ORIGINAL ARTICLES

ebph

Administrative databases as a tool for identifying healthcare demand and costs in an over-one million population

Fabiana Madotto[(1)], Michele A. Riva[(1)], Carla Fornari[(1, 2)], Luciana Scalone[(1)], Roberta Ciampichini[(3)], Chiara Bonazzi[(4)], Lorenzo G. Mantovani[(1, 3, 5)], Giancarlo Cesana[(1, 2)]

ABSTRACT

BACKGROUND: the aim of this study was to assess healthcare demand of specific groups of population and their costs borne by Italian Health System, using healthcare administrative databases.

METHODS: demographic, clinical and economic data were obtained from datasets available at the Regional Health System, combined into a data warehouse (DENALI), using a probabilistic record linkage to optimize the data matching process. The study population consisted of more than 1 million people registered in 2005 at one Local Healthcare Unit of Lombardy. Eight different segments were identified. Costs occurring in 2005 for hospital admissions, drug prescriptions, outpatient medical specialist visits were quantified in each segment.

RESULTS: healthy people accounted for 53% of the population and cost € 180 per-capita. Subjects with only one chronic disease made up 16% of the population and cost € 916 per-capita, those affected by several chronic diseases accounted for 13% and cost € 3 457 per-capita. Hospitalizations were the cost driver in five segments, ranging from 42% to 89% of total expenditures. Outpatient visits were the cost driver among healthy subjects (54%) and those with a possible chronic disease (42%), while drug costs ranged between 4% ("acute event") and 32% ("one chronic disease"). Overall, healthcare cost was € 809 per-capita.

CONCLUSIONS: healthcare costs were mainly determined by people affected by chronic conditions, even if "healthy people" ranked third for total expenditure. These costs need an appropriate identification of healthcare demand, that could be efficiently monitored through the use of administrative databases.

(1) CESP, Research Centre on Public Health, University of Milano-Bicocca, Monza, Italy
(2) Department of Statistics and Quantitative Methods, University of Milano-Bicocca, Milan, Italy
(3) CHARTA Foundation, Center for Health Associated Research and Technology Assessment, Milan, Italy
(4) Department of Health Sciences, University of Milano-Bicocca, Monza, Italy
(5) CIRFF, Center of Pharmacoeconomics, Federico II University of Naples, Naples, Italy

CORRESPONDING AUTHOR: Fabiana Madotto, Research Centre on Public Health, University of Milano-Bicocca, via Cadore 48, 20900 Monza (Italy). Tel: 0039 039 2333097, Fax: 0039 039 365378. e-mail: fabiana.madotto@unimib.it
DOI: 10.2427/8840

INTRODUCTION

Healthcare demand continues rising in most Western countries and the challenge is how to sustain the high costs resulting. Furthermore the recent global financial crisis is having an impact on the healthcare systems, especially in countries offering an universal coverage. In particular, in Italy the slow economic growth is limiting both public and private health expenditures and is making difficult to meet the health needs and demand of population [1]. In this country, policy-makers are responding to crisis through interventions aimed at increasing the efficiency of public spending and rationalizing costs. These actions may be improved by a more accurate knowledge of healthcare demand, that sometimes can provide a direct estimation of individual needs [2].

Managing demand for health care has been defined as the process of identifying where, how, why and by whom demand arises and then deciding on the best methods of dealing with it, so that the most efficient, appropriate and equitable approach can be developed and applied [3].

A study conducted by Lynn *et al.* suggests that the analysis of healthcare demand can be efficiently obtained by segmenting the target population on health prospects and priorities and identifying homogeneous groups with specific provision of services [4]. Population size and annual healthcare costs for each segment were estimated obtaining data from various sources (i.e. National Centre for Health Statistics, scientific and medical societies, surveys, publications) that, unfortunately, used different definitions and time periods. Hence such estimates were only rough approximations and the authors themselves encourage further research aimed at exploring and setting up more comprehensive and satisfactory models.

Indeed, one crucial point in constructing accurate and precise demand models is related to the type and amount of relevant data available [5-9]. Thus, administrative databases of national or Regional Health Systems (RHSs) may constitute a fundamental source of readily available and relatively inexpensive large amounts of good quality data (demographic, clinical, economic) referring to the general population [10-14]. In this way, these datasets allowed us to identify and monitor health needs and demand of the population and consequently optimize the use of economic resources available.

The aim of this study was to build a segmentation model that can be applied on healthcare administrative databases in order to assess demand and the main costs borne by Italian health system for the care of specific and distinguished groups of a general population.

METHODS

Data Source

We developed a model suitable to estimate consumption of healthcare resources and related costs occurring in one year within the RHS of Lombardy. Data were obtained from the administrative datasets available within the Lombardy RHS, regarded: hospital discharges, pharmaceutical prescriptions, outpatient claims (laboratory and diagnostic examinations, specialist medical visits). These datasets and demographic characteristics of people under the care of the Lombardy RHS were organized into a data warehouse named DENALI. Since the quality of data reported in these databases was not optimal, a probabilistic record linkage (PRL) was implemented to match the anonymized data of the different datasets belonging to the same individuals [15-17]. PRL uses available individual information to establish a link between database records belonging to the same subject and, unlike deterministic record linkage it does not require complete identifiers to match exactly. Some studies showed that this method provides the most accurate technique of matching files when they do not share a single common identifier or when there are errors or omissions in the identifiers [18-19].

Subjects and setting

The study was limited to the population enrolled on January 1st 2005 in one of the fifteen Local Healthcare Units of Lombardy, about 1 million inhabitants. The subjects who had been registered for less than two years were excluded from the analysis, while individuals who were born between January 1st 2003 and December 31st 2005 were included.

Modelling

To organize our model we followed some criteria previously adopted in the study by Lynn *et al.*, segmenting the target population into groups "sufficiently homogeneous to enable arranging a set of commonly needed supports and services to meet the expected (health) needs" of their members [4]. So, we stratified the population into eight different and sequential segments according to the following features: a) the set of population segments must be limited; b) the set of segments must include everyone, so that at every point in his/her life, every person fits into only one segment; c) when a subject meets the criteria for several segments, he is assigned to the first segment in the order established; d) individuals in each segment must have similar health care needs and priorities, to make the segment useful for planning, but each segment must be different enough to justify separate considerations. On the basis of these characteristics, we identified the following segments:

1. *Subjects unknown to the RHS*
   This group involved subjects that during a 3-year period (2003-2005) did not use any health care service.

2. *Maternity and infancy*
   This group was defined according to age at 1st June 2005. The maternity sub-group included women aged between 10 and 54 years, who during the year 2005 were hospitalized or received medical examinations for their pregnancy, and women and men who applied for fertility treatments. The infancy sub-group included children aged less than one year.

3. *Elderly*
   This group involved subjects aged 85 years or more at 1st June 2005.

4. *People affected by only one chronic disease (CD)*
   This group involved subjects that between the years 2003-2005 consumed healthcare resources for one of the following CDs: malignant tumours, cerebrovascular diseases, hypertension, ischemic heart diseases, peripheral angiopathies, chronic heart failure, non-ischemic heart diseases (vasculopathy, arrhythmias, etc.), continuous anticoagulant treatment, diabetes mellitus, dyslipidemia, nephropathies and end-stage renal diseases, chronic obstructive pulmonary diseases, gastroenteric diseases, endocrine disorders, neurological diseases, major psychiatric disorders (schizophrenia, psychosis), minor psychiatric disorders (anxiety, depression), rheumatic illnesses, hepatic disorders, inherited genetic defects and congenital metabolic diseases, HIV infection and AIDS, diseases for which they received an organ transplantation, continuous therapy with nonsteroidal anti-inflammatory drugs (NSAIDs), antianemic drugs (iron, folate, B12 and erythropoietin), glucocorticoid drugs. Patients were assigned to these sub-groups if during the observational period they had at least one hospital admission (according to ICD-9-CM) and/or outpatient visit (according to specific codes) and/or continuous use of drugs (according to ATC code on prescriptions) for only one of the diseases aforementioned [20-22]. A drug therapy was defined as continuous if the drug prescriptions covered a clinically meaningful number of days. The drug coverage period was computed using the defined daily dose and expressed as proportion of days covered (PDC), calculated as the number of days with drug on-hand divided by the number of days of observation [23].

5. *People affected by more than one CD*
   This group included subjects with more than one of the aforementioned CDs and it was further divided according to the number of diseases.

6. *People possibly affected by CD, or at an early stage*
   This group included subjects not belonging to the groups above but who occasionally (according to PDC) received drug prescriptions for the aforesaid CDs.

7. *People suffering from an acute event*
   This group involved people who, in 2005, were hospitalized for an accident or acute event, excluding cardiovascular disease and including traumas.

8. *Healthy people*
   All the remaining people not belonging to the aforementioned segments were included in this group. It was further sub-grouped according to the amount of costs generated by its members:
   - *Heavy consumers.* This subgroup included subjects who generated high level costs in at least one determinant (drugs, hospital admissions or outpatient claims) in 2005.

The threshold used to define "high level" for each determinant was the per-capita cost of the segment 4.

- *Moderate consumers.* Subjects not included among the heavy consumers, who generated a moderate level cost in at least one determinant of health expenditure. The thresholds used to define "moderate level" were the per-capita costs estimated within the healthy people segment.

- *Mild consumers.* This subgroup involved people costing more than 0 € but not included in the subgroups above.

- *Non consumers.* The subgroup included all the people costing nothing during the 2005.

Data Analyses

For each group, we estimated the number and percentage of subjects, median age, total and per-capita (mean ± standard error) annual cost. We considered only the health resources consumed during the year 2005. The health care expenditures were stratified according to three cost components: hospital admissions, outpatient claims and drugs.

RESULTS

Description of segments

The sample consisted in 1 031 684 subjects with a median age of 42 years and a male prevalence of 49% (Table 1). The "healthy people" accounted for more than half of the sample (52.9%), followed by "people affected by only one CD" (16.4%), and then by "people affected by more than one CD" (13.0%). The segments with a low prevalence were "maternity and infancy" (2.6%), "elderly" (1.7%) and "people suffering from an acute event" (0.9%). Overall, about 18% of the population did not use any public health resource in 2005 ("healthy non consumers" and "subjects unknown to the RHS").

As expected, the categories were related to age. The youngest was "people suffering from an acute event" (median age 25 years), then the "maternity" sub-group and "healthy people" segment (respectively 33 and 34 years) and the

"subjects unknown to the RHS" (39 years). This latter segment, together with "people suffering from an acute event", also had a higher prevalence of men than women, while the "elderly" category had the lowest prevalence of male (excluding the "maternity" sub-group).

As regards the segment "people affected by only one CD", the most common conditions detected were: hypertension (38.2% of subjects within the segment), major or minor psychiatric disorders (14.4%), neurological diseases (6.9%), endocrine disorders (6.3%), and chronic obstructive pulmonary diseases (6.2%) (Figure 1).

The subjects in "people affected by more than one CD" segment had a median age of 67 years and 50.0% were affected by two chronic conditions (Table 2). In this segment, the subjects affected by hypertension were 73.4%, followed by people with a continuous anticoagulant treatment (25.8%), affected by dyslipidemia (23.8%) or diabetes mellitus (21.2%) (Table 3). Moreover, the people affected by HIV/AIDS, inherited genetic defects (or congenital metabolic diseases), major psychiatric disorders and organ transplantation were the youngest, respectively with a median age of 44, 49, 53 and 53 years. Subjects affected by endocrine disorders, minor psychiatric disorders, treated with anti-anaemic drugs or with NSAIDs were predominantly women; instead HIV/AIDS was most prevalent in men.

"Healthy people" were segmented into four sub-group; the "mild consumers" were the half of this segment (50.6%). In spite of this, the sub-group "moderate consumers" accounted for 26.8% of the sample, "heavy consumers" for 5.1% and "non-consumers" for 11.6%.

Health expenditures and determinants of costs

The study population consumed healthcare public resources for a total amount of € 834 million, corresponding to € 808 per-capita, in 2005 (Table 1). The segment consuming more resources was "people affected by more than one CD" (55.4% of total expenditure), then "people affected by only one CD" (18.6%), followed by the "healthy people" (11.8%). People with multiple CDs were also the segment with the higher per-capita cost (€ 3 457), followed by "people suffering from an acute event" (€ 2 556), "maternity and infancy" (€ 2 067) and the "elderly" segment (€ 1 909). In all these groups more than 50% of expenditures were due to

TABLE 1

SEGMENT DEMOGRAPHIC CHARACTERISTICS AND COSTS					
SEGMENT AND SUB-GROUP	SUBJECTS N (%)	MALE %	AGE $Q_{0.5}$ $(Q_{0.25}$-$Q_{0.75})$[b]	TOTAL COST MILLION EURO (%)	PER-CAPITA COST (EURO) MEAN ± SE[c]
Unknown to the RHS[a]	63 137 (6.12)	60.53	39 (28 - 51)	0.00 (0.00)	0.00 ± 0.00
Maternity and infancy	27 053 (2.62)	22.66	26 (0 - 33)	36.71 (4.40)	1 357.13 ± 10.88
Maternity	15 273 (1.48)	0.01	33 (29 - 36)	31.58 (3.79)	2 067.74 ± 14.69
Infancy	11 780 (1.14)	52.01	0 (0 - 0)	5.13 (0.62)	435.81 ± 11.57
Elderly	17 487 (1.69)	23.94	89 (86 - 91)	33.38 (4.00)	1 908.96 ± 29.34
People with one CD[d]	169 649 (16.44)	44.51	56 (41 - 67)	155.48 (18.64)	916.48 ± 6.46
People with several CDs[d]	133 670 (12.96)	46.93	67 (58 - 75)	462.13 (55.40)	3 457.28 ± 19.81
People with possible CD[d] or at an early stage	65 352 (6.33)	46.16	53 (40 - 64)	23.35 (2.80)	357.30 ± 4.17
People with acute event	9 738 (0.94)	60.57	25 (10 - 43)	24.89 (2.98)	2 556.12 ± 34.18
Healthy people	545 598 (52.88)	51.46	34 (19 - 46)	98.24 (11.78)	180.06 ± 1.06
Non consumers	119 889 (11.62)	58.20	33 (20 - 44)	0.00 (0.00)	0.00 ± 0.00
Mild consumers	276 153 (26.77)	50.79	32 (15 - 44)	22.72 (2.72)	82.27 ± 0.28
Moderate consumers	96 513 (9.35)	47.29	36 (22 - 49)	30.17 (3.62)	312.62 ± 1.33
Heavy consumers	53 043 (5.14)	47.30	42 (31 - 56)	45.35 (5.44)	854.91 ± 9.93
Population	**1 031 684**	**48.81**	**42 (27 - 60)**	**834.19**	**808.57 ± 3.12**

[a]RHS=Regional Health System; [b]$Q_{0.25}$=First quartile; $Q_{0.5}$=Median; $Q_{0.75}$=Third quartile; [c]SE=Standard Error. [d]CD=Chronic disease

hospitalization (Figure 2). The "healthy people" had the lowest per-capita cost (€ 180), excluding the segment of "subjects unknown to the RHS", which by definition did not cost anything. The sub-group "heavy consumers" had a per-capita cost very similar to that of "people affected by only one CD" segment.

The segment "people affected by more than one CD" had a per-capita cost of € 1 910. Data in Table 2 show that the number of diseases was positively associated with per-

capita costs, in addition to age: with a rising number of CDs, the median age and the per-capita cost increase.

Costs for outpatient visits and examinations were similar to cost for drugs, about € 200 million each (24.0% of total expenditure). However, the economic burden of the three cost components varied considerably between the different segments; hospital admissions were the cost drivers in the segments "people suffering from an acute event" (88.7% of total

TABLE 2					
DEMOGRAPHIC CHARACTERISTICS AND COSTS OF "PEOPLE AFFECTED BY MORE THAN ONE CHRONIC DISEASE" SEGMENT, STRATIFIED BY THE NUMBER OF CONCOMITANT CHRONIC CONDITIONS					
N. OF CDS[a]	SUBJECTS N (%)	MALE %	AGE $Q_{0.5}$ $(Q_{0.25}-Q_{0.75})$[b]	TOTAL COST MILLION EURO (%)	PER-CAPITA COST (EURO) MEAN ± SE[c]
2	66 887 (50.04)	43.38	65 (54 - 73)	127.75	1 909.97 ± 20.31
3	31 709 (23.72)	46.81	68 (60 - 75)	99.52	3 138.40 ± 34.16
4	16 283 (12.18)	50.77	70 (62 - 76)	76.97	4 727.08 ± 60.49
5	9 220 (6.90)	54.74	71 (63 - 77)	59.94	6 501.49 ± 96.46
6	4 798 (3.59)	57.71	72 (65 - 78)	40.86	8 516.21 ± 157.88
7	2 486 (1.86)	57.12	73 (66 - 78)	25.77	10 365.71 ± 269.31
8	1 258 (0.94)	58.66	73 (67 - 79)	15.20	12 083.39 ± 354.84
9 or more	1 029 (0.77)	61.42	74 (67 - 78)	16.12	15 666.25 ± 480.05
Segment	133 670	46.93	67 (58 - 75)	462.13	3 457.28 ± 19.81

[a]CD=Chronic disease; [b]$Q_{0.25}$=First quartile; $Q_{0.5}$=Median; $Q_{0.75}$=Third quartile; [c]SE=Standard error

costs within the segment) "maternity and infancy" (70.8%), "elderly" (62.8%) and "people affected by more than one CD" (57.1%). Within the latter segment, hospital costs represented the main determinant of cost in each sub-group identified (details not shown). The lowest hospital costs were generated in the "healthy people" segment (27.2%).

As regards drugs, the "maternity and infancy" and the "people suffering from an acute event" segments generated the lowest amount of costs (respectively 5.6% and 2.7%), while those with the highest costs were the "people affected by only one CD" (32.0%) and people with two CDs (31.8%) (data not shown). Outpatient examinations and medical visits were the major component of costs for "healthy people", and "people possibly affected by CD or at an early stage" (53.8% and 42.2% of total costs in the respective segments) reached the relatively lowest amount within the "people suffering from an acute event" segment (8.6%).

DISCUSSION

This is one of the first studies developing a population segmentation model of demand analysis applied on a data warehouse merging different complementary healthcare administrative datasets, routinely collected within a national system based on universal coverage. This not only allowed the application of the criteria suggested by Lynn *et al.*, but also their use in a perspective of continuous monitoring of some indices of the population needs, their costs and their distribution. Actually the administrative datasets can be used for the periodic and relatively easy survey of quite a number of items – particularly hospital admissions, drug prescriptions, outpatient medical specialist visits and diagnostic tests – describing the health demand of very large populations in a definite time and setting. In particular our study is referred to the middle of the last decade, just before of the economical crisis, in an urban, highly industrialized and well off area North of Milan.

At the time of the study the Lombardy health care data warehouse was not fully implemented and some important data about costs – i.e. the often very expensive drugs provided directly by the hospitals (to treat cancer, coagulation and rheumatic diseases, etc.) – were not available, although not limiting our analysis of the distribution of

TABLE 3

DEMOGRAPHIC CHARACTERISTICS AND COSTS OF "PEOPLE AFFECTED BY MORE THAN ONE CHRONIC DISEASE" SEGMENT, STRATIFIED BY CHRONIC CONDITIONS			
CHRONIC DISEASES	SUBJECTS N (%)	MALE %	AGE $Q_{0.5}$ $(Q_{0.25}$-$Q_{0.75})^A$
Hypertension	98 063 (73.36)	47.88	69 (62 - 76)
Anticoagulant treatment	34 479 (25.79)	56.84	70 (63 - 77)
Dyslipidemia	31 843 (23.82)	53.78	66 (59 - 73)
Diabetes mellitus	28 274 (21.15)	52.81	69 (61 - 75)
Gastroenteric diseases	26 063 (19.50)	47.84	68 (59 - 75)
Neurological diseases	23 672 (17.71)	43.99	66 (53 - 75)
Malignant tumours	22 378 (16.74)	50.42	69 (60 - 75)
Non-ischemic heart diseases	17 854 (13.36)	55.86	73 (65 - 78)
Endocrine disorders	16 107 (12.05)	21.34	64 (54 - 72)
Chronic obstructive pulmonary diseases	15 353 (11.49)	54.61	70 (61 - 77)
Ischemic heart diseases	14 942 (11.18)	69.76	70 (63 - 77)
Minor psychiatric disorders	12 252 (9.17)	31.05	66 (54 - 74)
Cerebrovascular diseases	10 024 (7.50)	56.00	73 (66 - 79)
Major psychiatric disorders	10 009 (7.49)	42.01	53 (38 - 66)
Peripheral angiopathies	7 538 (5.64)	50.08	68 (59 - 74)
Rheumatic illnesses	6 679 (5.00)	27.97	70 (62 - 75)
Treatment with antianemic drugs	5 749 (4.30)	34.21	65 (48 - 75)
Chronic heart failure	5 500 (4.11)	55.53	75 (68 - 80)
Treatment with NSAIDs	5 293 (3.96)	32.06	70 (62 - 76)
Hepatic disorders	4 188 (3.13)	60.82	64 (53 - 72)
Nephropathies	3 497 (2.62)	64.37	73 (65 - 79)
Treatment with glucocorticoid drugs	2 989 (2.24)	49.05	65 (54 - 72)
End-stage renal diseases	1 889 (1.41)	56.38	65 (55 - 73)
HIV/AIDS	674 (0.50)	74.63	44 (40 - 56)
Organ transplantation	582 (0.44)	65.12	56 (45 - 64)
Inherited genetic defects or congenital metabolic diseases	437 (0.33)	51.03	49 (30 - 65)

[a] $Q_{0.25}$=First quartile; $Q_{0.5}$=Median; $Q_{0.75}$=Third quartile

health and disease. The cost of the above mentioned CD may be underestimated (the use of very costly medicines out of the hospital was rather low), but is probably more accurate than the esteemed percentages and costs reported in other studies [4]. Now the healthcare data warehouse is fully implemented and the value of the segmentation method here proposed may be further verified. Also data related to the cost of general practitioners, community psychiatric services, long-term care are not considered, because they are paid not on a fee for service, but on a capitation basis, so not influencing the distributions of costs and diseases. Finally it must be taken into account that, in Italy, about 25% of health costs are paid directly out of pocket by the citizens, in form of prescription charges, private insurances,

FIGURE 1

PREVALENCE FOR EACH CONDITION IDENTIFIED IN THE SEGMENT "PEOPLE AFFECTED BY ONLY ONE CHRONIC DISEASE"
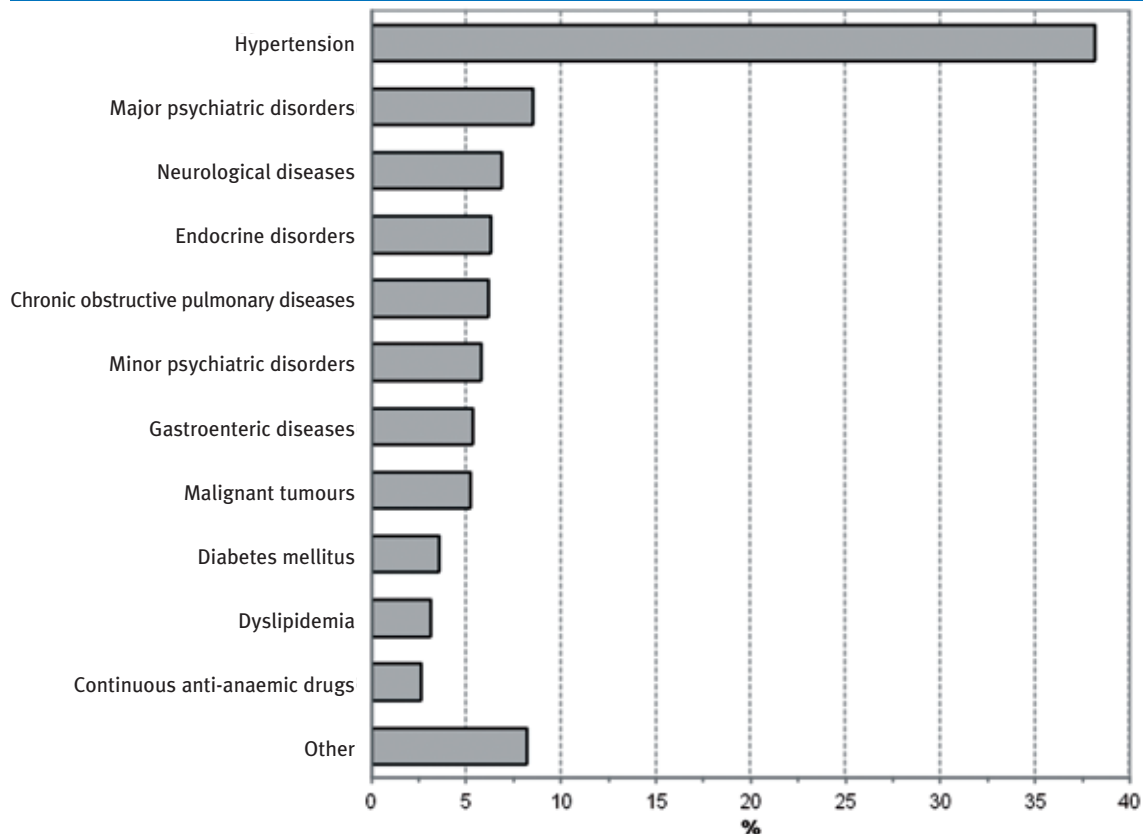
Note. The bar "Other" includes chronic diseases with a prevalence between 0% and 2%: chronic heart failure, nephropathies, end-stage renal diseases, organ transplantation, genetic defects and congenital diseases, continuous glucocorticoid therapy, ischemic heart diseases, HIV/AIDS, cerebrovascular diseases, hepatic disorders, non-ischemic heart diseases, continuous NSAID therapy, rheumatic illnesses, continuous anticoagulant treatment, peripheral angiopathies

dentistry, cosmetic medicine, and totally private services. These costs are not computed in our analysis, which only evidences a 6% of people unknown to the RHS, probably healthy, and probably paying everything for themselves.

Our data showed that 74% of healthcare expenditure was determined by the 30% of the population, because affected by a CD. As expected, the most expensive segment was the one involving subjects with several CDs, with a positive relationship between the number of conditions and health expenditure, mostly attributable to hospitalizations. The high prevalence of chronic conditions and related costs are due to the ageing of the population and the utilization of more effective and expensive treatments which improve life expectancy. It is important to know which are the prevalent diseases and with the joint effort

of scientist and politicians to realize the more efficient health care programs aimed at slowing down disease progression and shortening the disability period are required [24]. Data here presented do not have the precision of epidemiologic or clinical survey, but give substantial indication for future research in these fields.

As expected, among people affected by one CD, the most prevalent condition was hypertension, and our data showed that psychiatric disorders are an increasing pathological condition in the studied community and probably in similar settings, typical of developed countries.

It should be noted that "healthy people", although they had the lowest per-capita cost, ranked third for total expenditures, mainly for medical visits and examinations. This is

FIGURE 2

DETERMINANTS OF THE HEALTHCARE COST IN THE SEGMENTS

■Hospitalizations ■Drugs □Visits/Examinations
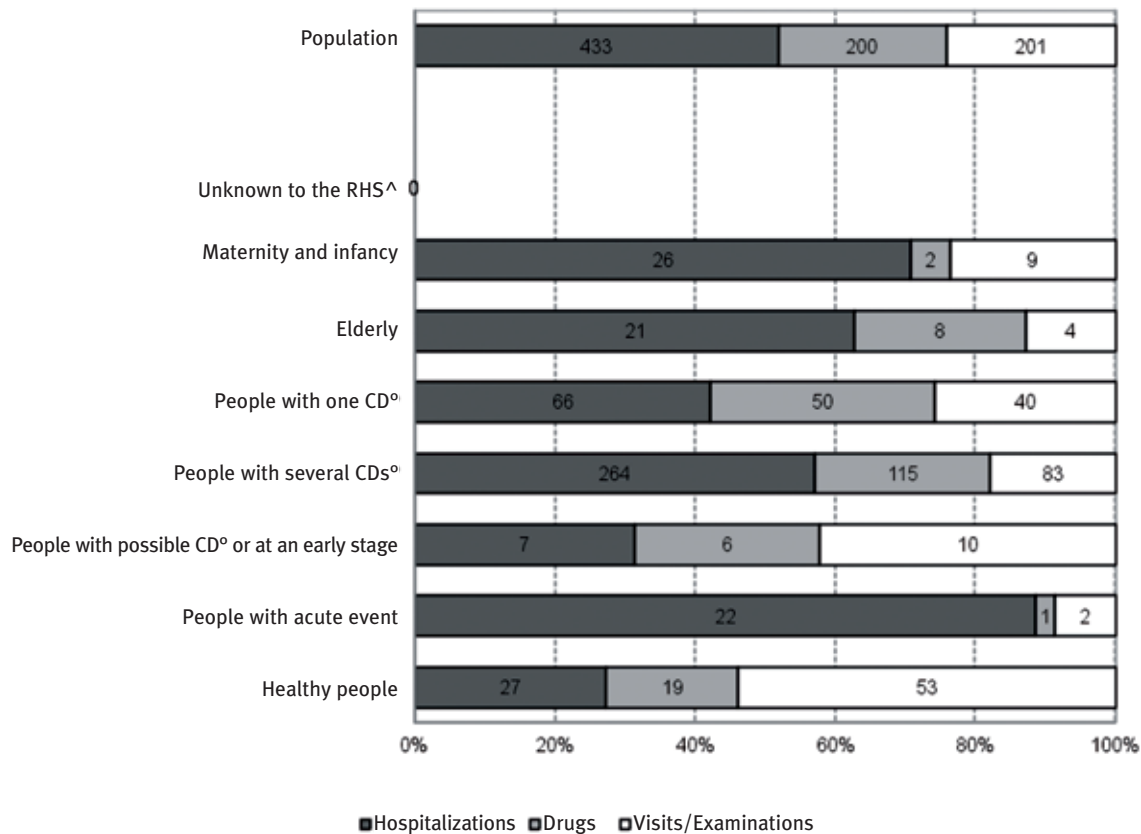
^RHS=Regional Health System; °CD=Chronic disease
Note. Numbers reported in each bar represent Euros (in million) spent by the Regional Health System, while the length of each bar represents the percentage of determinants within each segment

a segment that can be carefully considered in order to improve prevention and savings. Analyzing the sub-groups of this segment in detail, a considerable proportion of apparently healthy subjects consumed resources as well as patients affected by CDs.

In the whole population, drug expenditure accounted for about 24% of the total costs. This value is quite similar to that shown by OECD Health Data for Italy in 2005 and it is above the OECD average [25]. Programs could be implemented to reduce pharmaceutical costs, considering they are made up of two components: price and quantity of the product. Strategies to reduce both these elements may include: interventions enforcing the diffusion of generic drugs instead of brand named ones and programs involving patients and physicians to reduce inappropriate pharmaceutical consumption and prescriptions, in particular for CDs [26].

Six percent of the population was classified as "people possibly affected by CD or at an early stage". In addition to subjects with recently diagnosed (early stage) and unconfirmed chronic conditions, this segment of the population may include people with a low compliance to pharmaceutical treatment or with a drug under-prescribed by the physician. The ambiguities of this segment suggest a hypothetical inappropriateness of care. Efforts should be made to correctly classify people belonging to this "ambiguous" segment.

In the segment "people suffering from an acute event", almost 90% of costs were attributable to unavoidable hospital admissions. Although the total cost was relatively lower than the other segments since these subjects represent less than 1% of the population, it may justify investments towards programs able to improve first aid interventions and to avoid

these acute events through the prevention of work, domestic and road accidents.

In conclusion, even if this survey could not provide accurate and definitive data on healthcare demand and costs, which can only be estimated in relation to the time and the socio-cultural context, however the proposed model and data sources are promising for a more reliable evaluation of the impact of health programmes, and of the changes that are required.

References

[1] De Belvis AG, Ferrè F, Specchia ML, et al. The financial crisis in Italy: implications for the healthcare sector. Health Policy 2012; 106: 10-6

[2] Wright J, Williams R, Wilkinson JR. Development and importance of health needs assessment. BMJ 1998; 316: 1310-3

[3] Pencheon D. Managing demand: matching demand and supply fairly and efficiently. BMJ 1998; 316: 1665-7

[4] Lynn J, Straube BM, Bell KM, et al. Using population segmentation to provide better health care for all: the "Bridges to Health" model. Milbank Q 2007; 85: 185-208

[5] Sears JM, Krupski A, Joesch JM, et al. The use of administrative data as a substitute for individual screening scores in observational studies related to problematic alcohol or drug use. Drug Alcohol Depend 2010; 111: 89-96

[6] Crane SJ, Tung EE, Hanson GJ, et al. Use of an electronic administrative database to identify older community dwelling adults at high-risk for hospitalization or emergency department visits: the elders risk assessment index. BMC Health Serv Res 2010; 10: 338

[7] Miyata H, Hashimoto H, Horiguchi H, et al. Assessment of hospital performance with a case-mix standardized mortality model using an existing administrative database in Japan. BMC Health Serv Res 2010; 10: 130

[8] Mapel DW, Dutro MP, Marton JP, et al. Identifying and characterizing COPD patients in US managed care. A retrospective, cross-sectional analysis of administrative claims data. BMC Health Serv Res 2011; 11: 43

[9] Mantovani LG, Fornari C, Madotto F, et al. Burden of acute myocardial infarction. Int J Cardiol 2011; 150: 111-2

[10] Shahian DM, Silverstein T, Lovett AF, et al. Comparison of clinical and administrative data sources for hospital coronary artery bypass graft surgery report cards. Circulation 2007; 115: 1518-27

[11] Suissa S, Garbe E. Primer: administrative health databases in observational studies of drug effects-advantages and disadvantages. Nat Clin Pract Rheumatol 2007; 3: 725-32

[12] Schmitt J, Maywald U, Schmitt NM, et al. Cardiovascular comorbidity and cardiovascular risk factors in patients with chronic inflammatory skin diseases: a case-control study utilising a population-based administrative database. Ital Journal of Public Health 2008; 5: 187-193

[13] Baglio G, Sera F, Cardo S, et al. The validity of hospital administrative data for outcome measurement after hip replacement. Ital Journal of Public Health 2009; 6: 114-127

[14] De Portu S, Fanin R, Patriarca F, et al. The burden of multiple myeloma: assessment on occurrence, outcomes and cost using a retrospective longitudinal study based on administrative claims database. Ital Journal of Public Health 2011; 8: 325-30

[15] Fellegi IP, Sunter AB. A theory for record linkage. J Am Stat Assoc 1969; 64: 1183-210

[16] Newcombe HB, Kennedy JM, Axford SJ, James AP. Automatic linkage of vital records. Science 1959; 30: 954-9

[17] Newcombe HB, Kennedy JM. Making maximum use of the discriminatory power of identifying information. In: Communications of ACM 1962; 5: 563-6

[18] Newgard CD. Validation of probabilistic linkage to match de-identified ambulance records to a state trauma registry. Acad Emerg Med 2006; 13: 69-75

[19] Fornari C, Madotto F, Demaria M, et al. Record-linkage procedures in epidemiology: an Italian multicentre study. Epidemiol Prev 2008; 32: 79-88

[20] Regione Lombardia. Classificazione delle malattie, dei traumatismi, degli interventi chirurgici e delle procedure diagnostiche e terapeutiche, versione italiana della ICD-9-CM (International Classification of Disease - 9th revision – Clinical Modification). Milan: Regione Lombardia, 2002

[21] Regione Lombardia. Prestazioni di assistenza specialistica ambulatoriale erogabili nell'ambito del Servizio Sanitario Regionale e relative tariffe. Milan: Regione Lombardia, 2004

[22] WHO Collaborating Centre for Drug Statistics Methodology. Available from: http://www.whocc.no/ [Accessed May 14, 2012]

[23] Peterson AM, Nau DP, Cramer JA, et al. A checklist for medication compliance and persistence studies using retrospective databases. Value Health 2007; 10: 3-12

[24] Bodenheimer T, Fernandez A. High and Rising Health Care Costs. Part 4: Can Costs Be Controlled While Preserving Quality? Ann Intern Med 2005; 143: 26-31

[25] OECD Health Data 2010. Statistics and Indicators. Available from: http://www.oecd-ilibrary.org/social-issues-migration-health/data/oecd-health-statistics_health-data-en [Accessed May 14, 2012]

[26] Bodenheimer T. High and Rising Health Care Costs. Part 3: The Role of Health Care Providers. Ann Intern Med 2005; 142: 996-1002

*