authors had financial relationships with any organization that might have an interest in the submitted work in the previous three years, or engaged in other relationships or activities that could appear to have influenced the submitted work.

References

1. Solon JA, Feeney JJ, Jones SH, Rigg RD, Sheps CG (1967) Delineating episodes of medical care. American J of Public Health 57: 401-408.
2. Hornbrook MC, Hurtado AV, Johnson RE (1985) Health care episodes: definition, measurement and use. Med Care Rev 42: 163-218.
3. Wingert TD, Kralewski JE, Lindquist TJ, Knutson DJ (1995) Constructing episodes of care from encounter and claims data: some methodological issues. Inquiry 632: 430-443.
4. Tobin J (1958) Estimation of relationships for limited dependent variables. Econometrica 26: 24-36.
5. Wooldridge JM (2010) Econometric analysis of cross section and panel data. Cambridge MIT Press.
6. Goldstein H (2011) multilevel statistical models. 4th ed. Chichester: John Wiley and Sons Ltd.
7. Shephard N, Falcaro M, Zeggini E, Chapman P, Hinks A, et al. (2003) Linkage analysis of cross-sectional and longitudinally derived phenotypic measures to identify loci influencing blood pressure. BMC Genet 4: S26.
8. Twisk J, Rijmen F (2009) Longitudinal tobit regression: a new approach to analyze outcome variables with floor or ceiling effects. J Clinical Epidemiol 62: 953-958.
9. Snijders T, Bosker R (1999) Multilevel analysis: an introduction to basic and advanced multilevel modelling. London: Sage Publications.
10. Rabe-Hesketh S, Skrondal A (2008) Multilevel and longitudinal modeling using Stata. 2nd ed. College Station.
11. Rabe-Hesketh S, Skrondal A, Pickles A (2004) GLLAMM Manual. Working Paper 160. U.C. Berkeley Division of Biostatistics Working Paper Series, Berkeley CA.
12. Poen E (2009) The tobit model with feedback and random effects: a Monte-Carlo study. CeDEx Discussion Paper Series 2009-14. Nottingham: University of Nottingham.
13. Rabe-Hesketh S, Skrondal A, Pickles A (2005) Maximum likelihood estimation of limited and discrete dependent variable models with nested random effects. J Econometrics 128: 301-323.
14. Starfield B, Weiner J, Mumford L, Steinwachs D (1991) Ambulatory Care Groups: a categorization of diagnoses for research and management. Health Services Research 26: 53-74.
15. The Johns Hopkins Bloomberg School of Public Health The Johns Hopkins ACG® system bibliography.
16. Hughes JS, Averill RF, Eisenhandler J, Goldfield NI, Muldoon J, et al. (2004) Clinical Risk Groups (CRGs). A classification system for risk-adjusted capitation-based payment and health care management. Med Care 42: 81-90.
17. Averill RF, Goldfield NI, Eisenhandler J, Muldon JH, Hughes JS, et al. (1999) Development and evaluation of clinical risk groups (CRGs). 3M HIS Research Report.
18. The Johns Hopkins Bloomberg School of Public Health. The Johns Hopkins ACG System.
19. Iezzoni L (1997) Assessing quality using administrative data. Annals Intern Med 127: 666-674.
20. Moise P (2001) Using hospital administrative databases for a disease-based approach to studying health care systems. OECD Technical Reports. Paris: OECD.
21. Terracini B (2008) [Exploiting electronic health archives for epidemiological purposes. An experience using a standardized approach to estimate diseases in various areas of Italy. Forward]. Epidemiol Prev 32: 3.
22. Barbieri P, Grieco N, Ieva F, Paganoni AM, Secchi P (2010) Exploitation, integration and statistical analysis of public health databases and STEMI archive in the Lombardia Region. Springer Milano 41-55.
23. Piergentili P (2006) [Methodology to identify quality indicators in primary health care. The case of anti-acid drugs (ATC class A02) in the ASSL10 – Veneto orientale]. Ricerca e Pratica 22: 243-250.
24. Basso B, Bomben L, D'Arenzo L, Del Pup E, Di Lorenzo A, et al. (2011) [Evaluation of the influence of specialists consultants on drug prescription in primary health care. Ace inhibitors vs angiotensin-receptor blocker in the ASS6 Friuli occidentale, Pordenone]. Ricerca e Pratica 27: 4-12.
25. Simon G, Francescutti C, Brusin S, Rosa F (1995) Variation in drug prescription costs and general practitioners in an area of North-East Italy. The use of current data. Epidemiol Prev 18: 224-229.
26. Bordignon M (1997) Medici di base e spesa farmaceutica: prolegomeni per la costruzione di un budget preventivo di spesa, in Petretto A. (a cura di). Bologna: Economia della Sanità, Il Mulino 221-253.
27. Sturkenboom MC, Romano F, Simon G, Correa-Leite ML, Villa M, et al. (2002) The iatrogenic costs of NSAID therapy: a population study. Arthritis Rheum 47: 132-140.
28. Borgnolo G, Simon G, Francescutti C, Lattuada L, Zanier L (2001) Antibiotic prescription in Italian children: a population-based study in Friuli Venezia Giulia, North-East Italy. Acta Paediatr 90: 1316-1320.
29. Franzo A, Simon G, Francescutti C (2006) Mortality associated with delay in operation after hip fracture: ... but Italian data seem to contradict study findings. BMJ 332: 1093.
30. Varas-Lorenzo C, Garcia-Rodriguez LA, Cattaruzzi C, Troncon MG, Agostinis L, et al. (1998) Hormone replacement therapy and the risk of hospitalization for venous thromboembolims: a population-based study in Southern Europe. American J Epidemiol 147: 387-390.
31. Wild S, Roli G, Green A, Sicree R, King H (2004) Global prevalence of diabetes: estimates for the year 2000 and projections for 2030. Diabetes Care 27: 1047-1053.
32. Davis P, Gribben B, Scott A, Lay-Yee R (2000) The "supply hypothesis" and medical practice variation in primary care: testing economic and clinical models of interpractitioner variation. Soc Sci Med 50: 407-418.
33. O'Donnel CA (2000) Variation in GP referral rates: what can we learn from the literature? Family Practise 17: 462-471.
34. Parchman ML (1995) Small area variation analysis: a tool for primary care research. Family Medicine 27: 272-276.

35. Rice N, Dixon P, Lloyd DCEF, Roberts D (2000) Derivation of a needs based capitation formula for allocating prescribing budgets to health authorities and primary care groups in England: regression analysis. BMJ 320: 284-288.

36. Orzella L, Chini F, Giorgi Rossi P, Borgia P (2010) Physician and patient characteristics associated with prescription and costs of drugs in the Lazio Region of Italy. Health Policy 95: 236-244.

37. Phelps CE, Mooney C (1993) Variations in medical practice use: causes and consequences. In: Arnould RJ, Rich RF, White W, editors. Competitive Approaches to Health Care Reform. Washington, DC: The Urban Institute Press.

38. Piergentili P, Paccagnella O (2009) Multivariate analysis models of costs referable to general practitioners based on administrative databases. Epidemiol Prev 33: 88-95.

39. Piergentili P, Paccagnella O (2006) Per una corretta stima dei costi riferiti ai medici di medicina generale. I dati di consumo di farmaci dell'ASSL10 del Veneto. SIMG: 30-36.

40. O'Sullivan C, Omar RZ, Forrest CB, Majeed A (2004) Adjusting for Casemix and social class in examining variation in home visits between practices. Family Practice 21: 355-363.

41. Davis P, Gribben B, Lay-Yee R, Scott A (2002) How much variation in clinical activity is there between general practitioners? A multi-level analysis of decision-making in primary care. J Health Serv Res Policy 7: 202-208.

42. Omar RZ, O'Sullivan C, Petersen I, Islam A, Majeed A (2008) A model based on age, sex, and morbidity to explain variation in UK general practice prescribing: cohort study. BMJ 337: a238.

43. O Sullivan C, Omar RZ, Ambler G, Majeed A (2005) Case-mix and variation in specialist referrals in general practice. British Journal of General Practice 55: 529-533.

44. Aguado A, Guinó E, Mukherjee B, Sicras A, Serrat J, et al. (2008) 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 Services Research 8: 53.

45. Ministero della salute. Attività gestionali ed economiche della ASL e Aziende ospedaliere. Annuario statistico del Servizio sanitario nazionale. Roma: Ministero della salute; 2011.

46. Ministero della salute. Rapporto sull'attività di ricovero ospedaliero - Dati SDO 2010, table F2. Rome, 2011.

Citation: *Piergentili P, Simon G, Paccagnella O, Grassetti L, Rizzi L, et al. (2015) Risk-Adjusted Models of Costs Referable to General Practitioners Based on Administrative Databases in the Friuli Venezia Giulia Region in Northern Italy. Arch Community Med Public Health 1(1): 012-021.*