



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA



ESCUELA TÉCNICA
SUPERIOR INGENIEROS
INDUSTRIALES VALENCIA



2017 Spanish Stata Users Group meeting

Madrid | 19 October

Instituto de Salud Carlos III

Construction and validation of a predictive model for the identification of complex chronic patients

Silvia Badal Carsi, Ruth Usó, David Vivas, Alexander Zlotnik
Universidad Politécnica de Valencia, Universidad Politécnica de Madrid

ÍNDICE

- I. INTRODUCCIÓN
- II. ANTECEDENTES
- III. OBJETIVOS
- IV. METODOLOGÍA
- V. RESULTADOS
- VI. DISCUSIÓN
- VII. CONCLUSIONES

I. INTRODUCCIÓN

I. INTRODUCCIÓN

- **Colaboración** entre el grupo de investigación en economía y gestión de salud de **INECO** de la UPV y la dirección del **Hospital Clínico Universitario de Valencia**.
- Los directivos del Hospital Clínico detectaron la necesidad de **mejorar** el sistema de identificación de Pacientes Crónicos Complejos (PCC).
- Diseñar un **algoritmo** de identificación de PCC en la CV que mejore los resultados del sistema existente.
- Los PCC son de gran importancia debido a que **consumen** la gran cantidad de **recursos** siendo un porcentaje muy pequeño de la población. Gracias a una buena identificación se podrá estudiar su comportamiento y aplicarles tanto una **gestión** como unos **cuidados** más adecuados con el objetivo de mejorar el sistema sanitario y la calidad de vida de los pacientes.

2. ANTECEDENTES

EFFECTO E INFLUENCIA DE LA
CRONICIDAD EN EL SNS

EVOLUCIÓN DEL GASTO SANITARIO EN
ESPAÑA Y CV


MODELOS DE AJUSTE DE RIESGO EN
SALUD: CLINICAL RISK GROUPS (CRG).

II. ANTECEDENTES

2.3 Modelos de ajuste de riesgo

- **Herramienta** para estratificar a la población según unos indicadores determinados.
- Su objetivo es aplicar **estrategias** sanitarias más efectivas según la morbilidad de la población.

Los Clinical Risk Groups (CRG)

- Establece un estado de salud mutuamente excluyente a cada paciente basándose en la **historia clínica y las características demográficas**
- ACRG3 con 9 Estados de salud y 6 niveles de gravedad. 

Estado de salud para el nivel de agregación ACRG3	
1	Estado sano
2	Enfermedad aguda significativa
3	Enfermedad crónica menor única
4	Enfermedad crónica menor en múltiples sistemas orgánicos
5	Enfermedad dominante o crónica moderada única
6	Enfermedad significativa crónica en múltiples sistemas orgánicos
7	Enfermedad dominante crónica en tres o más sistemas orgánicos
8	Neoplasias dominantes, metastásicas y complicadas
9	Necesidades sanitarias elevadas.

III. OBJETIVOS

HIPÓTESIS

OBJETIVO GENERAL

OBJETIVOS ESPECÍFICOS

III. OBJETIVOS

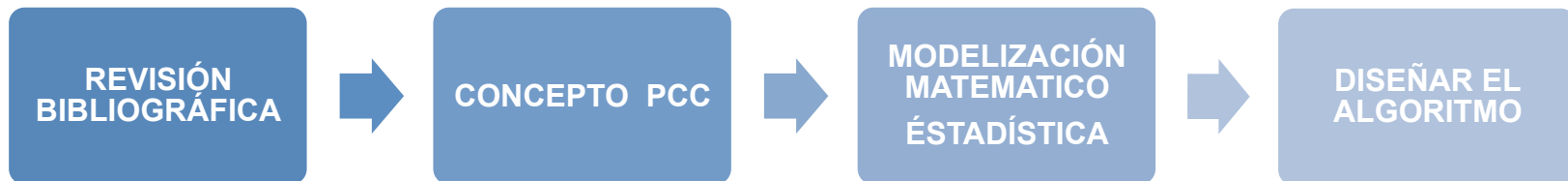
3.1 Hipótesis

Los pacientes crónicos complejos (PCC) se pueden identificar usando información digital a partir de las siguientes variables: edad, sexo, gasto farmacéutico ambulatorio, índice de riesgo clínico (CRG) y utilización de recursos sanitarios.

3.2 Objetivo general

Diseño de un algoritmo de identificación de pacientes crónicos complejos usando la información de CRGs y utilización de recursos.

3.2 Objetivos específicos



IV. METODOLOGÍA

DISEÑO
VARIABLES DEL ESTUDIO
FUENTES DE INFORMACIÓN
PROGRAMAS INFORMÁTICOS
ANÁLISIS ESTADÍSTICO

IV. METODOLOGIA

4.1 Diseño

4.2 Variables del estudio

4.3 Fuentes de información

4.4 Programas informáticos

4.5 Análisis estadístico

4.6 Plan de trabajo

IV. METODOLOGIA

4.1 Diseño

4.2 Variables del estudio

4.3 Fuentes de información

4.4 Programas informáticos

4.5 Análisis estadístico

4.6 Plan de trabajo

IV. METODOLOGIA

Estudio de ámbito poblacional, observacional, descriptivo, analítico, retrospectivo y de corte transversal.

- ***Ámbito del estudio:*** CV incluyendo los 24 departamentos de salud.
- ***Periodo de estudio:*** de enero del 2015 hasta diciembre del 2015.
- ***Sujetos de estudio:*** 4.709.115 ciudadanos registrados en el sistema de información poblacional de la CV, entre enero del 2015 y diciembre del 2015.
- ***Criterios de inclusión y exclusión de los sujetos de estudio:*** Se incluyen en el estudio los ciudadanos dados de alta en el SIP con médico asignado (empadronados más de un mes) de la CV.

4.2 V

4.3 Fu

IV. METODOLOGIA

4.1 Diseño

4.2 Variables del estudio

4.3 Fuentes de información

4.4 Programas informáticos

4.5 Análisis estadístico

IV. METODOLOGIA

4.1 Diseño

Estudio de ámbito poblacional, observacional, descriptivo, analítico, retrospectivo y de corte transversal.

- **Ámbito del estudio:** CV incluyendo los 24 departamentos de salud.
- **Periodo de estudio:** de enero del 2015 hasta diciembre del 2015.
- **Sujetos de estudio:** 4.709.115 ciudadanos registrados en el sistema de información poblacional de la CV, entre enero del 2015 y diciembre del 2015.
- **Criterios de inclusión y exclusión de los sujetos de estudio:** Se incluyen en el estudio los ciudadanos dados de alta en el SIP con médico asignado (empadronados más de un mes) de la CV.

4.2 Variables del estudio

- **Datos socio-demográficos:** clave de anonimización de pacientes, edad y sexo.
- **Datos de utilización del sistema sanitario o carga asistencial:** número de contactos en atención primaria, número de contactos en urgencias y número de ingresos hospitalarios.
- **Datos sobre el CRG**
- **Datos de problemas relacionados con los medicamentos (PRM)**
- **Datos de gasto farmacéutico:** importe por paciente de farmacia ambulatoria.

IV. METODOLOGÍA

4.3 Fuentes de información

Sistemas de información sanitarios de la CV disponibles en la Oficina de Farmacoeconomía de la Dirección General de Farmacia y Productos Sanitarios (DGFyPS) de la Conselleria de Sanitat Universal i Salut Pública.

4.4 Programas informáticos

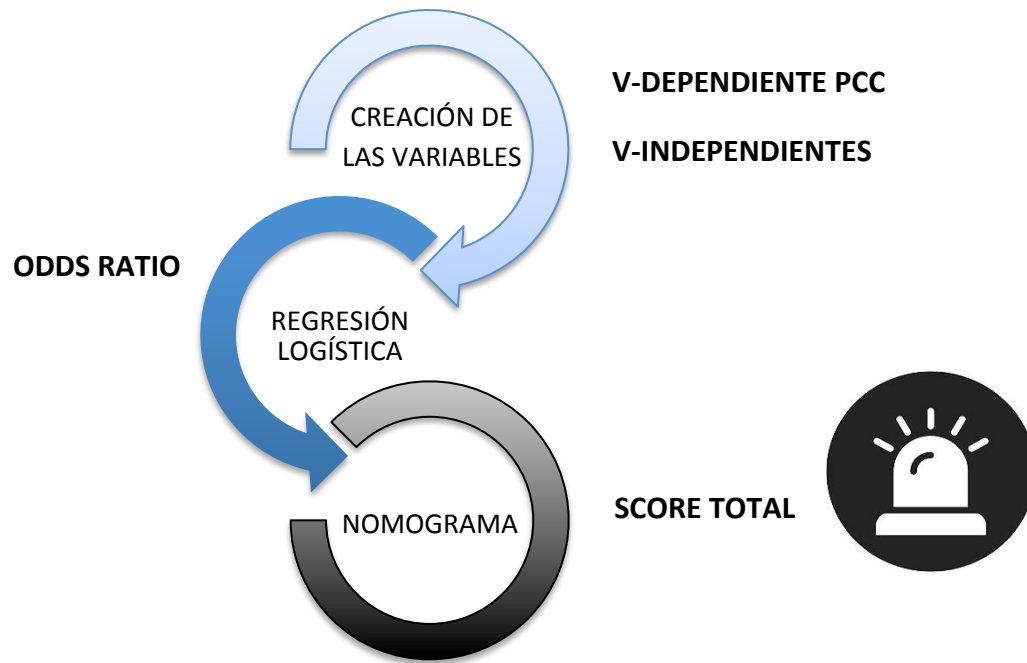
El programa **StataSE** versión 14 para el análisis descriptivo y estadístico de los datos y el programa **Microsoft Excel 2016** para expresar los gráficos, construir las tablas y presentar la información numérica.

4.6 Análisis estadístico

1. **Análisis descriptivo** de las variables
2. Desarrollo de un **modelo predictivo** mediante regresión logística binaria.
3. Realización de un **nomograma** de discriminación.

DISEÑO DEL ALGORITMO

Síntesis del algoritmo de identificación



V. RESULTADOS

DEFINICIÓN DE UN PCC
ANÁLISIS DE LA POBLACIONAL
DISEÑO DEL ALGORITMO

5.1 DEFINICIÓN DE UN PCC

- Un PCC es aquella persona cuya **GESTIÓN CLÍNICA ES ESPECIALMENTE DIFÍCIL**.
- Son personas con **COMORBILIDAD, POLIMEDICACIÓN Y EDADES AVANZADAS**.
- Una **ENFERMEDAD CRÓNICA** muy grave y en curso o varias enfermedades crónicas. Además de otras **SITUACIONES COMPLEJAS**, por ejemplo, enfermedades minoritarias o disfunciones intelectuales.
- La disminución de autonomía del paciente y la polimedicación que puede derivar en **EFFECTOS SECUNDARIOS E INTERACCIONES ENTRE FÁRMACOS**.
- **USO RECURRENTE DE LOS SERVICIOS SANITARIOS** con una **FRECUENCIA ALTA DE INGRESOS DE LARGA DURACIÓN**.
- Su atención requiere de **VARIOS PROFESIONALES** y a su vez un **GRAN NÚMERO DE PRUEBAS DIAGNÓSTICAS**.

5.2 ANÁLISIS POBLACIONAL

Distribución de la población por estado de salud y gravedad.

Para este estudio se ha utilizado una base de datos de la CV con un total de **4.709.115 pacientes**. Los pacientes del estudio han sido clasificados en función de la morbilidad mediante el software CRG.

El nivel de agregación utilizado en el **ACRG3 del modelo P** o predictivo del CRG donde se estructura a la población por estado de salud y nivel de gravedad.

Distribución de toda la población por estado de salud.



- Estado sano
- Enfermedad aguda significativa
- Enfermedad crónica menor única
- Enfermedad crónica menor en múltiples sistemas orgánicos
- Enfermedad dominante o crónica moderada única
- Enfermedad significativa crónica en múltiples sistemas orgánicos
- Enfermedad dominante crónica en tres o más sistemas orgánicos
- Neoplasias dominantes, metastásicas y complicadas
- Necesidades sanitarias elevadas

5.2 ANÁLISIS POBLACIONAL

Distribución del uso de recursos por estado de salud y gravedad.

- Se ha estudiado la distribución de los contactos en **atención primaria, ingresos hospitalarios y urgencias**.
- Los pacientes con un mayor uso de los recursos sanitarios se localizan en los grupos de mayor **morbilidad y complejidad** de los ACRG, llegando a superar veinticuatro veces la media de la población en el caso de los ingresos hospitalarios.
- Un mayor uso de recursos implica un mayor gasto sanitario

MEDIA DE INGRESOS HOSPITALARIOS POR ESTADO DE SALUD Y GRAVEDAD EN LA CV EN 2015

Estado de Salud	Nivel de gravedad							TOTAL
	0	1	2	3	4	5	6	
1 Estado sano	0,21	0,01						0,17
2 Enfermedad aguda significativa	0,44							0,44
3 Enfermedad crónica menor única		0,23	0,71					0,28
4 Enfermedad crónica menor en múltiples sistemas orgánicos		0,25	0,29	0,49	0,74			0,32
5 Enfermedad dominante o crónica moderada única		0,25	0,45	0,41	0,68	0,4	0,98	0,31
6 Enfermedad significativa crónica en múltiples sistemas orgánicos		0,36	0,49	0,58	0,76	0,94	1,56	0,5
7 Enfermedad dominante crónica en tres o más sistemas orgánicos		0,46	0,69	0,88	1,26	1,74	2,11	0,92
8 Neoplasias dominantes, metastásicas y complicadas		0,47	0,7	0,96	1,53	1,84		0,97
9 Necesidades sanitarias elevadas		0,42	0,48	0,81	1,02	1,57	2,33	0,75
TOTAL	0,24	0,21	0,49	0,58	0,85	0,93	1,74	0,29

5.2 ANÁLISIS POBLACIONAL

Distribución los PRM por estado de salud y gravedad.

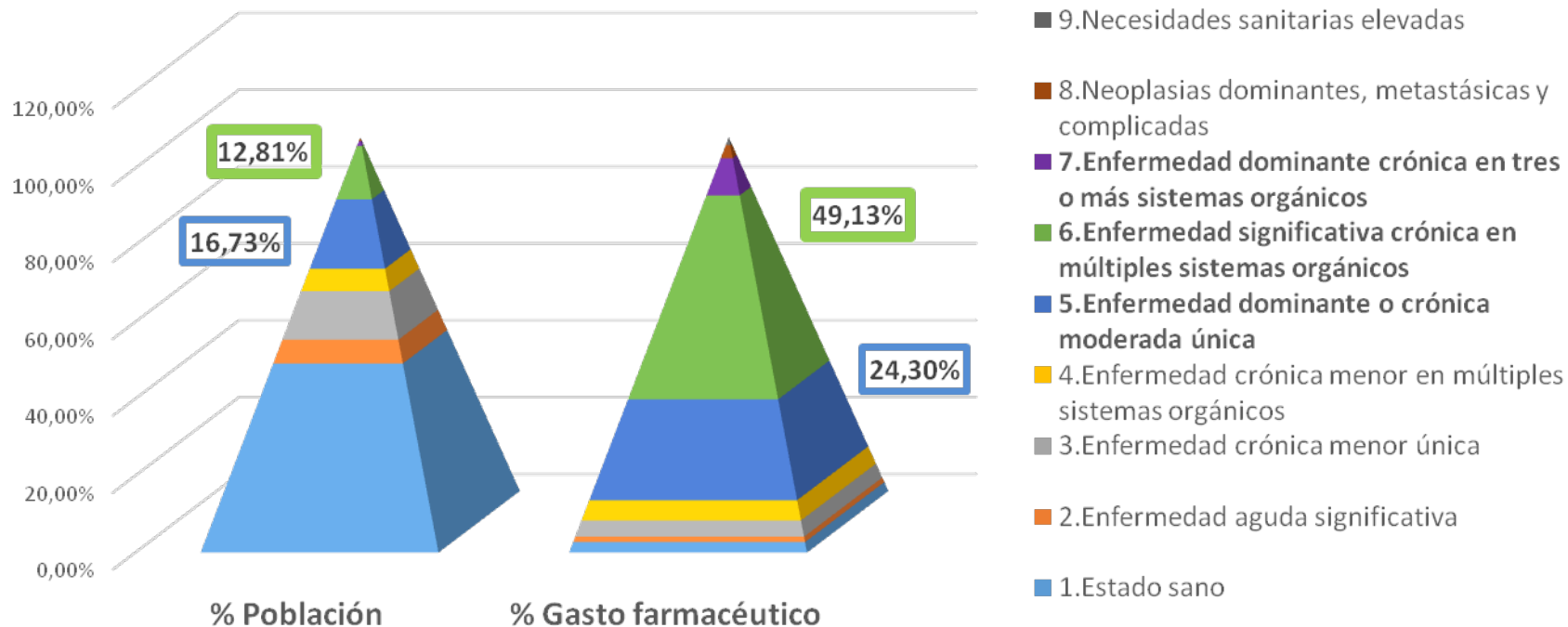
- PRM incluyen las alergias, las interacciones, posología, contraindicaciones, CIE-ATC, acontecimientos adversos y duplicidades.
- Conforme empeora el estado de salud del paciente el número de medicamentos que consume aumenta generando un mayor número de PRMs.

MEDIA DE URGENCIAS POR ESTADO DE SALUD Y GRAVEDAD EN LA CV EN 2015

Estado de Salud	Nivel de gravedad							TOTAL
	0	1	2	3	4	5	6	
1 Estado sano	0,04	0						0,03
2 Enfermedad aguda significativa	0,18							0,18
3 Enfermedad crónica menor única		0,27	0,34					0,28
4 Enfermedad crónica menor en múltiples sistemas orgánicos		0,71	1,52	1,36	2,16			1,05
5 Enfermedad dominante o crónica moderada única		1,38	2,29	3,83	6,2	5,74	4,84	1,8
6 Enfermedad significativa crónica en múltiples sistemas orgánicos		4,21	6,12	7,76	10,07	11,89	14,46	6,16
7 Enfermedad dominante crónica en tres o más sistemas orgánicos		10,3	12,13	14,09	16,05	18,25	20,69	13,81
8 Neoplasias dominantes, metastásicas y complicadas		2,14	3,38	5,49	8,93	10,18		5,29
9 Necesidades sanitarias elevadas		3,14	3,84	6,3	8,13	10,49	12,73	5,62
TOTAL	0,06	1,19	3,4	6,01	9,53	10,98	15,14	1,43

5.2 ANÁLISIS POBLACIONAL

Distribución del gasto farmacéutico por estado de salud y gravedad.

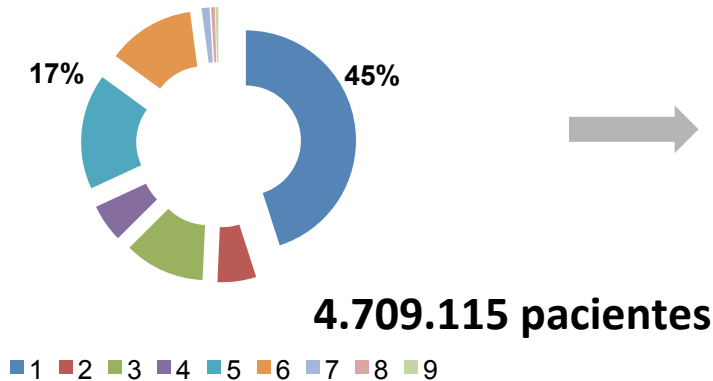


5.3 DISEÑO DEL ALGORITMO

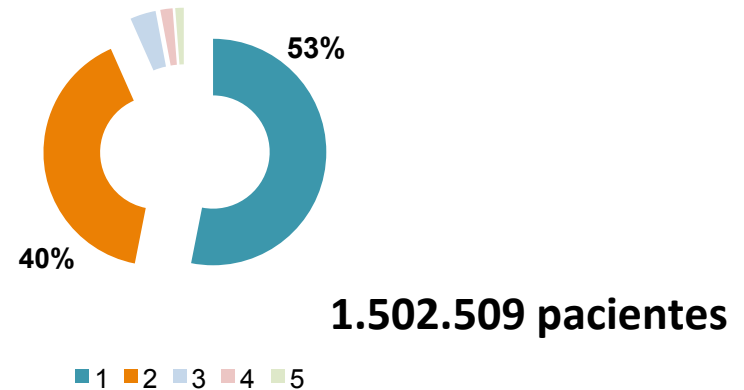
5.3.1 Selección de las variables

- No se incluirán las variables sociodemográficas, edad y sexo.
- Exclusión de los pacientes “no crónicos” del estudio.

Distribución de toda la población por estado de salud.



Distribución de la población crónica por estado de salud.



5.3 DISEÑO DEL ALGORITMO

5.3.1 Selección de las variables

Variable dependiente PCC

Se han clasificado como PCC a los pacientes que superan el percentil 95 (P95) de los recursos, PRM y coste, teniendo en cuenta a toda la población.



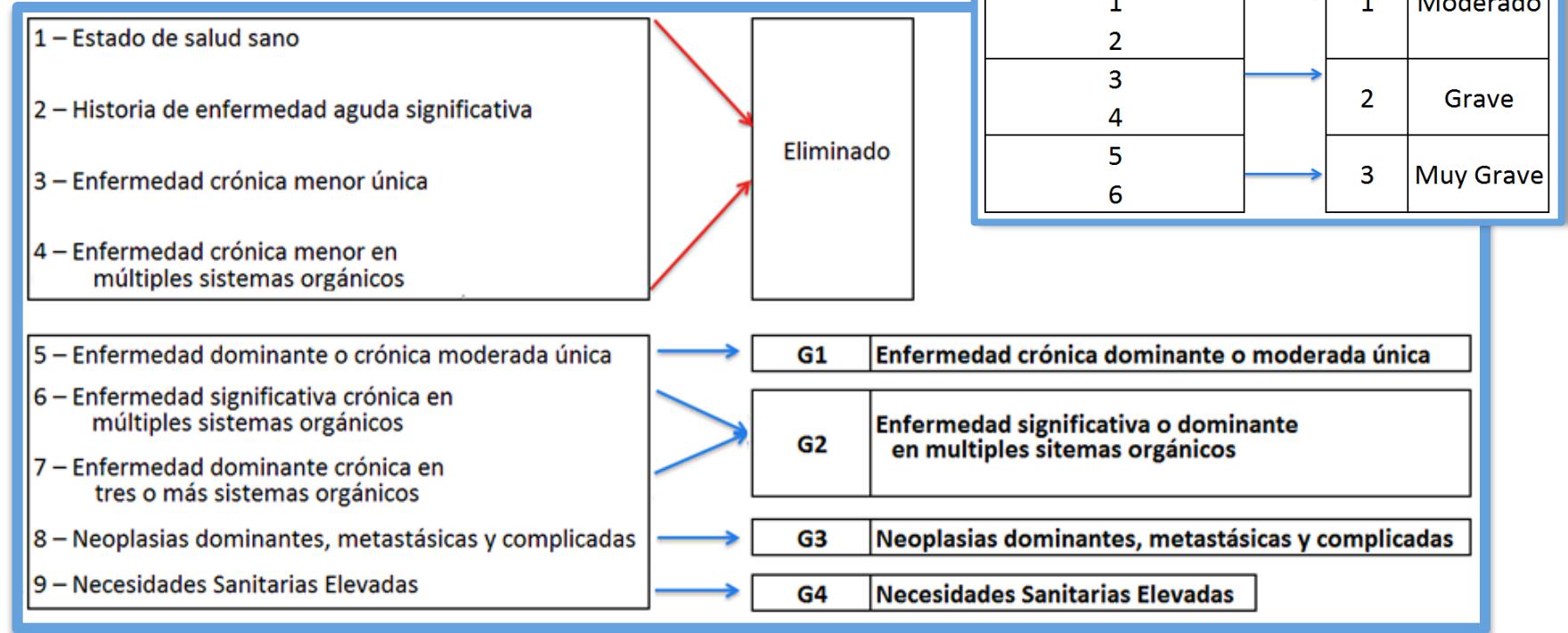
Variable independiente

Uso de recursos, los PRM, el coste y la clasificación CRG.

	Primaria	Hospital	Urgencias	PRM	Coste farmacia
Media	7	0	0	1	279
Percentil 95	24	1	2	9	1.492
Máximo	344	47	74	121	160.777

5.3 DISEÑO DEL ALGORITMO

5.3.1 Selección de las variables



5.3 DISEÑO DEL ALGORITMO

5.3.1 Selección de las variables

Variables finales

Variable **dependiente**: "V_PCC".

Variables **independientes**:

Continuas:

- Nº de contactos en atención primaria
- Nº de contactos en urgencias
- Nº de ingresos hospitalarios
- PRM
- coste farmacéutico.

Categorías



5.3 DISEÑO DEL ALGORITMO

5.3.2 Modelo de regresión logística (generación del modelo)

```
// Creamos la muestra train (75% de los datos)
set seed 1
g train = (runiform() < 0.75)
```

```
// Realizamos la regresión logística odds ratio.
logistic v_PCC i.G1_Cronica i.G2_Cronicas i.G3_Neoplasias i.G4_Necesidades
NUM_CONT_AMB NUM_ING_HOSP NUM_URG PRM_TOTAL IMPORTE_NUM if train
```

5.3 DISEÑO DEL ALGORITMO

5.3.2 Modelo de regresión logística (pruebas postestimación)

```
// TRAIN
// Área bajo la curva ROC; Test de bondad de ajuste; Tabla de clasificación
lroc if train, nograph
estat gof if train
estat classification if train

// TEST
// Las mismas pruebas pero para el 25% restante, el test
lroc if !train, nograph
estat gof if !train
estat classification if !train

//Generación del nomograma
nomolog if train
```

5.3 DISEÑO DEL ALGORITMO

5.3.2 Modelo de regresión logística

Debido a las dimensiones de la población, se realizarán las pruebas con una **muestra del 20% del total**.

A su vez, la regresión se aplicará a un conjunto denominado **train**, una muestra aleatoria del 75% de los datos, y después al conjunto **test**, el 25% restante.

V_PCC	Odds Ratio	Std. Err.	z	P>z	[95% Conf.Interval]
G1_Cronica					
1	0,013	0,013	-4,250	0	0,002 0,095
2	0	0	-1,180	0,236	0 1,316
3	1,000	(empty)			
G2_Cronicas					
1	0,303	0,077	-4,670	0	0,183 0,500
2	1,114	0,231	0,520	0,603	0,742 1,671
3	1,668	0,350	2,440	0,015	1,106 2,516
G3_Neoplasias					
1	0,191	0,139	-2,280	0,023	0,046 0,792
2	1,514	0,376	1,670	0,095	0,930 2,464
3	2,352	0,806	2,500	0,013	1,202 4,602
Nº contactos atención primaria	1,028	0,002	13,950	0	1,024 1,031
Nº contactos hospitalarios	1,690	0,035	25,500	0	1,623 1,760
Nº de urgencias	1,488	0,021	27,660	0	1,447 1,531
PRM totales	1,090	0,004	24,680	0	1,083 1,098
Coste farmacéutico	1,000	0	6,730	0	1,000 1,000
constante	0	0	-35,820	0	0 0

*note: G1_Cronica != 0 predicts failure perfectly
 G1_Cronica dropped and 2163 obs not used

5.3 DISEÑO DEL ALGORITMO

5.3.2 Modelo de regresión logística

Post-estimación

- **Discriminación:** Área bajo la curva ROC.
- **Calibración:** Test de bondad de ajuste de Hosmer Lemeshow.
- **Sensibilidad y especificidad**

Post-etimación					
	Nº de observaciones	AUROC	Test de bondad de ajuste Prov>chi2	Sensibilidad	Especificidad
Train	223792	0.9963	0	24.24%	99.95%
Test	74546	0.9920	0	28.00%	99.94%

5.3 DISEÑO DEL ALGORITMO

5.3.3 Nomograma de discriminación

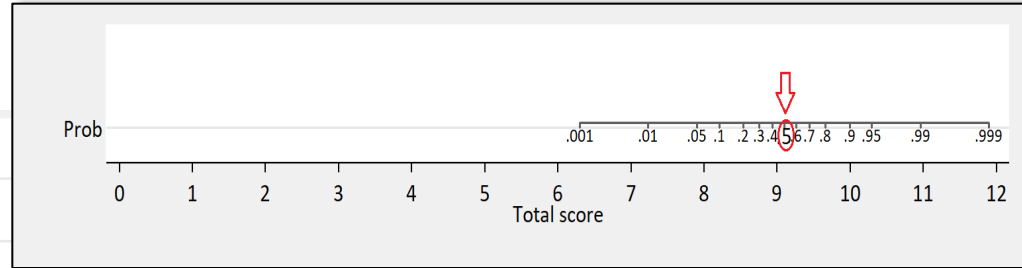
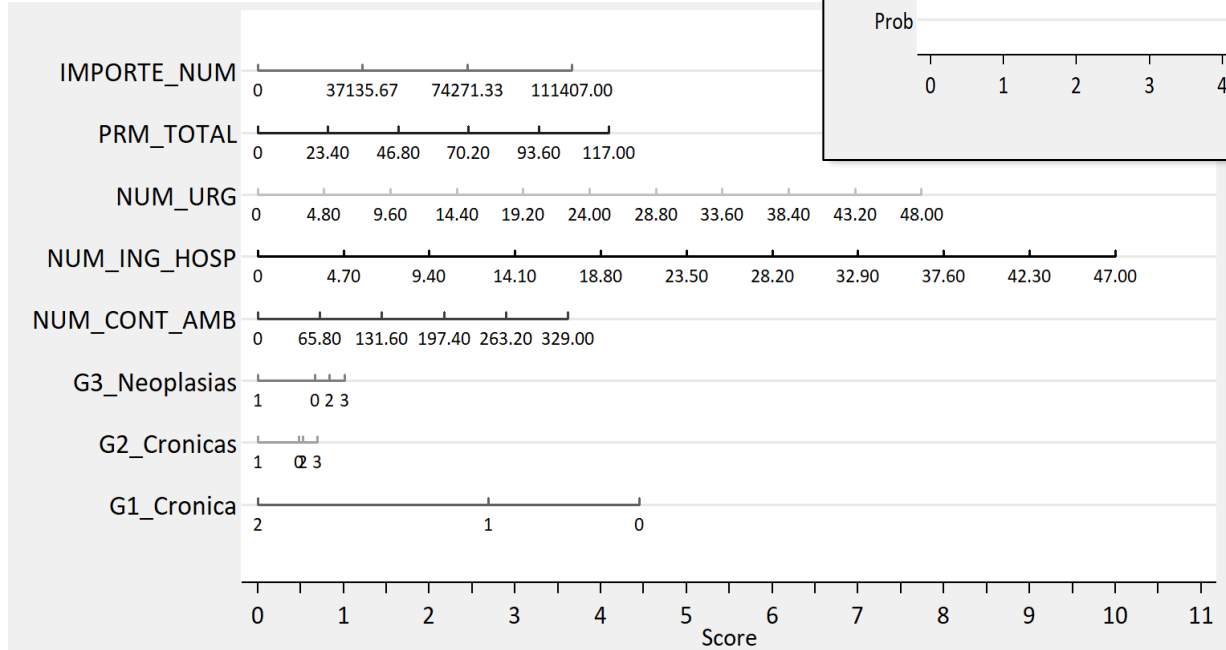
Un nomograma es una **representación gráfica de los odds ratio**, coeficientes de la regresión. A la hora de representarlos, sigue los siguientes pasos:

1. Establece una puntuación para todas las variables del modelo
2. Obtiene una puntuación total sumando todas las puntuaciones de las variables.
3. Obtiene la probabilidad de evento a partir de la puntuación total.

En nuestro caso, dado que la calibración no es adecuada, se utiliza el nomograma para discriminar entre PCC o no. Para ello, marcamos como límite entre sí y no la puntuación total de 0.5.

5.3 DISEÑO DEL ALGORITMO

5.3.3 Nomograma de discriminación



LA PUNTUACIÓN TOTAL
CORRESPONDIENTE A LA
PROBABILIDAD 0.5% ES **9,3**

VI. DISCUSION

Limitaciones del estudio

Aplicaciones futuras

VI. DISCUSION

Limitaciones del estudio

- Las fuentes de información se corresponden con la situación real de Conselleria de Sanitat, pueden presentar **errores en la codificación** de los diagnósticos, procedimientos médicos y prescripciones farmacéuticas.
- No se disponían de variables referentes a situación social y de dependencia

VI. DISCUSION

Aplicación futura

- **Alerta asistencial de cronicidad** en la historia clínica electrónica (HCE, ABUCASIS). Divide la cronicidad en niveles:
 - Nivel 1: Estados de salud 1 y 2.
 - Nivel 2: Estados de salud 3, 4 y 5.
 - Nivel 3: Estados de salud 6 gravedad 1, 2, 3 y 4.
 - **Nivel 4:** Estado de salud 6 gravedad 5 y 6, Estados 7, 8 y 9 → Paciente Crónicos Complejos
- La aplicación futura del algoritmo propuesto en este trabajo es formar parte de la alerta asistencial de cronicidad reformando el sistema actual para poder obtener mejores resultados.

VII.CONCLUSIONES

VII. CONCLUSIONES

1. El sistema de identificación de cronicidad aplicado por la Consellería de Sanitat no satisface los requerimientos de los usuarios y es necesario una nueva herramienta más potente.
2. Los **Pacientes Crónicos Complejos (PCC)** representan un porcentaje muy bajo de la población y consumen un alto porcentaje de los recursos sanitarios. Son personas de edades avanzadas que sufren una enfermedad crónica muy grave o varias enfermedades crónicas. Presentan además comorbilidades, pluripatología, polimedicación y edades avanzadas.
3. Se ha corroborado que un **aumento de la carga de morbilidad y gravedad** de las enfermedades supone un **incremento proporcional del consumo de recursos**

VII. CONCLUSIONES

4. A medida de **umenta la morbilidad** y el estado del paciente se agrava, el número de medicamentos aumenta simultáneamente provocando en un aumento en los **PRM** y el **gasto farmacéutico**
5. La **inclusión de nuevas variables** al modelo de identificación ha aumentado la precisión del sistema de identificación.
6. El algoritmo desarrollado mediante un **modelo de regresión logística** es validado para identificar PCC.
7. El **nomograma** es una herramienta de discriminación adecuada para determinar si un paciente es o no es PCC.
8. El algoritmo resultante supone una **mejora sobre el sistema** en uso.

MUCHAS GRACIAS

dvivas@upv.es