APPENDIX

Patient Classification

FINGER is a patient classification system that estimates the individual risk of each person in 3 steps: classification of diseases into chronic disease groups (CDGs), assignment of a morbidity score, and assignment of scores related to other variables.

1.- Formation of chronic disease groups.

CDGs are based on freely available categorisation systems created by the American Agency for Healthcare Research and Quality (AHRQ). Specifically, the AHRQ generated a set of tools to support research in a broad range of fields related to healthcare management and planning under the Healthcare Cost and Utilization Project. These tools include the Chronic Condition Indicator (CCI)(s1) and the Clinical Classifications Software (CCS)(s2). The algorithm of the CCI categorises all ICD-9-CM diagnosis codes (more than 15,000) into two groups: chronic conditions (those with an expected duration of at least 12 months) and non-chronic conditions. The CCS groups classify all the ICD-9-CM codes into 284 clinically homogeneous groups.

Applying CCI and CCS to all the diagnoses (ICD-9-CM codes) from our population, we selected only those identified as chronic problems by the CCI, being a total of 6,259,725 (32.37%). Subsequently, the chronic conditions in the study database were classified using the CCS. Given that many of CCS groups correspond to acute conditions, our diagnoses were classified into only 181 of the CCS categories. Then, in order to achieve a description of morbidity that better suited our goals, the original CCS groups were modified in three cases:

- Creating a new independent category for obesity.
- Including all cases of heart failure in a single group, rather than dividing them into two groups as a function of whether individuals had hypertension.

- Distinguishing psychosis from other type of diagnoses related to depression.

Subsequently, we collapsed the remaining and modified CCS into CDGs so as to obtain a more manageable number of categories, applying the following criteria:

- Selected conditions were considered separately due to their particular importance, frequency or interest, in terms of health policy in general, and specifically at the Basque Health System (Osakidetza), regarding the development of specific care programmes (s3,s4). These were diabetes mellitus, obesity, hypertension, congestive heart failure (CHF) and metastasis. When feasible, this process was carried out directly using the CCS groups (for example, cases of diabetes mellitus with and without complications were grouped into one CDG). However, sometimes we needed to create a new group (Obesity) or to reclassify ICD 9-CM codes (for example, mood disorders, hypertension and CHF).
- For other conditions, the CCS categories were grouped as a function of body system, with three exceptions:
    - Mental health problems were divided into two groups: schizophrenia and psychosis; and other disorders.
    - Traumatic spinal cord injuries were included in the nervous system CDG, while other types of trauma were excluded.
    - A new CDG created to encompass eight miscellaneous CCS categories which had not been grouped (allergic reactions; complication of devices; complications of medical care; gangrene; lymphadenitis; poisoning by nonmedicinal substances; unspecified benign neoplasm; unclassified residual codes)

The stage of formation of CDGs finishes with a total of 27 CDGs generated. The distribution of the population across these groups is shown in Table 1.

2.- Allocation of scores to the CDGs and estimation of morbidity scores for each patient.

To avoid over-adjustment problems, we randomly divided the database into two subsets: the first was used for designing and calibrating the FINGER system, and the second exclusively for validation. With the calibration sample, we ran linear regression analysis using the presence or not of each CDG in the list of diagnoses of each patient in the first year (that is, a total of 27 dummy variables) as explanatory variables, and healthcare costs by this individual in the second year as the dependent variable. When malignant neoplasms and metastasis were diagnosed concurrently in the same person, only the latter condition was considered.

We provided one score for each CDG using the estimators obtained. As our goal is to relatively rank the risk of patients, we assigned 10 points to the CDG with the highest estimator (Metastasis), and proportionally lowered scores to the other CDGs. Our scores are rounded to integers. As a result, four CDGs had a score of zero, given that their modified estimators had values below 0.5. The regression was repeated, each time removing the CDG with the lowest score, until all remaining categories had a score greater than or equal to 1 (Table 2). Each patient obtained a chronic morbidity score that corresponded to the sum of the scores of CDGs for which they had a diagnosis. For each patient, any given CDG did not count more than once; that is, multiple diagnoses that all corresponded to the same CDG did not change an individual's score.

This way we deal with the trade-off between predictability and simplicity. We understand having only integers as scores reduces the predictive power of the model. However, our goal is to make the model simple enough to reduce the reluctance to its use by clinicians and check whether the price of simplicity is acceptable in terms of a small reduction of predictive power.

3.- Adding demographic and previous use information to get the final score for each patient.

In the last step we carried out new regression analyses with the calibration sample as follows:

$$ECost_I = \alpha + \beta * CDGscore_I + \sum_{j=1}^{20} \delta_{ij} * DEMO_{ij} + \gamma * HOSP_i + \rho * EMERG_i + \eta * DIALISIS_i + \varepsilon_i$$

where the explanatory variables were:

- $CDGscore_I$: individual CDG score calculated in stage 2.
- $DEMO$:Sex and age group: we consider 10 age groups (14 to 17; 18 to 34; 35 to 44; 45 to 54; 55 to 64; 65 to 69; 70 to 74; 75 to 79; 80 to 84; and ≥85 years). Each of these groups is subdivided into male and female.
- Number of hospital admissions, excluding those with a main diagnosis on discharge of obstetric or traumatic conditions. We consider five levels depending on the number of hospital admissions: 0, 1, 2, 3, and 4 or more.
- Number of emergency department visits per year: 0; 1 or more.
- Inclusion on the chronic dialysis programme.

On the basis of this analysis and clinical criteria and recalling that we sought to obtain a system that was simple and easy to use, we decided to assign the scores that are listed in Table 2. In this case, we took as reference the estimator for 1 point in CDGscore and proportionally estimated the scores to the other variables. Also we rounded such scores to integers. The differences between sex groups (adjusted by age) only were higher than 0.5 in people over 70 years old. By agreement of the researchers' team the score of dialysis was lowered due to its extreme value, and we conferred it 10 points (the same quantity that "Metastasis").

Characteristics of the population of study. Comparisons between calibration and validation samples.

Table S1

	Calibration		Validation		
	n	**%**	**n**	**%**	**P-value**
Total	973,400	50.0%	973,484	50.0%	
Sex (Male)	474,420	48.7%	474,739	48.8%	0.69
Chronic disease groups (CDGs)					
Infectious and parasitic diseases	2,534	0.26%	2,599	0.27%	0.37
Malignant neoplasms	16,795	1.73%	16,774	1.72%	0.91
Other endocrine disorders	28,774	2.96%	29,088	2.99%	0.19
Diabetes mellitus	42,501	4.37%	42,196	4.33%	0.28
Hyperlipidaemia	71,138	7.31%	72,046	7.40%	0.01
Diseases of the blood and blood-forming organs	4,537	0.47%	4,652	0.48%	0.23
Diseases of the nervous system	20,567	2.11%	20,586	2.11%	0.94
Diseases of the sense organs	33,317	3.42%	32,960	3.39%	0.16
Other heart diseases	10,759	1.11%	10,996	1.13%	0.11
Hypertension	98,408	10.11%	99,285	10.20%	0.04
Congestive heart failure	5,620	0.58%	5,756	0.59%	0.21
Stroke	10,036	1.03%	10,180	1.05%	0.31
Other vascular diseases	7,347	0.75%	7,295	0.75%	0.67
Other respiratory diseases	1,681	0.17%	1,558	0.16%	0.03
Chronic obstructive pulmonary disease	14,465	1.49%	14,689	1.51%	0.19
Asthma	21,968	2.26%	22,112	2.27%	0.50
Diseases of the digestive system	19,990	2.05%	19,774	2.03%	0.27
Diseases of the genitourinary system	44,089	4.53%	43,576	4.48%	0.07
Diseases of the skin and	10,041	1.03%	10,090	1.04%	0.74
Diseases of the musculoskeletal	55,179	5.67%	55,315	5.68%	0.69
Congenital anomalies	6,860	0.70%	6,934	0.71%	0.54
Other mental illnesses	61,976	6.37%	62,431	6.41%	0.19
Alcohol and substance abuse	5,581	0.57%	5,557	0.57%	0.82
Schizophrenia and psychosis	4,789	0.49%	4,842	0.50%	0.60
Metastasis	1,341	0.14%	1,327	0.14%	0.80
Obesity	9,725	1.00%	9,778	1.00%	0.71
Other miscellaneous conditions	6,001	0.62%	6,018	0.62%	0.89
Dialisis previous year	451	0.05%	469	0.05%	0.58
2nd Year					
Admisssions 1+	33,869	3.48%	33,884	3.48%	0.96
Prolongued stay	10,053	1.03%	10,288	1.06%	0.10
Very prologued stay	6,890	0.71%	7,072	0.73%	0.12
Death	9,155	0.94%	9,392	0.96%	0.08

Characteristics of the population of study. Comparisons between calibration and validation samples.

Table S2

	Calibration		Validation		
	Mean	**sd**	**Mean**	**sd**	**P-value**
Age	48.08	18.88	48.11	18.89	0.20
Emergency visits in previous year	0.26	0.74	0.26	0.75	0.80
Hospital admissions in previous year	0.09	0.42	0.09	0.42	0.52
FINGER Scores					
Age and sex	0.92	1.14	0.92	1.14	0.33
Diagnoses	0.78	1.69	0.78	1.69	0.26
Age & sex + diagnoses	1.70	2.39	1.70	2.39	0.21
Age & sex + diagnoses + resource use in previous year	2.05	3.00	2.06	3.01	0.24
Healtcare cost (2nd year)	1,119.97	3262.06	1,126.37	3273.78	0.17

s1. HCUP Chronic Condition Indicator (CCI). Healthcare Cost and Utilization Project (HCUP). 2009. Agency for Healthcare Research and Quality, Rockville, MD. http://www.hcup-us.ahrq.gov/toolssoftware/chronic/chronic.jsp. Accessed July 20, 2017.

s2. HCUP Clinical Classifications Software (CCS) for ICD-9-CM. Healthcare Cost and Utilization Project (HCUP). 2006-2009. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov/toolssoftware/ccs/ccs.jsp. Accessed July 20, 2017.

s3. Departamento de Sanidad y Consumo del Gobierno Vasco. Estrategia para afrontar el reto de la cronicidad en Euskadi. Vitoria-Gasteiz: Editado por Eusko Jaularitza - Gobierno Vasco; 2010. . Accessed July 20, 2017. Available from: https://www.osakidetza.euskadi.eus/r85-skorga01/es/contenidos/informacion/estrategia_cronicidad/es_cronicos/estrategia_cronicidad.html

s4. Gutiérrez M, Berraondo I, Bilbao J, Gorroñogoitia A, Gutiérrez M, Gutiérrez B, et al. Análisis y desarrollo del plan de actividades preventivas de Atención Primaria. Revisión de la oferta preferente y del programa informático de soporte (PAP). Investigación Comisionada. Vitoria-Gasteiz . Departamento de Sanidad y Consumo. Gobierno Vasco, 2011 . Informe nº Osteba D-11-01. Accessed July 11, 2017. Available from: http://www.osakidetza.euskadi.eus/contenidos/informacion/2011_osteba_publicacion/es_def/adjuntos/D-11-01%20Plan%20de%20actividades%20preventivas.pdf