

*Estudio de los perfiles de progresión de la enfermedad renal
crónica en una población intervenida con un programa de
prevención secundaria*

JESÚS DAVID RAMOS MONTAÑA



UNIVERSIDAD NACIONAL DE COLOMBIA
FACULTAD DE CIENCIAS
DEPARTAMENTO DE ESTADÍSTICA
BOGOTÁ, D.C.
25 DE NOVIEMBRE DE 2016

*Estudio de los perfiles de progresión de la enfermedad renal
crónica en una población intervenida con un programa de
prevención secundaria*

JESÚS DAVID RAMOS MONTAÑA

TRABAJO FINAL PARA OPTAR AL TÍTULO DE
MAGÍSTER EN CIENCIAS - BIOESTADÍSTICA

DIRECTOR
CAMPO ELÍAS PARDO, PH.D.
DOCTOR EN ESTADÍSTICA

CODIRECTOR
RAFAEL MAURICIO SANABRIA, M.SC.
MAGÍSTER EN EPIDEMIOLOGÍA CLÍNICA Y GERENTE GENERAL DE ASUNTOS
CIENTÍFICOS DE RENAL THERAPY SERVICE



UNIVERSIDAD NACIONAL DE COLOMBIA
FACULTAD DE CIENCIAS
DEPARTAMENTO DE ESTADÍSTICA
BOGOTÁ, D.C.
25 DE NOVIEMBRE DE 2016

Título en español

Estudio de los perfiles de progresión de la enfermedad renal crónica en una población intervenida con un programa de prevención secundaria

Title in English

A study of the progression profiles of the chronic kidney disease in a population monitored with a secondary prevention program

Resumen: se combina el análisis armónico cualitativo (AAC) con el análisis factorial múltiple (AFM) para determinar distintas clases de trayectorias de pacientes con enfermedad renal crónica (ERC) e identificar variables clínicas y sociodemográficas asociadas a la evolución temporal de la enfermedad.

Abstract: qualitative harmonic analysis (QHA) and multiple factor analysis (MFA) are combined to determine different types of trajectories of patients with chronic kidney disease (CKD) and to identify clinical and sociodemographic variables associated with the temporal evolution of the disease.

Palabras clave: enfermedad renal crónica, trayectorias de progresión de la enfermedad renal crónica, análisis armónico cualitativo, análisis factorial múltiple, análisis de correspondencias difuso, clasificación de datos cualitativos longitudinales.

Keywords: chronic kidney disease, progression trajectory of chronic kidney disease , harmonic qualitative analysis, multiple factorial analysis, fuzzy correspondence analysis, qualitative longitudinal data clustering.

Nota de aceptación

Trabajo de tesis

Aprobado

“Mención Meritoria”

Jurado

Luis Guillermo Díaz, Ph.D. Universidad Nacional de Colombia

Jurado

Alvaro Mauricio Montenegro, Ph.D. Universidad Nacional de Colombia

Director

Campo Elías Pardo, Ph.D. Universidad Nacional de Colombia

Codirector

Rafael Mauricio Sanabria, M.Sc. Gerente general de asuntos científicos de Renal Therapy Service

Bogotá, noviembre 25 de 2016

A mis hijas, pero sobre todo, a ti Angie.

Agradecimientos

Inicialmente deseo extender mis agradecimientos al profesor Campo Elías Pardo, director de este trabajo, por ser mentor y maestro y al doctor Rafael Mauricio Sanabria, codirector de mi trabajo y gerente de asuntos científicos de Renal Therapy Services, por su oportuno y coherente acompañamiento y por permitir el acceso a los datos sobre los cuales se desarrolló esta investigación. Agradezco también a cada uno de los miembros del proyecto Estudio de la progresión de la ERC en una población intervenida con un programa de prevención secundaria en Colombia, el doctor Ricardo Sánchez, el profesor Edilberto Cepeda y los estudiantes Jhonathan Medina y Sergio Páez por sus valiosos aportes a la elaboración de este trabajo.

Índice general

Índice general	V
Índice de tablas	IX
Índice de figuras	XIII
Introducción	XVII
1. La enfermedad renal crónica (ERC)	1
1.1. Los estadios de la ERC	2
1.2. Factores de riesgo para la ERC	4
1.3. Prevalencia e incidencia de la ERC	5
1.4. Diagnóstico y tratamiento de la ERC	5
1.5. Trayectorias de la tasa de filtrado glomerular (TFG)	8
2. Métodos estadísticos utilizados	11
2.1. Notación	11
2.2. Análisis armónico cualitativo (AAC)	13
2.2.1. Principios del AAC	13
2.2.2. El AAC como un análisis de correspondencias particular	14
2.2.3. Consideraciones prácticas del AAC	15
2.2.4. Obtención de tipologías de trayectorias	17
2.3. Análisis de correspondencias difuso (ACD)	17
2.4. Análisis factorial múltiple (AFM)	19
2.4.1. Factores extraídos por el AFM	20
2.4.2. Análisis de conjuntos de variables	21
2.4.3. Inclusión de conjuntos de variables cualitativas al AFM	22
2.4.4. Distancias entre individuos y fórmula de transición general	22

2.4.5. El AFM para una variable longitudinal cualitativa	23
2.5. El AFM y el análisis multicanónico	25
3. Combinación del AAC y el AFM (AAC-AFM)	27
3.1. Propiedades del AAC-AFM	28
3.1.1. Distancia entre individuos	28
3.1.2. Distancia entre categorías	29
3.1.3. Relaciones de transición	29
3.1.4. Inercias de la combinación	30
3.1.5. Relación entre periodos de tiempo	31
4. Aplicación del AAC-AFM	33
4.1. Descripción de trayectorias de progresión de la ERC para una población de pacientes en un programa de prevención secundaria	33
4.2. Los datos de progresión a través de la ERC	34
4.2.1. La variable longitudinal cualitativa (VLC) activa	34
4.2.2. Variables transversales y longitudinales ilustrativas de la ERC	35
4.3. Etapa de preprocesamiento de los datos	36
4.3.1. Ajustes de información y reestructuración de la tabla VLC	36
4.3.2. La tabla P: proporción de tiempo de permanencia en un estadio de la ERC	40
4.4. Análisis descriptivo de perfiles de permanencia global en estadios de la ERC	43
4.5. Análisis factoriales aplicados	46
4.5.1. Aplicación del análisis armónico cualitativo	47
4.5.1.1. Clases de progresión de la ERC a partir del AAC	49
4.5.2. Aplicación del análisis factorial múltiple sobre el AAC	51
4.5.2.1. Análisis de los periodos de corte del AAC-AFM y grupos de VTI	56
4.6. Clases de progresión de la ERC a partir del AAC-AFM	60
4.6.1. Trayectorias promedio de las clases de progresión de la ERC	63
4.6.1.1. Caracterización de las trayectorias promedio de las clases a partir de la VLC activa	65
4.6.2. Los perfiles de progresión de las clases de la ERC	67
4.6.2.1. Perfil de progresión de la ERC: clase 1	68
4.6.2.2. Perfil de progresión de la ERC: clase 2	69
4.6.2.3. Perfil de progresión de la ERC: clase 3	70
4.6.2.4. Perfil de progresión de la ERC: clase 4	71

4.6.2.5.	Perfil de progresión de la ERC: clase 5	73
4.6.2.6.	Perfil de progresión de la ERC: clase 6	74
4.6.2.7.	Perfil de progresión de la ERC: clase 7	75
4.6.3.	Comparación de las clases de la ERC mediante las VTI	77
4.6.3.1.	Distribución de las TFG inicial, promedio y final en las clases de la ERC	77
4.6.3.2.	Distribución de la edad y los días aportados al estudio en las clases de la ERC	80
4.6.3.3.	Distribución de las categorías de las VTI cualitativas en las clases de la ERC	81
4.6.4.	Clases de progresión de la ERC a través de trayectorias de las VLI . .	86
4.6.4.1.	Trayectorias medias de las VLI: clase 1	88
4.6.4.2.	Trayectorias medias de las VLI: clase 2	89
4.6.4.3.	Trayectorias medias de las VLI: clase 3	90
4.6.4.4.	Trayectorias medias de las VLI: clase 4	91
4.6.4.5.	Trayectorias medias de las VLI: clase 5	92
4.6.4.6.	Trayectorias medias de las VLI: clase 6	93
4.6.4.7.	Trayectorias medias de las VLI: clase 7	94
4.6.5.	Comparación de las clases de la ERC mediante las VLI	95
4.6.5.1.	Trayectorias de la TFG para las clases de progresión de la ERC	95
4.6.5.2.	Trayectorias de la creatinina para las clases de progresión de la ERC	96
4.6.5.3.	Trayectorias del bun para las clases de progresión de la ERC	96
4.6.5.4.	Trayectorias de la hemoglobina para las clases de progresión de la ERC	97
4.6.5.5.	Trayectorias del índice de masa corporal para las clases de progresión de la ERC	98
4.6.5.6.	Trayectorias de la tensión sistólica para las clases de pro- gresión de la ERC	99
4.6.5.7.	Trayectorias de la tensión diastólica para las clases de pro- gresión de la ERC	100
5.	Conclusiones	103
A.	Aspectos éticos y de confidencialidad del Estudio	113
B.	Variables transversales y longitudinales ilustrativas de la ERC	115
B.1.	Análisis descriptivo de las VTI de la población con ERC	117

B.2. Variables categóricas transversales ilustrativas	117
B.3. Variables cuantitativas transversales ilustrativas	119
B.4. VTI cuantitativas por categorías de las VTI categóricas	121
C. Permanencias y transiciones entre estadios de la ERC y censuras	127
D. Primeros planos factoriales del AAC y del AAC-AFM	129
E. Caracterización de las clases de progresión a partir de la VLC y las VTI	131
F. Permanencias y transiciones de las clases en estadios y censuras de la ERC	139
G. Síntesis descriptiva de las clases de progresión	143

Índice de tablas

1.1. Clasificación de los estadios de ERC según guías <i>K/DIGO</i>	3
1.2. Factores de riesgo para la ERC	4
4.1. VLC activa: cambios de estadio de la ERC	35
4.2. VLC activa: cambios de estadio de la ERC. Primer ajuste	37
4.3. VLC activa: duración y transición entre estadios de la ERC. Segundo ajuste	39
4.4. Longitud de cortes del periodo de estudio y proporción de cambios de estadio retenidos	42
4.5. Tabla A: tiempos de permanencia en estadios de la ERC y censuras por periodos de corte	43
4.6. Tabla P: proporción de tiempos de permanencia en estadios de la ERC y censuras por periodos de corte	43
4.7. Proporción de inercia proyectada sobre los primeros ejes del AAC	47
4.8. Primeros valores propios y pesos de los AAC separados	53
4.9. Proporción de inercia proyectada sobre los primeros ejes del AAC-AFM	53
4.10. Correlación de los periodos de corte con los primeros ejes del AAC-AFM	56
4.11. Ayudas para la interpretación de los periodos de corte y de los grupos de variables ilustrativas para el AAC-AFM	58
4.12. Coeficientes <i>RV</i> de los periodos de corte y de los grupos de variables ilustrativas del AAC-AFM	59
4.13. Coeficientes <i>Lg</i> de los periodos de corte y de los grupos de variables ilustrativas del AAC-AFM	59
4.14. Clases de progresiones de la ERC a partir del AAC-AFM	61
B.1. Tabla de variables transversales ilustrativas del AAC-AFM	115
B.2. Tabla de variables longitudinales ilustrativas del AAC-AFM	115
B.3. Definición operativa de las variables ilustrativas del AAC-AFM	116
B.4. Síntesis de la distribución de las categorías de las VTI categóricas	118

B.5. Medidas descriptivas de las VTI cuantitativas	119
C.1. Frecuencias de permanencia en estadios de la ERC y censuras	127
C.2. Transiciones entre estadios y censuras por periodos de corte	128
E.1. Caracterización de la clase 1 según valores test de la VLC	131
E.2. Caracterización de la clase 1 según valores test de las VTI categóricas	131
E.3. Caracterización de la clase 1 según valores test de las VTI cuantitativas	131
E.4. Caracterización de la clase 2 según valores test de la VLC	132
E.5. Caracterización de la clase 2 según valores test de las VTI categóricas	132
E.6. Caracterización de la clase 2 según valores test de las VTI cuantitativas	132
E.7. Caracterización de la clase 3 según valores test de la VLC	132
E.8. Caracterización de la clase 3 según valores test de las VTI categóricas	132
E.9. Caracterización de la clase 3 según valores test de las VTI cuantitativas	132
E.10. Caracterización de la clase 4 según valores test de la VLC	133
E.11. Caracterización de la clase 4 según valores test de las VTI categóricas	133
E.12. Caracterización de la clase 4 según valores test de las VTI cuantitativas	133
E.13. Caracterización de la clase 5 según valores test de la VLC	134
E.14. Caracterización de la clase 5 según valores test de las VTI categóricas	134
E.15. Caracterización de la clase 5 según valores test de las VTI cuantitativas	134
E.16. Caracterización de la clase 6 según valores test de la VLC	135
E.17. Caracterización de la clase 6 según valores test de las VTI categóricas	135
E.18. Caracterización de la clase 6 según valores test de las VTI cuantitativas	135
E.19. Caracterización de la clase 7 según valores test de la VLC	135
E.20. Caracterización de la clase 7 según valores test de las VTI categóricas	135
E.21. Caracterización de la clase 7 según valores test de las VTI cuantitativas	135
E.22. Síntesis de las clases de progresión según las VTI categóricas	136
E.23. Síntesis de las clases de progresión según las VTI cuantitativas	137
F.1. Permanencia de las clases en estadios y censuras de la ERC	139
F.2. Frecuencias de transiciones entre estadios y censuras por periodo: clase 1	139
F.3. Frecuencias de transiciones entre estadios y censuras por periodo: clase 2	140
F.4. Frecuencias de transiciones entre estadios y censuras por periodo: clase 3	140
F.5. Frecuencias de transiciones entre estadios y censuras por periodo: clase 4	140
F.6. Frecuencias de transiciones entre estadios y censuras por periodo: clase 5	141
F.7. Frecuencias de transiciones entre estadios y censuras por periodo: clase 6	141

F.8. Frecuencias de transiciones entre estadios y censuras por periodo: clase 7 . . .	141
G.1. Síntesis de la caracterización de la clase 1	143
G.2. Síntesis de la caracterización de la clase 2	144
G.3. Síntesis de la caracterización de la clase 3	144
G.4. Síntesis de la caracterización de la clase 4	145
G.5. Síntesis de la caracterización de la clase 5	145
G.6. Síntesis de la caracterización de la clase 6	146
G.7. Síntesis de la caracterización de la clase 7	146

Índice de figuras

1.1. Ejemplos de trayectorias de progresión de la ERC en el tiempo	9
2.1. Tabla global de proporción de tiempo de permanencia en cada categoría de la VLC	12
4.1. Histogramas de frecuencias absolutas y acumuladas de transiciones de la ERC. 2190 días	41
4.2. Perfil global de tiempo de permanencia en estadios de la ERC y censuras por periodos de corte	44
4.3. Transiciones entre estadios de la ERC por periodos de corte	44
4.4. Histograma de valores propios del AAC	47
4.5. Primer plano factorial del AAC con individuos y estadios por periodos proyectados	48
4.6. Histograma de índices de nivel para clasificación en el AAC	50
4.7. Clases de progresión de la ERC en el primer plano factorial del AAC	51
4.8. Valores propios de análisis separados del AFM	52
4.9. Histograma de valores propios del AAC-AFM	53
4.10. Primer plano factorial del AAC-AFM con individuos y centroides de las VTI proyectados	54
4.11. Círculo de correlación de periodos de corte y estadios del AAC-AFM	55
4.12. Proyección de los periodos de corte en el primer plano factorial del AAC-AFM	57
4.13. Histograma de índices de nivel para clasificación en el AAC-AFM	61
4.14. Clases de progresión de la ERC en el primer plano factorial del AAC-AFM	62
4.15. Trayectorias promedio de las primeras cuatro clases de progresión de la ERC a partir del AAC-AFM	64
4.16. Trayectorias promedio de las tres últimas clases de progresión de la ERC a partir del AAC-AFM	65
4.17. Perfil de progresión de la ERC de la clase 1	68
4.18. Transiciones entre estadios y/o censuras por periodos de corte de la clase 1	68

4.19. Perfil de progresión de la ERC de la clase 2	70
4.20. Transiciones entre estadios y/o censuras por periodos de corte de la clase 2	70
4.21. Perfil de progresión de la ERC de la clase 3	71
4.22. Transiciones entre estadios y/o censuras por periodos de corte de la clase 3	71
4.23. Perfil de progresión de la ERC de la clase 4	72
4.24. Transiciones entre estadios y/o censuras por periodos de corte de la clase 4	72
4.25. Perfil de progresión de la ERC de la clase 5	73
4.26. Transiciones entre estadios y/o censuras por periodos de corte de la clase 5	73
4.27. Perfil de progresión de la ERC de la clase 6	75
4.28. Transiciones entre estadios y/o censuras por periodos de corte de la clase 6	75
4.29. Perfil de progresión de la ERC de la clase 7	76
4.30. Transiciones entre estadios y/o censuras por periodos de corte de la clase 7	76
4.31. Distribución de las TFG iniciales para las clases de progresión	77
4.32. Distribución de las TFG promedios para las clases de progresión	78
4.33. Distribución de las TFG finales para las clases de progresión	79
4.34. Distribución de las edades para las siete clases de progresión de la ERC	80
4.35. Distribución de las edades para las siete clases de progresión de la ERC	81
4.36. Perfil de las clases de progresión según género	84
4.37. Perfil de las clases de progresión según nivel de escolaridad	84
4.38. Perfil de las clases de progresión según diagnóstico de cáncer	84
4.39. Perfil de las clases de progresión según antecedente de enfermedad cardiovascular	84
4.40. Perfil de las clases de progresión según causa de egreso del estudio	85
4.41. Perfil de las clases de progresión según causa de la ERC	85
4.42. Perfil de las clases de progresión según diagnóstico de diabetes	85
4.43. Perfil de las clases de progresión según diagnóstico de HTA	85
4.44. Trayectorias medias crudas de las VLI de la clase 1	86
4.45. Trayectorias de las VLI de la clase 1	88
4.46. Trayectorias de las VLI de la clase 2	89
4.47. Trayectorias de las VLI de la clase 3	90
4.48. Trayectorias de las VLI de la clase 4	91
4.49. Trayectorias de las VLI de la clase 5	92
4.50. Trayectorias de las VLI de la clase 6	93
4.51. Trayectorias de las VLI de la clase 7	94
4.52. Trayectorias de la TFG estimada para las clases de progresión de la ERC	95

4.53. Trayectorias de la creatinina estimada para las clases de progresión de la ERC	96
4.54. Trayectorias del BUN para las clases de progresión de la ERC	97
4.55. Trayectorias de la hemoglobina para las clases de progresión de la ERC . . .	98
4.56. Trayectorias del imc para las clases de progresión de la ERC	99
4.57. Trayectorias de la tensión arterial sistólica para las clases de progresión de la ERC	100
4.58. Trayectorias de la tensión arterial diastólica para las clases de progresión de la ERC	101
B.1. Histogramas de frecuencias de las VTI cuantitativas	120
B.2. Variable edad por categorías de las VTI categóricas (I)	121
B.3. Variable edad por categorías de las VTI categóricas (II)	122
B.4. TFG inicial, media y final por género, HTA y diabetes	123
B.5. TFG inicial, media y final por causa de egreso del estudio	125
B.6. TFG inicial, media y final por causas de la ERC	126
D.1. Primeros planos factoriales del AAC	129
D.2. Primeros planos factoriales del AAC-AFM	130

Introducción

La enfermedad renal crónica (ERC) es una afección de los riñones y se define como una disminución de la función renal. Esta disminución se expresa a través de una tasa de filtrado glomerular (TFG) o un aclaramiento de la creatinina estimadas menor a $60 \text{ ml/min/1.73m}^2$, o como la presencia de daño renal, para ambos casos, de forma permanente durante al menos tres meses. En la ERC la función renal declina progresivamente en un periodo largo de tiempo, y generalmente, lleva al individuo que la padece a una fase terminal, denominada insuficiencia renal crónica terminal (IRCT), donde se presenta fallo renal. Al ocurrir este evento se hace necesario remitir al individuo a un tratamiento sustitutivo de dicha función (terapias de reemplazo renal, TRR) o un trasplante renal, que de no realizarse a tiempo, puede llevar a un desenlace fatal (Cabrera 2004).

Las TRR se presentan indeseables en el sentido de que afectan negativamente la calidad de vida de los individuos, no garantizan la supervivencia de los mismos y resultan excesivamente costosas. Retrasar en la mayor medida el tiempo de llegada a esta fase de la ERC se constituye en uno de los principales propósitos de los programas de seguimiento y prevención de esta enfermedad. Sin embargo, esta tarea resulta compleja por cuanto es necesario tener un conocimiento amplio del modo como evoluciona la población intervenida a través de su ERC y de los factores clínicos y socioeconómicos que pueden acelerar o desacelerar la progresión a la IRCT. Conocida esta información, especialistas y en general el personal médico, pueden entonces generar estrategias para prevenir complicaciones, retrasar la llegada a IRCT mediante diagnóstico precoz y realizar un manejo terapéutico adecuado. De aquí la importancia de desarrollar estudios sobre la progresión de la ERC (Levey et al. 2002).

Con base en valores de la TFG, en Levey et al. (2002) y posteriormente en Levey & Coresh (2012) la *National Kidney Foundation* de Estados Unidos a través del proyecto *K/DIGO* (*Kidney Disease Improving Global Outcomes*) propone una clasificación en estadios de la ERC. Estos estadios son ordenados decrecientemente, de menor a mayor gravedad: el estadio 1 corresponde a una TFG estimada superior o igual a 90 (pero con indicios de daño renal), mientras que el estadio 5 corresponde a una TFG estimada menor a 15 y significa que el individuo ha llegado a IRCT y se hace necesaria la remisión a TRR o trasplante.

El objetivo principal de este trabajo es determinar, describir y comparar las trayectorias que describen un conjunto de individuos adscritos a un programa de prevención y seguimiento a través de los distintos estadios de la ERC. De manera general, los estadios de la ERC pueden verse como el conjunto de posibles estados que pueden asumir los individuos durante un intervalo de tiempo determinado. Si se considera la progresión de los individuos a través de estos estadios como una secuencia de estados, unido al conjunto de instantes en el tiempo donde se realiza la transición de un estado a otro, entonces lo que se obtiene es la realización de una variable longitudinal cualitativa (VLC). Esta VLC no es más que la trayectoria de un individuo a través de su ERC en un tiempo continuo acotado. El estudio de estas trayectorias individuales puede resultar de suma importancia clínica, pues permite reconstruir la historia natural de la ERC de la población bajo estudio; identificar los factores de progresión (clínicos o sociodemográficos) que aceleran o retrasan la evolución de estadios determinados a estadios de mayor grado y finalmente, detectar grupos de trayectorias individuales similares (progresivos lentos, normales o rápidos, por ejemplo).

La naturaleza longitudinal y cualitativa de las trayectorias exige el uso de técnicas exploratorias no convencionales del análisis multivariado de datos, como por ejemplo, el *análisis armónico cualitativo* (AAC) (Daville & Saporta 1980, Deville 1982, Deville & Saporta 1983). El AAC permite obtener una tipología multivariada de trayectorias individuales en el tiempo, y es básicamente, equivalente a un *análisis de correspondencias* (AC) realizado sobre una tabla de individuos \times categorías \times periodos de tiempo (Barbary 1996). El análisis se realiza sobre la VLC codificada convenientemente, codificación que representa simultáneamente los tiempos de permanencia de los individuos en cada estadio y el paso por cada uno de ellos. Este tiempo/transición por los diferentes estadios genera una estructura

que permite identificar aquellos individuos con trayectorias de cambios similares (Saporta 1996), que unido a una caracterización a partir de variables ilustrativas (variables que no tienen peso sobre el análisis), producen unos perfiles, denominados *perfiles de progresión de la ERC*.

Particularmente, la VLC se codifica dividiendo el tiempo de estudio en un conjunto de intervalos (periodos de corte), con igual o diferente duración, y calculando la proporción de tiempo que pasa cada individuo en cada uno de los estadios por cada periodo de corte. Al realizar esto, se obtiene una tabla particular que describe las trayectorias de los individuos, tabla que luego es sometida a un AC particular denominado *análisis de correspondencias difuso* (ACD), propuesto por Chevenet et al. (1994). Este AC, en el contexto del *análisis factorial general*, constituye un *análisis en componentes principales ponderado* (Escofier & Pagès 1992).

En Corrales & Pardo (2015) se sugiere combinar el AAC con *el análisis factorial múltiple* (AFM), método desarrollado por Escofier & Pagès (1984) que permite analizar tablas donde individuos están descritos por conjuntos de variables. Dentro de los principales propósitos de aplicar el AFM sobre el AAC están: obtener una síntesis parcial (por periodos de tiempo) y una síntesis global de toda la información; equilibrar la influencia de los distintos periodos de tiempo dentro del AAC y finalmente, acceder a medidas de asociación para cuantificar la semejanza global entre tales periodos.

A partir de las coordenadas factoriales del AAC-AFM (combinación del AAC-AFM), se logran obtener las distintas clases de progresión de la ERC. La metodología para realizar esto supone aplicar un *análisis de clasificación jerárquica con el criterio agregación de mínima varianza de Ward* (Lebart et al. 2006) sobre las coordenadas factoriales del AAC-AFM. Esta clasificación sobre los primeros factores del AAC-AFM exige determinar el número óptimo de ejes (para clasificación) a retener. Un criterio de selección es seleccionar aquel número de ejes que permitan obtener una partición cuya clase mayoritaria sea la de menor tamaño entre todas las clases mayoritarias de las demás particiones (Barbary & Pinzón 1998). Obtenidas las distintas clases de progresión de la ERC, a partir de la partición seleccionada, éstas se consolidan a través de un *K-means* (Lebart et al. 2006) y se caracterizan a través de la misma VLC y de distintas variables transversales y longitudinales, variables denominadas *variables transversales ilustrativas* (VTI) y *variables longitudinales ilustrativas* (VLI) respectivamente .

Se logra entonces, a partir del AAC-AFM y de la clasificación automática, determinar y caracterizar grupos de trayectorias de progresión de la ERC para un conjunto de 2445 individuos, cuyo seguimiento se realizó por cinco años para la mayoría de ellos. Además, también se logra identificar qué VTI o VLI se constituyen en factores de progresión de la ERC (de retardo o de aceleración de un estadio a otro).

Este documento se divide en cinco capítulos organizados de la siguiente manera:

En el capítulo 1 se realiza una breve descripción de los aspectos fundamentales de la enfermedad renal crónica con relación a su definición, sus estadios, sus factores de riesgo, su prevalencia e incidencia y el estudio de sus trayectorias. El capítulo 2 presenta los aspectos teóricos y metodológicos más relevantes del AAC y el AFM. Por su parte, el capítulo 3 está dedicado a describir la manera como se combinan el AAC y el AFM, mostrando cómo las fórmulas clásicas del análisis factorial general se adaptan a dicho método. En el capítulo 4, se presentan y discuten los principales resultados obtenidos tras aplicar el AAC-AFM sobre los datos de progresión de la ERC. Puntualmente en este capítulo, se describe el modo como las distintas clases de la progresión a través de la ERC se determinan, se reportan sus características principales y finalmente, se comparan entre ellas. Por último, en el capítulo 5 se sintetizan las conclusiones encontradas a partir de los resultados obtenidos de la aplicación del AAC-AFM sobre los datos de la ERC y se proponen algunos desarrollos futuros y oportunidades a partir del trabajo presentado.

CAPÍTULO 1

La enfermedad renal crónica (ERC)

La función renal es el proceso fisiológico a través del cual los riñones eliminan del cuerpo: toxinas, desechos metabólicos y el exceso de agua y iones, al tiempo que, reabsorben las sustancias que no deben ser eliminadas. Este proceso se logra filtrando (a través de una permeabilidad selectiva) la sangre que llega a una red de capilares ubicados en los riñones (denominados capilares glomerulares). De este modo, elementos como el sodio, el potasio y la creatinina, entre otros, son excretados del cuerpo, mientras que otros elementos necesarios para el organismo, como la hemoglobina y la albúmina, son retenidos (Mezzano & Aros 2005).

La disminución de la función renal se expresa por una declinación progresiva de la *tasa de filtrado glomerular* (TFG). Esta tasa expresa el volumen de fluido filtrado por unidad de tiempo desde los capilares glomerulares hasta la *cápsula de Bowman* (unidad renal donde se realiza el filtrado). Generalmente, la TFG se mide en mililitros por minuto por metro cuadrado ($ml/min/m^2$) y se puede estimar a través de distintas ecuaciones que incorporan variables demográficas y clínicas (Haynes & Winearls 2010). Dentro de las ecuaciones de estimación de la TFG más utilizadas está la *fórmula de Cockcroft-Gault* y la *ecuación abreviada del estudio MDRD*.

La fórmula de Cockcroft-Gault (Himmelfarb & Sayegh 2010) permite estimar el aclaramiento de la creatinina, que a su vez es un marcador endógeno utilizado para la estimación del la TFG. La fórmula se expresa de la siguiente manera:

$$\widehat{C}_{cr} = \frac{(140 - E \times P) \times 0.85 \text{ (si es mujer)}}{Scr \times 72} \quad (1.1)$$

donde \widehat{C}_{cr} es el aclaramiento de creatinina estimado en ml/min , E es la edad en años, P es el peso en Kg y Scr es el nivel de creatinina en la sangre en mg/dl .

Por otro lado, la ecuación abreviada MDRD (*Modification of Diet in Renal Disease*) (Levey et al. 2002), también involucra el nivel de creatinina en la sangre y la edad, pero a la vez está condicionada por el género y la raza:

$$\widehat{TFG} = 186 \times Scr^{-1.154} \times E^{-0.203} \times 1,21 \text{ (Si es raza negra)} \times 0,742 \text{ (Si es mujer)} \quad (1.2)$$

En 2009, *Chronic Kidney Disease-Epidemiology Collaboration* (CKD-EPI) propone una fórmula basada en una población con valores de TFG más elevados y métodos de creatinina estandarizados, conocida como CKD-EPI (Eknoyan et al. 2013), sustituyendo la MDRD, pues presenta una mejor exactitud al permitir una cuantificación de la TFG por encima de $60 ml/min/m^2$.

1.1. Los estadios de la ERC

La *enfermedad renal crónica* (ERC) se define como una disminución progresiva de la función renal o como la presencia de daño renal permanente durante al menos tres meses (Levey & Coresh 2012). Específicamente, cuando un individuo presenta una TFG estimada (e.g. un aclaramiento estimado de la creatinina) menor a $60 ml/min/1.73m^2$ o tiene indicios de daño renal durante el tiempo observado, el individuo se diagnostica con ERC. El daño renal se diagnostica habitualmente mediante marcadores dentro de los que se destaca una excreción urinaria de albúmina o proteínas elevada (albuminuria y proteinuria respectivamente) (Cabrera 2004).

Una vez diagnosticada la ERC, se presenta una declinación permanente de la TFG en función del tiempo. Con base en valores del TFG, la *National Kidney Foundation* de Estados Unidos a través del proyecto *K/DIGO* (*Kidney Disease Improving Global Outcomes*)

propone una clasificación de los estadios de la ERC (Levey & Coresh 2012), clasificación que se presenta en la tabla 1.1:

Estadio	Descripción	TFG ($ml/min/1.73m^2$)
1	Daño renal con TFG normal a alta	≥ 90
2	Daño renal con TFG ligeramente disminuida	60 - 89
3.a	TFG ligera a moderadamente disminuida	45 - 59
3.b	TFG moderada a gravemente disminuida	30 - 44
4	TFG severamente disminuida	15 - 29
5	Fallo renal	< 15

TABLA 1.1. Clasificación de los estadios de ERC según guías *K/DIGO* de la *National Kidney Foundation*, (Levey & Coresh 2012).

Algunos de los más importantes propósitos de establecer estas etapas evolutivas de la ERC son: a) clasificar la ERC independientemente de la(s) causa(s) subyacentes; b) asociar los estadios con las posibles complicaciones de la ERC; c) estratificar los riesgos de progresión y, d) estratificar el riesgo de desarrollar complicaciones cardiovasculares (Levey & Coresh 2012).

Los casos en los estadios 1 y 2 (evidencia de daño renal por más de tres meses y TFG ligeramente disminuida respectivamente) podrían beneficiarse del diagnóstico precoz y del inicio de medidas preventivas para frenar la progresión de la ERC y del manejo adecuado de enfermedades coadyuvantes como la tensión arterial alta (HTA), la diabetes y las patologías cardiovasculares, pues existen evidencias de que estas afecciones constituyen factores desencadenantes y comórbidos de la ERC. En los estadios 3a y 3b (no se necesita evidencia de daño renal) se presenta un riesgo aumentado de progresión de la ERC a un estadio mayor o a un estadio terminal y de complicaciones cardiovasculares; en estos estadios los pacientes deben ser evaluados integralmente desde el punto de vista cardiovascular y renal, con el ánimo de recibir un tratamiento adecuado en ambos niveles y tratar de desacelerar la progresión de la ERC. Pacientes que alcanzan el estadio 4 (disminución severa la TFG) presentan una probabilidad muy alta de pasar al estadio 5 y de sufrir complicaciones cardiovasculares graves; se debe preparar al paciente para una posible *terapia de reemplazo renal* (TRR). La manifestación más avanzada para un paciente afectado por la ERC es la *insuficiencia renal crónica terminal* (IRCT). En esta situación, que corresponde al estadio 5, se presenta fallo renal por lo cual se establece la necesidad de implementar un tratamiento sustitutivo de la función renal (e.g. hemodiálisis o diálisis peritoneal) o un trasplante renal, que de no realizarse puede llevar a un desenlace fatal, pues estos constituyen trata-

mientos de soporte de vida para quienes padecen la ERC en este estadio (Cabrera 2004, Mezzano & Aros 2005).

1.2. Factores de riesgo para la ERC

Existen unas situaciones y unos factores de riesgo para el aumento de la progresión de la ERC (Norris et al. 2006). Los factores de riesgo deben tomarse en cuenta al momento de prevenir la ERC, o en su defecto, para realizar un diagnóstico precoz de la misma. Algunos factores de riesgo aumentado de la ERC son: edad mayor o igual a sesenta años, HTA, diabetes, obesidad, enfermedad cardiovascular, TFG levemente disminuida, bajo peso al nacer, entre otras. Los factores de riesgo se pueden dividir en factores de susceptibilidad, factores iniciadores y factores de progresión. Los primeros, los factores de susceptibilidad, son condiciones que aumentan el riesgo de desarrollar ERC; los factores iniciadores están implicados en el inicio del daño renal y los factores de progresión, determinan la progresión de la ERC, esto es, condicionan el tiempo de permanencia de un individuo en un determinado estadio y la consecuente progresión a un estadio de mayor grado. En la tabla 1.2, tomada y modificada de Cabrera (2004), se presentan los factores de riesgo de la ERC más conocidos.

Factores de susceptibilidad	<ul style="list-style-type: none"> - Edad mayor a 60 años - Historia familiar ERC - Masa renal disminuida - Bajo peso al nacer - Raza afroamericana - Diabetes - Hipertensión arterial
Factores iniciadores	<ul style="list-style-type: none"> - Enfermedades autoinmunes - Infecciones urinarias - Fármacos nefrotóxicos - Diabetes - Hipertensión arterial
Factores de progresión	<ul style="list-style-type: none"> - Protenuira persistente - HTA mal controlada - Diabetes mal controlada - Tabaco - Anemia - Enfermedad cardiovascular - Dislipidemia - Diferir envío a nefrólogo

TABLA 1.2. Factores de riesgo para la ERC

1.3. Prevalencia e incidencia de la ERC

La ERC es una enfermedad catalogada como catastrófica y de alto costo, que ha alcanzado en años recientes proporciones de epidemia. Según la Organización Mundial de la Salud afectaba, para el 2015, a cerca del 10 % de la población mundial. En países como México, Estados Unidos, Taiwan y Japón, el número de pacientes con IRCT se ha triplicado en las últimas décadas, con prevalencias mayores a los 1800 casos por millón de habitantes e incidencias medias aproximadas de 400 por millón de habitantes por año. En Europa, cuya incidencia de IRCT es menor que en esos países, se ha observando un aumento del 5% anual de pacientes en TRR, mientras que a nivel latinoamericano, para el 2011, se estimaba una incidencia de 613 por millón anual. Para Colombia, Cuenta de Alto Costo (2015) en dicho año reportó 34.469 personas en TRR, lo que condujo a una prevalencia de 715 por millón de habitantes (892 para hombres y 566 para las mujeres) y una incidencia de 123 por millón anual.

Sin embargo, los estudios epidemiológicos de la ERC para los estadios anteriores al fallo renal son escasos y poco documentados (Cabrera 2004). Para 2010, Mills et al. (2015) estimaron que a nivel mundial la prevalencia de la ERC (estadios 1 al 5 inclusive) para adultos de 20 años o más, era del 10.4% para hombres y 11.8% para las mujeres, y que el número total de adultos diagnosticados con ERC en estadios de 1 a 5 se estimaba en aproximadamente 497 millones (esto representaba para ese año el 8.3% de la población mundial). A nivel Nacional, para el 2015, el número de pacientes con ERC en alguno de los estadios de progresión era de 979.409, la prevalencia del 2%, siendo más frecuente en las mujeres que en los hombres y una incidencia de 5.3 por cada 1000 habitantes (Cuenta de Alto Costo 2015).

1.4. Diagnóstico y tratamiento de la ERC

La ERC es una enfermedad de relativamente fácil reconocimiento. Con los tratamientos adecuados, se puede prevenir su desarrollo y reducir la velocidad de progresión a estadios terminales. Así mismo, es posible detectar y tratar sus complicaciones comórbidas (enfermedad cardiovascular, anemia, deterioro físico y mental, etc.). Con base en un modelo

conceptual de la ERC (Levey et al. 2009) se plantean una serie de recomendaciones de intervención que permiten mejorar su pronóstico y evolución, que se pueden resumir en:

- Detección precoz y clasificación del daño renal.
- Diagnóstico, tratamiento, criterios de interconsulta, derivación y seguimiento de la ERC.
- Control de la progresión y de las complicaciones.

Generalmente, la ERC es asintomática en los primeros estadios y se detecta cuando el paciente está siendo valorado por otra condición comórbida. La detección precoz y determinación aproximada del daño renal debe realizarse sobre pacientes con aumento del riesgo de la ERC (factores de susceptibilidad, iniciación y progresión), a través de la estimación de la TFG y la albuminuria. Una detección temprana puede resultar importante a la hora de optimizar las posibilidades de tratamiento, retrasar la progresión, disminuir la morbimortalidad y reducir los costos de tratamiento de la ERC (Himmelfarb & Sayegh 2010).

El diagnóstico de la ERC no debe basarse en una determinación única de la TFG o la albúmina y debe confirmarse su persistencia por al menos tres meses. Este diagnóstico consta de una clasificación lo más precisa del estadio de la ERC (se recomienda para esto la clasificación de la ERC según las guías *K/DIGO* en Levey & Coresh (2012, pág. 168)) y de la identificación de las situaciones comórbidas asociadas. Algunas de estas condiciones asociadas a la ERC, identificadas y descritas en distintas fuentes, son:

La anemia: en esta condición el individuo presenta escasez de glóbulos rojos (hematíes) en el torrente sanguíneo. Los hematíes contienen la *hemoglobina*, proteína encargada de captar y transportar el oxígeno y el dióxido de carbono en el cuerpo. La disminución de estas células puede ser consecuencia de diversas razones, entre ellas el inadecuado funcionamiento de los riñones, los cuales bajo esta condición no producen suficiente *eritropoietina* (EPO), hormona que estimula la producción de glóbulos rojos en la médula ósea. Guías, como la *K/DIGO*, sugieren realizar conteos de hemoglobina a todos los individuos con ERC, independientemente de su causa o del estadio donde estén. Se considera que hay anemia, y por ende un posible mal funcionamiento de los riñones, cuando la concentración de

hemoglobina es menor a 13.5 *g/dl* en hombres o menor a 11.5 *g/dl* en mujeres (Levey & Coresh 2012).

Por otro lado, está la tensión arterial alta. La tensión arterial mide la fuerza que ejerce la sangre contra las paredes de los vasos sanguíneos. Si existe exceso de líquidos (como por ejemplo, por el mal funcionamiento de los riñones) en el torrente sanguíneo, la tensión arterial puede aumentar, forzando el trabajo del corazón y dañando los vasos sanguíneos, incluidos los ubicados en los riñones. Al dañarse los vasos sanguíneos de los riñones por la alta tensión, estos dejan de eliminar el exceso de líquidos en el cuerpo, aumentando nuevamente la presión arterial, generando un ciclo peligroso. Es bien conocido que individuos en estadio *E3* o mayores de la *ERC* presentan tensiones arteriales por encima de 140/90 *mmHg* (Himmelfarb & Sayegh 2010), es decir tensiones sistólicas (presión arterial cuando el corazón está latiendo) mayores a 140 y tensiones diastólicas (presión arterial cuando el corazón está en reposo) mayores a 90.

La obesidad y el sobrepeso son considerados factores de riesgo para desarrollar la *ERC* en poblaciones saludables, mientras que la obesidad, durante los primeros estadios de esta enfermedad, es un factor de riesgo cardiovascular (Himmelfarb & Sayegh 2010). Si bien aún son pocos los estudios que han podido explicar satisfactoriamente la relación de esta condición con la progresión de la *ERC* en sus estadios terminales, en Hsu et al. (2006) los autores logran mostrar, a partir de un grupo de 1471 individuos, que un índice de masa corporal alto (*imc*) es un factor de riesgo para desarrollar fallo renal. De manera particular concluyen que, en comparación con personas con un peso normal (con *imc* entre 18.5 y 24.9 *Kg/m²*), el riesgo relativo ajustado de alcanzar la insuficiencia renal va en aumento a medida que el *imc* también aumenta. Por ejemplo, para personas con sobrepeso (*imc* entre 25 y 29.9 *Kg/m²*) el riesgo es de 1.87, mientras que para personas con obesidad extrema (*imc* mayor a 40 *Kg/m²*) el riesgo reportado es de 7.07.

Para el tratamiento de la *ERC* se recomiendan las siguientes acciones (Subdirección General de Calidad y Cohesión et al. 2015): a) tratamiento de las enfermedades renales subsidiarias; b) control de los factores de riesgo y comorbilidades; c) prevención de la toxicidad sobre los riñones y profilaxis higiénica, dietética y vacunal y, d) estimación de la progresión de la *ERC* y control de los factores de progresión y sus complicaciones.

En cuanto al control de la progresión y de las complicaciones de la *ERC*, debe tenerse en cuenta que se considera progresión cuando se avanza de un estadio determinado a un estadio

de mayor gravedad, o cuando existe un descenso de la TFG mayor a $5\text{ml}/\text{min}/1.73\text{m}^2$ en un año o a $10\text{ml}/\text{min}/1.73\text{m}^2$ en cinco años. También se considera progresión, cuando sobre una medida basal o previa de la TFG esta disminuye en un 25 %. Nuevamente, la idea principal es desarrollar estrategias que desaceleren la velocidad, traten complicaciones, detecten signos de alarma, y de ser necesario, preparen al paciente para una TRR o un trasplante.

1.5. Trayectorias de la tasa de filtrado glomerular (TFG)

La hipótesis generalmente aceptada entre la comunidad dedicada al estudio de la ERC es que, en largos periodos de tiempo, la progresión longitudinal de la TFG decrece de manera aproximadamente lineal; variados modelos matemáticos de la trayectoria en el tiempo de esta tasa descansan sobre este supuesto (Mitch et al. 1976, Levey et al. 1988, Hunsicker et al. 1997). En este contexto, la pendiente es una adecuada medida de la tasa de progresión, y el modelo a través del cual se describe tiene variadas ventajas: estadísticamente fiable y fácil de interpretar, un software bastante desarrollado, bases para la investigación clínica, pero sobre todo, ha permitido a los especialistas predecir de manera más o menos acertada el momento en el cuál el paciente podría alcanzar la IRCT.

Sin embargo, el supuesto de linealidad en la declinación progresiva de la TFG deja de ser válido cuando diversas realidades son puestas en juego, sobre todo si el periodo de observación de la ERC es largo (generalmente en años) Haynes & Winearls (2010). Algunos ejemplos de progresiones de esta enfermedad en el tiempo son presentadas en la figura 1.1:

La línea 1, de color verde, muestra un decrecimiento lineal típico de la TFG en un periodo mayor a 10 años. La línea 2 (azul) muestra una progresión escalonada, mientras que la línea 3 (roja) es una progresión estable de la función renal. Finalmente, la línea punteada 4, de color verde, describe una declinación drástica de la TFG, propia de un fallo renal. Se muestra entonces una gran variedad en las tasas de progresión. Por ejemplo, progresiones rápidas de la ERC (línea 4) son consecuencia de predictores independientes como la proteinuria, la hipertensión y la raza negra, mientras que progresiones estables son observadas en periodos cortos de tiempo y corresponden particularmente a pacientes de descendencia europea (Hu et al. 2012). Por último en algunos individuos, estables por algún tiempo en su función lineal, se pueden presentar saltos (línea 2) que coinciden con

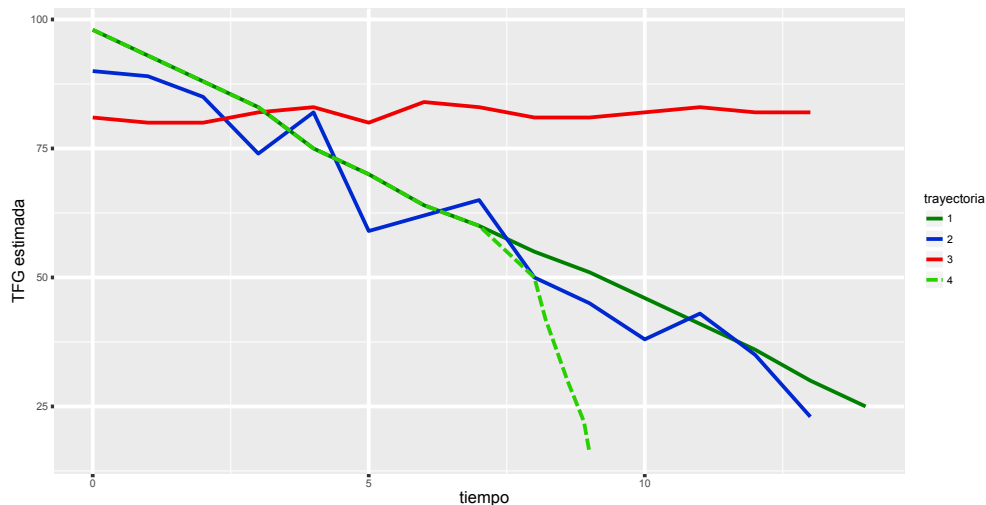


FIGURA 1.1. Ejemplos de trayectorias de progresión de la ERC en el tiempo.

cuadros de sepsis (respuesta sistémica a una infección), hipotensión o como reacción a un tratamiento con agentes nefrotóxicos.

Teniendo en cuenta las situaciones anteriores, es razonable pensar que la progresión longitudinal media de la ERC no describe una trayectoria lineal, al menos en un periodo corto de observación. Alejarse de la linealidad, implica que los modelos planteados bajo este supuesto podrían conllevar a predicciones del tiempo de ocurrencia de la IRCT poco fiables estadísticamente, con muchas consecuencias clínicas indeseables. Recientes estudios han puesto en tela de juicio este supuesto de linealidad y han comenzado a plantear modelos matemáticos no lineales (e incluso no progresivos) para ajustar la tasa de filtrado a situaciones más reales. En este sentido, se destaca el trabajo de Li et al. (2012), quienes logran mostrar que la probabilidad de que las trayectorias individuales de una proporción grande de individuos afroamericanos (352 de 846) se alejen de la linealidad, o no tengan un periodo de progresión, es mayor al 90%. Por otro lado, Hu et al. (2012), a través de modelos Bayesianos de efectos mixtos, observan la presencia de pendientes positivas de las TFG estimadas para pacientes con ERC hipertensiva durante las fases de ensayo, pendientes que no pudieron ser explicadas por la variación aleatoria bajo el modelo planteado. Zhong et al. (2014) conduce un estudio de caso-control con niños (de uno a dieciséis años) para comprobar si la TFG declinaba no linealmente cuando la ERC se aproximaba al estadio cinco. La conclusión principal es que para los casos (niños en TRR) se observó una declinación del 26% más rápida de la TFG antes (18 meses) de llegar al estadio cinco en

comparación con los controles (niños con ERC pero sin TRR cuando el correspondiente caso inició TRR).

Métodos estadísticos utilizados

2.1. Notación

A continuación se presenta la notación, adaptada conjuntamente de Deville & Saporta (1983) y Chevenet et al. (1994), de los principales conceptos necesarios para la comprensión de los métodos utilizados en este trabajo:

Sea $[0, T]$ un intervalo de tiempo, denominado *periodo de estudio*; $t = 1, 2, \dots, T$ periodos (cortes) de tiempo en el que se segmenta el periodo de estudio; $j = 1, 2, \dots, m_t$ categorías de una *variable longitudinal cualitativa* (VLC) medidas en el periodo t , para $i = 1, \dots, n$ individuos que las seleccionaron. A partir de estos elementos se definen las siguientes tablas:

- A_t : tabla de duraciones para el periodo t de n individuos (filas) por m_j categorías (columnas).
- A : tabla de duraciones en el periodo de tiempo, formada por la yuxtaposición de las T tablas A_t , constituida por n filas y $m = \sum_{t=1}^T m_t$ columnas.

El elemento a_{ij}^t de la matriz A , representa el tiempo en el que el individuo i permanece en la categoría j durante el periodo t , , mientras que $a_{ij} = \sum_{j=1}^{m_t} a_{ij}^t$ equivale a la duración en cada periodo t .

Se tiene entonces una VLC medida en diferentes instantes t del periodo de estudio $[0, T]$. Por otro lado, A_t contiene las categorías que asume el i -ésimo individuo para cada periodo t .

A partir de A_t y A es posible obtener correspondientemente dos nuevas tablas:

- P_t : tabla de porcentajes de permanencia para el periodo t de n individuos (filas) por m_j categorías (columnas).
- P : tabla de porcentajes de permanencia global formada por la yuxtaposición de las T tablas P_t , constituida por n filas y m columnas.

El elemento $p_{ij}^t = \frac{a_{ij}^t}{a_{i.}^t}$ de P (ver figura 2.1), representa la proporción del periodo t en el que el individuo i permanece en la categoría j de la VLC.

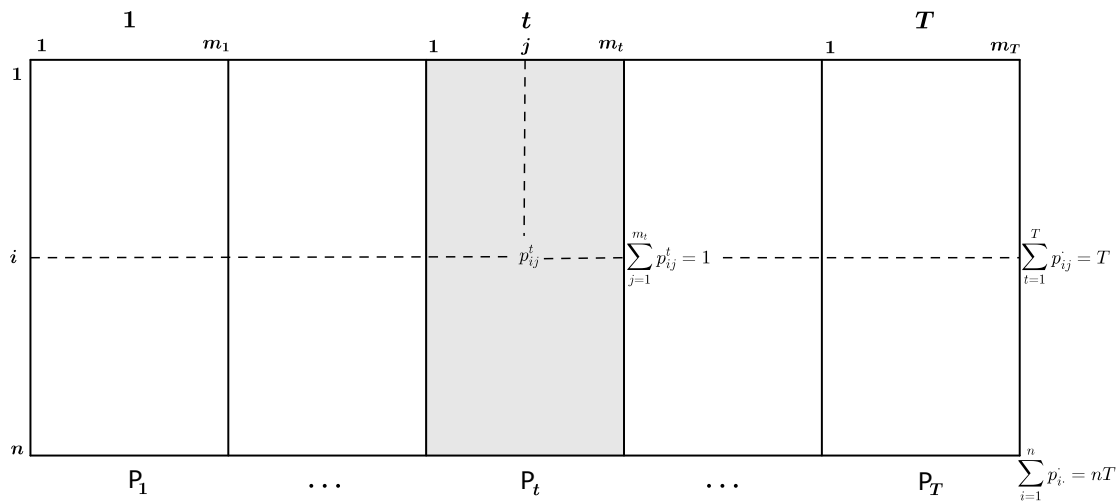


FIGURA 2.1. Tabla global de proporción de tiempo de permanencia en cada categoría de la VLC .

Dentro de las principales propiedades de la tabla P se destacan:

- $p_{i.} = \sum_{j=1}^{m_t} p_{ij}^t = 1$ (La suma de las proporciones de permanencia en un periodo particular t es 1).
- $p_{i.} = \sum_{t=1}^T p_{ij}^t = T$ (La suma de las proporciones de permanencia en cada uno de los t periodos es igual al número de periodos).
- $p_{..} = \sum_{t=1}^T p_{i.} = nT$ (La suma total de la tabla es igual al producto del número de individuos por el número de periodos de tiempo).

Finalmente, se define

$$\bar{p}_j^t = \frac{1}{n} \sum_{i=1}^n p_{ij}^t \quad (2.1)$$

que representa el promedio de proporciones de tiempo que permanecen los n individuos en la categoría j durante el periodo particular t .

2.2. Análisis armónico cualitativo

El análisis armónico cualitativo (AAC) es un método exploratorio multivariado de datos longitudinales cualitativos. Fue presentado por primera vez en un artículo por Daville & Saporta (1980) y luego divulgado con mayor detalle por Deville (1982). Por otro lado, Saporta (1996) realiza una presentación de este método aplicada al análisis de datos biográficos.

El objetivo principal del AAC es estructurar la información de los individuos de una muestra durante un periodo de tiempo determinado. Tales individuos se describen por la cronología de una transición de cambios en etapas o estados, entre un conjunto finito de etapas posibles. Esta transición entre etapas definen una trayectoria temporal que caracteriza a los individuos, de este modo el propósito final del AAC es conseguir una tipología de estas trayectorias.

En Daville & Saporta (1980), Deville & Saporta (1983) muestran que los datos longitudinales cualitativos pueden ser analizados ya sea, tomándolos como procesos escalares (a través del análisis armónico) o como un conjunto finito de variables transversales (a través de un análisis canónico generalizado o un análisis de correspondencias), que dicho en otras palabras, muestra la equivalencia entre el AAC y un análisis factorial de correspondencias particular (Barbary & Pinzón 1998).

2.2.1. Principios del AAC

Se considera la VLC como una realización de un proceso estocástico cualitativo en un tiempo continuo, denotado $X^t(i)$ de m categorías, donde $X^t(i)$ es el estado, entre los m estados posibles que asume un individuo i en un instante t . El siguiente paso es asociar a

todas las categorías de cruce de los diferentes estados y los instantes de observación con un código real $Y(t) = f_t(X^t(i))$. Una codificación en los reales del proceso cualitativo es una función que se define como:

$$f : X \times T \longrightarrow \mathbb{R} \quad (2.2)$$

$$(x, t) \longrightarrow y = f_t(x) \quad (2.3)$$

En otras palabras, el proceso cualitativo es cuantificado, de tal modo, que se obtiene un proceso aleatorio con valores reales.

Una vez hallado el proceso $Y(t)$, la siguiente etapa consiste en realizar una *descomposición espectral* del operador de varianza del proceso (descomposición también denominada análisis armónico, de ahí el nombre del método). De este modo, $Y(t)$ se descompone en una serie de variables aleatorias independientes del tiempo y en una serie de funciones del tiempo, denominadas armónicas del proceso (Barbary 1997).

Si se observa con detenimiento las etapas descritas del AAC del proceso $X^t(i)$ emulan las directrices del análisis factorial salvo que, el primero se desarrolla en el contexto del álgebra de operadores, esto es, cuantificar una(s) variable(s) cualitativa(s) para luego realizar una descomposición apropiada de su matriz. Esta situación lleva a pensar en la existencia de algún tipo de relación entre el análisis armónico y los métodos clásicos de análisis de datos.

2.2.2. El AAC como un análisis de correspondencias particular

Deville & Saporta (1983) demuestran que el AAC es una generalización de los métodos clásicos de análisis de datos si se divide el periodo de observación del proceso, tomando intervalos de tiempo en los cuales ningún individuo cambia de estado.

El análisis armónico en tiempo discreto, bajo la restricción del caso, se convierte en el *análisis de correspondencias* (AC) (Benzécri 1973, Escofier & Pagès 1992, Greenacre & Blasius 2006) de una tabla disyuntiva particular de tamaño $n \times mp$. Aquí p es el número de periodos donde el proceso es estable, es decir, donde ningún individuo cambia de estado. Esta tabla es disyuntiva en el sentido de que sus elementos sólo toman dos valores posibles: 1, si el individuo está en el estado considerado durante el periodo y 0, en caso contrario. De

este modo, el AAC es análogo a un *análisis de correspondencias múltiples* (ACM) (Escofier & Pagès 1992, Greenacre & Blasius 2006). Una consecuencia directa de esta solución es el tamaño de la tabla disyuntiva a analizar, que puede resultar inmensa dados los posibles valores que toman los parámetros n , m y p . En la práctica esta situación conlleva a un análisis complejo en términos operativos y de interpretación.

Como alternativa a este enfoque, los autores proponen tomar el periodo de estudio $[0, T]$ y dividirlo en un número razonable de periodos (de igual o distinta longitud) sin tomar en cuenta los cambios de estados de los individuos dentro de dichos periodos. La tabla resultante se transforma luego en una tabla, que es construida calculando la proporción tiempo de permanencia de cada individuo en cada uno de los estados posibles para cada uno de los t posibles periodos de tiempos (Barbary 1997, Barbary & Pinzón 1998). Finalmente, sobre esta tabla se aplica un AC, método que proporciona un medio para resumir la información a través de la selección de los factores de mayor inercia.

No es difícil observar que la tabla de proporción de permanencia sobre la que se aplica el AC es la tabla P definida en la sección 2.1. Esta tabla, como se verá más adelante, se puede interpretar como una *tabla de codificación difusa* que se analiza a través de un AC particular denominado *análisis de correspondencias difuso*.

El cálculo de proporciones de permanencia de los individuos por estados por periodo de tiempo se puede tomar como una codificación de los valores originales. Esta codificación conserva de manera completa la información relativa a las duraciones individuales en los estados, pero no el orden cronológico de los cambios de estado cuando ocurren en un mismo intervalo de codificación.

2.2.3. Consideraciones prácticas del AAC

Existen dos decisiones determinantes que deben tomarse al momento de aplicar el AAC: la longitud del periodo de estudio $[0, T]$ y el número y amplitud de los intervalos de observación en que se divide el periodo de estudio.

La longitud de $[0, T]$ es importante pues determina el grado de *censura* (individuos con trayectorias menores a la longitud del tiempo de observación global que presentan periodos sin datos) y la *pérdida de información* (individuos con trayectorias mayores a la longitud $[0, T]$ que pierden parte de su información). Cuando la longitud de $[0, T]$ tiende a ser muy

grande la censura aumenta considerablemente, mientras que si la longitud de $[0, T]$ tiende a ser pequeña, la pérdida de información es grande.

En Pinzón (1998) se propone establecer un compromiso entre censura y pérdida de información con el ánimo de aportar una solución al problema de la longitud del periodo de estudio. Se propone variar sistemáticamente los tiempos de análisis, determinar el grado de censura y pérdida de información y optar por aquel periodo cuya longitud logre un balance entre ellos. Puede resultar importante complementar esta solución (estadística) a través de un criterio basado en argumentos propios de la disciplina involucrada y de la temática de estudio y sus objetivos.

Una vez el periodo de estudio es definido, este debe ser dividido en un número óptimo de intervalos para el análisis. Si el número de intervalos es grande (muchos intervalos de amplitud pequeña), la pérdida de información es mínima, sin embargo, la carga de trabajo aumenta significativamente (se tiene una tabla con muchas celdas vacías) y no se asegura ganancia en la información. Por otro lado, si el número de intervalos es pequeño (pocos intervalos de amplitud grande), una parte valiosa de la información disponible se pierde (Barbary & Pinzón 1998, Robette & Thibault 2008).

La división, discreta y sincrónica, del periodo de estudio es modulable y ofrece diversas estrategias que permiten conservar de la mejor manera la información continua y asincrónica de la variable original. Esto permite adoptar diversas lógicas de división del periodo de estudio, algunas estadísticas, otras propias de la disciplina particular involucrada en el análisis. Como argumento estadístico, en (Barbary & Pinzón 1998) por ejemplo, se propone tomar una decisión con base en la distribución de frecuencias de las transiciones entre estados en el tiempo, de tal modo, que dichas frecuencias de periodo a periodo sean más o menos las mismas. Otra alternativa es dividir el periodo de estudio de acuerdo al tiempo histórico o de calendario, donde la amplitud de los intervalos son escogidos de acuerdo a fechas determinantes en la evolución temporal de los individuos. El caso es que la longitud de cada uno de los periodos no necesariamente deben ser iguales, incluso, para algunos tipos de estudio, por ejemplo el de datos biográficos, esta situación es recomendable.

2.2.4. Obtención de tipologías de trayectorias

Es muy probable, que en la mayoría de estudios, el interés central no esté en describir las trayectorias de los individuos de modo separado, sino detectar y describir diferentes tipos de trayectorias individuales. En otras palabras, se desea realizar un análisis tipológico de las trayectorias, esto es, realizar un análisis para identificar las tipología del curso de vida que capturan las trayectorias individuales en su totalidad, y no sólo en términos de los eventos que las componen. Es este sentido, una tipología de trayectoria se constituye en una clase de individuos con trayectorias particulares similares y el trabajo se reduce a un problema de clasificación (Barbary 1997, Barbary & Pinzón 1998, Robette & Thibault 2008).

Realizado el AC sobre la tabla P de proporciones de permanencia, se realiza una *clasificación automática* (ver Everitt et al. 2011, Lebart et al. 2006) de las trayectorias a partir de las coordenadas de los primeros ejes del AC. Una vez obtenidas las clases, estas se caracterizan a través de un conjunto de variables longitudinales o transversales (que pueden ser cualitativas o cuantitativas) denominadas *variables ilustrativas* y por la misma VLC.

El número de ejes factoriales a retener para la clasificación y el número de clases son dos elementos importantes a tener en cuenta durante este proceso. Para obtener un número óptimo de ejes de clasificación se recomienda realizar varias veces el análisis, variando dicho parámetro y tomando aquella partición de la población que origine la menor cantidad de individuos dentro la clase mayoritaria (Barbary 1996). En cuanto a la elección del número adecuado de tipologías a caracterizar, se recomienda tomar como criterio de decisión observar los saltos del histograma de índices de una clasificación jerárquica, o en su defecto, su dendograma.

2.3. Análisis de correspondencias difuso

El *análisis de correspondencias difuso* (ACD) es una extensión del ACM, propuesto por Chevenet et al. (1994) inicialmente para el análisis multivariado de información biológica y ambiental. En este contexto, el objetivo central del ACD es describir la afinidad de una especie con diferentes categorías de una variable dada a través de puntajes positivos.

De manera general, el ACD permite analizar *tablas de codificación difusa*, es decir, tablas donde un conjunto de individuos pueden asumir varias categorías de una misma variable y cada una de sus entradas está constituida por puntajes positivos (incluido el cero) que representan el grado de pertenencia o relación del individuo con la categoría que asume. En este sentido, la información de un individuo a través de las distintas categorías de una misma variable se torna “difusa”, a diferencia de la codificación en una tabla disyuntiva usual, donde la pertenencia o no a una categoría particular son propiedades mutuamente excluyentes (1: pertenencia; 0: en todos los casos contrarios).

El soporte teórico del ACD descansa en el análisis en componentes principales ponderado (Tenenhaus & Young 1985), notado $ACP(X, M, D)$. Esta tripleta indica que una tabla de datos X es analizada a través de un *análisis en componentes principales* (ACP) utilizando como pesos y métricas las matrices M y D . M se constituye en la matriz de métrica del espacio de los individuos (filas) y de pesos en el espacio de las columnas (variables), mientras que D corresponde a la matriz de métrica del espacio de las columnas y a la matriz de pesos del espacio de las filas.

Dado que Benzécri (1973) introduce el ACM como un ACP simultáneo de dos nubes de puntos ponderados, el ACM puede definirse como un análisis en componentes ponderado (Escofier & Pagès 1992, Pagès et al. 2004). Para el ACM, las matrices X , M y D son:

- $X = nZD_p^{-1}$, con término general $x_{ik} = \frac{n}{n_k} z_{ik}$
- $M = \frac{1}{ns} D_p$, con término general $m_k = \frac{n_k}{ns}$
- $D = \frac{1}{n} I_n$, con término general $d_i = \frac{1}{n}$

Aquí Z corresponde a una tabla disyuntiva completa (TDC) que es una tabla indicadora de término general $z_{ikj} = 1$ si el individuo i asume la categoría k de la j -ésima variable cualitativa ($j = 1, \dots, s$) y $z_{ikj} = 0$ de modo contrario; $D_p = \text{diag}(B)$, donde $B = Z'Z$; n_k es el número e individuos que asumen la categoría k y I_n una matriz diagonal de orden n . De este manera, en notación del ACP ponderado, el ACM queda definido como $ACP(nZD_p^{-1}, \frac{1}{ns} D_p, \frac{1}{n} I_n)$.

En razón de que el ACD se considera una extensión del ACM, entonces se puede concluir que que el ACD se puede ver como un análisis en componente principales ponderado (Gámez 2012). Para el caso particular, el ACD es el ACM de la tabla P de proporción de

permanencia global, por tanto, es posible desarrollar sobre ella un $ACP(X, M, D)$. Para el ACD las matrices X , M y D involucradas en el análisis son:

- X , es la matriz de datos con término general $x_{ij} = \frac{p_{ij}^t}{p_j^t} - 1$
- $M = \frac{1}{T} \text{diag}(\overline{p_j^t})$, la matriz de métrica en el espacio de las filas y de pesos en el espacio de las columnas.
- $D = \frac{1}{n} I_n$, la matriz de pesos en el espacio de las filas y de métrica en el espacio de las columnas.

2.4. Análisis factorial múltiple

El análisis factorial múltiple (AFM) es un método para el análisis multivariado de tablas de n individuos medidos simultáneamente por T grupos de variables (Escofier & Pagès 1984, 1992, Pagès et al. 2004). A cada grupo de variables se le hace corresponder una subtabla y la yuxtaposición de las T subtablas genera una única tabla. En este caso se habla de una tabla múltiple estructurada por conjuntos de variables que admite grupos de variables de distinta naturaleza. Para poder aplicar el AFM es indispensable que las variables de un mismo grupo sean de la misma naturaleza: cuantitativas o cualitativas.

Geoméricamente individuos, variables y grupos de variables pueden verse como nubes de elementos ubicados en distintos espacios euclidianos sobre el que se realiza el análisis: en R^m se ubica la nube de individuos N_m ; en R^n se ubica la nube de variables N_n y en R^{n^2} donde se sitúa la nube de grupos de variables N_n^t .

En principio, es posible que individuos y variables estén afectados por pesos: para un individuo i el peso se denota p_i , bajo la condición $\sum_{i=1}^n p_i = 1$, y m_k el peso asociado una variable k (inicialmente se asume que $m_k = 1$). Los pesos de los individuos se recogen en la matriz $D = \text{diag}(p_i)$, M_t para los pesos de las variables del grupo t y M para todos los pesos de las K variables.

Dentro de los objetivos del AFM están equilibrar la influencia de cada grupo en el análisis global y buscar, como en otros métodos factoriales, un subespacio de menor dimensión al espacio original, donde se represente óptimamente las distancia entre individuos y la estructura de correlación entre variables y entre grupos de ellas.

El AFM se realiza en dos etapas:

Etapla uno. Se realizan varios ACP (estandarizados o no) separados, uno para cada una de las subtablas asociadas a cada uno de los grupos de variables, extrayéndose el primer mayor valor propio λ_t^1 correspondiente a cada conjunto t . Si el conjunto está constituido por variables cualitativas se realiza un ACM. A esta etapa se le denomina *análisis separados*.

Etapla dos. Todas las variables de un conjunto particular t son ponderados por el inverso de λ_t^1 , es decir por $\frac{1}{\lambda_t^1}$. Realizada esta ponderación sobre cada uno de los T conjuntos se realiza un ACP sobre la tabla ponderada. A esta etapa se le denomina *análisis global*.

El aporte que hace cada grupo al análisis global depende del número de variables que contenga: a mayor número, mayor es su influencia, y a su estructura interna: mientras más fuerte es su estructura, mayor es la correlación entre sus variables, y por tanto, mayor es la influencia en la construcción de los factores comunes de variabilidad (Escofier & Pagès 1992). Esta es la razón por la cual se pondera, cada uno de los grupos, con el correspondiente inverso del mayor valor propio extraído durante el análisis separados. Esta ponderación equilibra la influencia de cada una de los grupos en la obtención del primer factor.

2.4.1. Factores extraídos por el AFM

Dentro del AFM un *factor* o eje extraído del análisis global para los n individuos se define como

$$F_s = XM u_s \quad (2.4)$$

donde u_s es el vector propio asociado al valor propio λ de la s -ésima dimensión que resulta de la diagonalización de la matriz $X'MXD$.

Mientras que un factor para las K variables es:

$$G_s = X'D v_s \quad (2.5)$$

donde v_s es el vector propio asociado al valor propio λ de la s -ésima dimensión que resulta de la descomposición en valores y vectores propios de la matriz $XDX'M$.

2.4.2. Análisis de conjuntos de variables

El AFM permite un estudio global de los distintos t conjuntos de variables en el espacio Euclídeo R^{n^2} , a través del operador $W_t D$, donde $W_t = X_t M_t X_t$ es una matriz $n \times n$ y $M_t = \text{diag}(\frac{1}{\lambda_t^1})$. La ponderación de cada conjunto por el inverso de su primer valor propio permite balancear la influencia de cada uno ellos sobre el primer eje del análisis global, dado que se estandariza la inercia sobre dicho eje a 1.

Por otro lado, a cada conjunto de variables t le es asociada una nube N_n^t de individuos en el espacio R^{m_t} , denominada *nube parcial*. Individuos pertenecientes al grupo t , denominados *individuos parciales* y denotados i^t , son analizados precisamente con respecto a t . Dado que R^K puede ser tomado como una suma directa de los R^{m_t} espacios, esto es $R^K = \oplus R^{m_t}$, las N_n^t nubes pueden ser colocadas en un mismo espacio y luego proyectadas sobre los ejes del análisis global (Escofier & Pagès 1992, Bécue-Bertaut & Pagès 2008).

De manera análoga como se hace en el análisis global, en el análisis parcial de conjuntos de variables es posible definir *factores parciales* como

$$F_s^t = \ddot{X}_t M u_s \quad (2.6)$$

donde \ddot{X}_t corresponde a una matriz que contiene las coordenadas de los puntos de la nube N_n^t sobre el conjunto X_t . Elementos de dicha matriz que no contengan dichas coordenadas son llevados a cero.

Las coordenadas de los individuos i^t a lo largo del eje s son denotadas $F_s(i^t)$. Estas coordenadas pueden ser calculadas a partir del factor parcial F_s^t ya que $F_s^t(i) = F_s(i^t)$:

$$F_s^t(i) = F_s(i^t) = \frac{1}{\sqrt{\lambda_s}} \frac{1}{\sqrt{\lambda_t^1}} \sum_{k \in m_t} x_{ik} G_s(k) \quad (2.7)$$

conocida como *fórmula de transición restringida*.

2.4.3. Inclusión de conjuntos de variables cualitativas al AFM

El AFM permite la inclusión al análisis conjuntos de variables cualitativas y de frecuencias, sin embargo, se limita la explicación al primer caso. Esto se logra gracias a la equivalencia entre el ACM y el ACP no estandarizado (Bécue-Bertaut & Pagès 2008) .

Los resultados del ACM pueden ser obtenidos a partir de aplicar un ACP a la tabla de término general $\frac{(z_{ikt}-b_{kt})}{b_{kt}}$, donde $b_{kt} = \sum_{i=1}^n p_i z_{ikt}$ y $\sum_{k \in m_t} b_{kt} = m_t$. Cada categoría l del conjunto t es ponderada por $\frac{b_{kt}}{\lambda_t^1 m_t}$ y el individuo i por p_i .

En el AFM, las categorías son representadas por sus correspondientes centros de gravedad, esta cualidad permite trabajar simultáneamente con categorías e individuos. De manera particular, cada categoría se representa por un punto global (el centro de gravedad de los individuos) y por un punto parcial para conjunto de variables (el centro de gravedad de los individuos parciales).

2.4.4. Distancias entre individuos y fórmula de transición general

El AFM induce una distancia entre los individuos i y i' que corresponde a una suma ponderada separada de distancias restringidas por el tipo de conjuntos de variables (Escofier & Pagès 1992). Si T_C representa una clase compuesta por conjuntos de variables cuantitativas y T_N una clase compuesta por conjuntos cualitativos, la distancia se define como:

$$d^2(i, i') = \sum_{t \in T_C} \frac{1}{\lambda_t^1} \sum_{k \in t} \left(\frac{x_{ikt} - x_{i'kt}}{s_{m_t}} \right)^2 + \sum_{t \in T_N} \frac{1}{\lambda_t^1} \sum_{k \in t} \frac{1}{m_t b_{kt}} (z_{ikt} - z_{i'kt})^2 \quad (2.8)$$

De esta manera un conjunto cuantitativo $t \in T_C$, contribuye a la distancia (entre i y i') a través de un ACP sobre la tabla t , mientras que un conjunto categórico $t \in T_N$, contribuye ella a través de un ACM sobre la tabla t , sin embargo, la ponderación $\frac{1}{\lambda_t^1}$ balancea la influencia de las tablas.

Si el análisis se reduce a conjuntos de variables cualitativas la expresión 2.8 se convierte en:

$$d^2(i, i') = \sum_{t \in T_N} \frac{1}{\lambda_t^1} \sum_{k \in t} \frac{1}{m_t b_{kt}} (z_{ikt} - z_{i'kt})^2 \quad (2.9)$$

La relación entre $F_s(i)$ y $G_s(k)$ está dada por la *fórmula de transición general*:

$$F_s(i) = \frac{1}{\lambda_s} \sum_{t \in T_C} \frac{1}{\lambda_t^1} \left[\sum_{k \in t} x_{ikt} G_s(k) \right] + \frac{1}{\lambda_s} \sum_{t \in T_N} \frac{1}{\lambda_t^1} \left[\sum_{k \in t} z_{ikt} G_s(k) \right] \quad (2.10)$$

que reducida al caso donde sólo se consideran conjuntos de variables cualitativas queda expresada como:

$$F_s(i) = \frac{1}{\lambda_s} \sum_{t \in T_N} \frac{1}{\lambda_t^1} \left[\sum_{k \in t} z_{ikt} G_s(k) \right] \quad (2.11)$$

2.4.5. El AFM para una variable longitudinal cualitativa

Es posible encontrar, en las ciencias sociales y en las ciencias de la vida, estudios donde una o varias variables, incluso de distinto tipo, son observadas en diferentes momentos del tiempo sobre una muestra de una población determinada. Estos estudios, de naturaleza longitudinal, permiten desde su diseño metodológico la generación de grupos de variables. Por ejemplo, se puede considerar como un grupo a todas aquellas variables analizadas dentro de un mismo intervalo de tiempo. Otra posibilidad, es que una única variable sea observada en distintos períodos, y que sea considerada un grupo en cada uno de esos intervalos. En este tipo de diseños el AFM puede constituirse en una herramienta útil para la exploración y descripción de los individuos y de la estructura de las variables (Abascal et al. 1999).

Cuando el diseño establece que los datos del estudio (en una o varias variables) son recolectados para una misma muestra de individuos en distintos instantes del tiempo, se dice que el estudio está basado en un *diseño longitudinal panel* (Kalton & Citro 1995). La ventaja de este tipo de diseño sobre otros diseños longitudinales (por ejemplo, los diseños de medidas repetidas) es la posibilidad de cuantificar y cualificar las variaciones individuales y grupales en el tiempo.

Si un estudio panel está restringido al análisis de una única VLC, cada periodo de estudio (panel) constituye un grupo conformado por las categorías de esta variable. Más aún, si la VLC describe los tiempos de permanencia de un individuo en un estado de un conjunto de estados considerandos, el resultado es una tabla de permanencia A , como la definida en la sección 2.1. Aquí cada subtabla A_t corresponde a los grupos, los cuales contienen los momentos de transición en el periodo t considerando.

Al realizar el análisis parcial del AFM sobre cada subtabla A_t , se está estudiando la VLC en cada momento del tiempo, esto es, se está realizando un análisis transversal de ella en momentos separados.

En el momento que se realiza el análisis global del AFM sobre la tabla A ponderada, se analiza la VLC de manera simultánea, considerando al mismo tiempo cada uno de los periodos en el que fueron observadas. Esta situación lleva a obtener un marco de referencia común, lo que permite estudiar la evolución (trayectoria) de los individuos en el tiempo y la posible interrelación entre periodos.

Al igual que otros métodos factoriales, el AFM proporciona representaciones gráficas y algunas ayudas de interpretación de los datos:

- Plano factorial de correlación entre la VLC (y sus categorías) con los factores, lo que permite interpretar los ejes factoriales.
- Nubes parciales de individuos caracterizados en cada panel de estudio.
- Plano factorial de individuos caracterizado por la VLC y por cada periodo de tiempo: nubes parciales y nube global, donde se puede observar los individuos a través de la VLC (trayectorias).
- Cosenos cuadrados, coordenadas parciales y globales de individuos y categorías, contribuciones a la inercia (Escofier & Pagès 1992).
- Inclusión de variables suplementarias al análisis que enriquecen la caracterización de las trayectorias de los individuos.

2.5. El AFM y el análisis multicanónico

La presencia de factores comunes en un análisis factorial puede ser detectada a través del cálculo de las correlaciones entre los factores globales y los factores del mismo rango correspondientes a cada uno de los t grupos analizados, obtenidos en los respectivos análisis parciales. Si la correlación entre el factor global y los factores parciales es fuerte, el factor global se constituye en una tendencia que está presente a lo largo de todos los grupos. Si la correlación es alta sólo para uno de los grupos, el factor global será un factor particular de dicho grupo. Si esto sucede, se procede a eliminar el mismo como elemento activo del análisis, debido a las diferencias en relación al resto de grupos.

De este modo el AFM puede ser considerado como un análisis multicanónico: sus factores globales son las *variables generales* de un *análisis multicanónico* (Escofier & Pagès 1992), que se interpretan como variables relacionadas con el conjunto de los grupos de variables, mientras que los factores parciales pueden ser interpretados como las *variables canónicas*, que resultan de la combinación lineal de las variables de un grupo más relacionada con la variable general, y se consideran las representaciones de los factores globales en los distintos grupos.

Sea z una variable general y K_t las variables del grupo t . Para analizar su correlación es posible establecer una medida de correlación, notada L_g . Por ejemplo, Carroll (1968) propone que L_g corresponde al cuadrado del coeficiente de correlación, es decir, el coeficiente de correlación entre z y la combinación lineal de las variables del grupo t más correlacionadas con z , notada v_t .

Para el caso del AFM, la medida de correlación entre una variable general y un grupo de variables se define como:

$$L_g(z, K_t) = \sum_{k \in K_t} \text{Inercia de la proyección de } v_k \text{ sobre } z \quad (2.12)$$

También es posible analizar la correlación entre grupos de variables. Sean t y h dos grupos de variables distintos. Para el AFM, una medida de relación global entre dichos grupos está dado por (Abascal Fernández & Landaluce Calvo 2002):

$$RV(t, h) = \frac{\langle W_t D, W_h D \rangle}{\|W_t D\| \|W_h D\|} \quad (2.13)$$

donde $W_t D = P_t P_t' D$, con P_t es la tabla asociada al conjunto t (resp. h), D es una matriz diagonal de pesos de los individuos y $\|W_t D\| = \sqrt{\sum_s (\lambda_t^s)}$ corresponde a la norma de $W_t D$ (resp. h), es decir, la raíz cuadrada de la suma de los primeros s - valores propios del análisis parcial de la tabla t .

El numerador de 2.13, $\langle W_t D, W_h D \rangle$ es el producto escalar entre variables representantes de t y h y es equivalente a L_g (Escofier & Pagès 1992), es decir:

$$L_g(t, h) = \langle W_t D, W_h D \rangle \quad (2.14)$$

Así, RV se obtiene a partir de los coeficientes de correlación lineal entre dos variables cualesquiera. Varía entre 0 (no existe relación entre las variables de los dos grupos) y 1 (las nubes de grupos son homotéticas).

Combinación del AAC y el AFM

Sea P una tabla de código difuso, formada por la yuxtaposición de P_t tablas:

$$P = [P_1, \dots, P_t, \dots, P_T]$$

. Cada P_t contiene m_t ($m = \sum_t m_t$) categorías de una variable longitudinal cualitativa (VLC) en diferentes periodos del tiempo t ($t = 1, \dots, T$). La tabla P está constituida por n filas de individuos y m columnas que corresponden a las categorías de una VLC tomada en T periodos de tiempo, y donde un individuo i , puede tomar en cada periodo t , las categorías de la variable con un diferente grado de pertenencia o relación.

La idea de complementar el análisis armónico cualitativo (AAC) con el análisis factorial múltiple (AFM) es realizar una ponderación extra de la tabla P global de proporciones de tiempo de permanencia, obtenida de A (Corrales 2010, Corrales & Pardo 2015). La estrategia empleada es primero, realizar un AC a cada tabla P_t para extraer el primer valor propio de cada periodo t , para luego ponderar todas las columnas del periodo t de la tabla global P por el inverso del correspondiente primer valor propio. Realizado esto, se realiza un ACP sobre dicha tabla global. Más concretamente:

Etapa 1: T análisis de correspondencias (AC) de las P_t tablas. De cada subtabla se retiene el primer valor propio λ_t^1 . Aquí el AC de P_t es un $ACP(X_t, M_t, D)$, donde:

- X_t es la matriz a analizar, con término general $x_j = \frac{p_j^t}{p_j^t} - 1$.
- $M = \frac{1}{T} \text{diag} \left(\overline{p_j^t} \right)$.

- $D = \text{diag}\left(\frac{1}{n}\right) = \frac{1}{n}I_n$.

Etapa 2: Un ACD a la tabla global P en la que cada columna j se pondera por $\frac{1}{\lambda_t^1}$ de la tabla P_t a la que pertenece. Esta segunda etapa del AAC-AFM se logra realizando un $ACP(X, M, D)$, donde:

- X es la matriz a analizar, con término general $x_{ij} = \frac{p_{ij}^t}{p_j^t} - 1$.
- $M = \frac{1}{T} \text{diag}\left(\frac{\overline{p_j^t}}{\lambda_t^1}\right)$.
- $D = \text{diag}\left(\frac{1}{n}\right) = \frac{1}{n}I_n$.

3.1. Propiedades del AAC-AFM

Dadas las características propias del AAC y del AFM y debido a la ponderación particular que se impone a la tabla analizada, el AAC-AFM adquiere una serie de propiedades específicas. Estas propiedades se deducen en su mayoría del *análisis factorial general* (AFG), marco general que describe el proceso común a todos los métodos en ejes factoriales (Lebart et al. 2006).

3.1.1. Distancia entre individuos

En el AFM la distancia entre dos individuos distintos i y l corresponde a una suma ponderada de las distancias inducidas por cada grupo de variables (Escofier & Pagès 1992). En Corrales (2010) el cuadrado de estas distancias para el AAC-AFM se define como:

$$d^2(i, l) = \sum_{t=1}^T \frac{1}{\lambda_t^1} \sum_{j=1}^{m_t} \frac{nT}{\sum_{i=1}^n p_{ij}^t} \left(\frac{p_{ij}^t}{T} - \frac{p_{lj}^t}{T} \right)^2 = \frac{1}{T} \sum_{t=1}^T \frac{1}{\lambda_t^1} \sum_{j=1}^{m_t} \frac{1}{p_j^t} (p_{ij}^t - p_{lj}^t)^2 \quad (3.1)$$

La expresión de la derecha permite deducir que dos individuos están cercanos si se seleccionan de aproximado la misma proporción de categorías de la VLC en los distintos t periodos de tiempo. Adicionalmente, esta distancia es ponderada por el primer valor propio de los análisis parciales realizados para cada grupo temporal.

3.1.2. Distancia entre categorías

Para dos categorías j y j' de los respectivos periodos t y t' se define (Corrales 2010) el cuadrado de su distancia como:

$$d^2(j^t, j^{t'}) = \sum_{i=1}^n n \left(\frac{p_{ij}^t}{\sum_{i=1}^n p_{ij}^t} - \frac{p_{ij'}^{t'}}{\sum_{i=1}^n p_{ij'}^{t'}} \right)^2 = \sum_{i=1}^n n \left(\frac{p_{ij}^t}{np_j^t} - \frac{p_{ij'}^{t'}}{np_{j'}^{t'}} \right)^2 = \sum_{i=1}^n \frac{1}{n} \left(\frac{p_{ij}^t}{p_j^t} - \frac{p_{ij'}^{t'}}{p_{j'}^{t'}} \right)^2 \quad (3.2)$$

De esta manera, dos categorías de dos periodos distintos estarán cercanas cuando el grupo de individuos que las han seleccionado tienen aproximadamente los mismos perfiles.

3.1.3. Relaciones de transición

La coordenada de la proyección de un individuo sobre un eje particular s es notada como $F_s(i)$, mientras que la coordenada de proyección de las columnas se representa como $G_s(p_t)$. Aplicando la fórmula general de transición del AFG es posible mostrar la relación existente entre ambas coordenadas de proyección dentro del AAC-AFM:

$$F_s(i) = \frac{1}{\sqrt{\lambda_s}} \sum_{t=1}^T \frac{1}{\lambda_t} \sum_{j=1}^{m_t} \frac{p_{ij}^t}{T} G_s(p_t) = \frac{1}{\sqrt{\lambda_s}} \frac{1}{T} \sum_{t=1}^T \frac{1}{\lambda_t} \sum_{j=1}^{m_t} p_{ij}^t G_s(p_t) \quad (3.3)$$

$$G_s(p_t) = \frac{1}{\sqrt{\lambda_s}} \sum_{n=1}^n \frac{1}{\lambda_t} \frac{p_{ij}^t}{np_j^t} F_s(i) \quad (3.4)$$

Estas relaciones, denominadas también *cuasibaricéntricas* (Escofier & Pagès 1984), permiten la representación simultánea de categorías e individuos y ayudan a la interpretación de resultados. Además de esto, permiten la inclusión de individuos y variables ilustrativas en el análisis factorial. En el caso particular del AFM, mediante estas fórmulas es posible acceder a una representación superpuesta de las T nubes de individuos, denominadas *nubes parciales*, definidos por los T análisis separados.

Cuando se restringe el estudio a las categorías de un periodo particular t , se obtiene la representación de la t -ésima nube parcial a través de la fórmula de transición:

$$F_s^t(i) = \frac{1}{\sqrt{\lambda_s}} \frac{1}{\lambda_t} \sum_{j=1}^{m_t} \frac{p_{ij}^t}{T} G_s(p_t) = \frac{1}{\sqrt{\lambda_s}} \frac{1}{\lambda_t} \frac{1}{T} \sum_{j=1}^{m_t} p_{ij}^t G_s(p_t) \quad (3.5)$$

Así, para un periodo t , un individuo i se ubica en el promedio, ponderado por los valores de su perfil, de las coordenadas de todos los puntos columna, dilatado por el inverso del producto de la norma del vector propio con el mayor valor propio del t -ésimo análisis. Esta representación superpuesta de los individuos puede ser extendida a las categorías, las cuales se localizan en el promedio de los individuos que las seleccionaron (Bécue-Bertaut & Pagès 2008).

De manera práctica mediante estas fórmulas de transición, la representación superpuesta de individuos y categorías, permite analizar en los planos factoriales la evolución en el tiempo de los individuos, de acuerdo a la posición de los puntos parciales en dichos planos. A través de los puntos medios (resumen de la información de los puntos parciales) es posible sintetizar toda la información en términos de todos los momentos del tiempo. El grado de proximidad de un individuo con su punto medio, permite determinar si el individuo ha mantenido preferencia por la misma categoría a lo largo de todos los periodos.

3.1.4. Inercias de la combinación

La variabilidad de la nube global respecto a su centro de gravedad recibe el nombre de *inercia total*. Esta inercia se descompone en dos tipos de inercia: la *inercia intra* y la *inercia entre*.

La inercia intra es la inercia de los puntos parciales con respecto a su punto medio en la nube global, esto es, la inercia de las nubes correspondientes a la imagen de cada individuo desde los diferentes puntos de vista, estudiados respecto a los centro de gravedad de cada individuo. Por otro lado, la inercia entre, es la inercia de los centros de gravedad de cada nube parcial respecto al centro de gravedad de la nube global.

El análisis de estas inercias permite realizar un estudio de la estructura de la nube de individuos, estructura inducida por la trayectoria a lo largo de los distintos periodos de tiempo. Por ejemplo, cuando el eje s muestra una estructura común para los diferentes periodos del tiempo, los puntos homólogos (que representan al mismo individuo en cada uno de los periodos) son muy próximos entre ellos y esta partición tiene mínima inercia intra.

En otras palabras, cuando los diferentes periodos de tiempo inducen similares estructuras sobre los individuos de la nubes parciales, puntos homólogos están cercanos entre sí. Esta propiedad global es medida a través de la razón inercia entre/inercia total y es cercana a 1 cuando el eje seleccionado representa una estructura común para los distintos periodos de tiempo.

La inercia total obtenida del AAC-AFM (Corrales 2010) se define como:

$$I = \frac{1}{nT} \sum_{t=1}^T \frac{1}{\lambda_t^1} \sum_{j=1}^{m_t} \overline{p_j^t} \sum_{n=1}^n \left(\frac{p_{ij}^t}{\overline{p_j^t}} - 1 \right)^2 = \frac{1}{nT} \sum_{t=1}^T \sum_{j=1}^{m_t} \sum_{n=1}^n \frac{1}{\lambda_t^1} \frac{\left(p_{ij}^t - \overline{p_j^t} \right)^2}{\overline{p_j^t}} \quad (3.6)$$

El término del cuadrado permite observar que la inercia total depende de las diferencias entre las proporciones observadas y las proporciones esperadas, lo mismo que del primer valor propio del respectivo análisis parcial.

Retomando lo descrito en la sección 2.5, con respecto a la relación entre el AFM y el análisis multicanónico, una medida de correlación entre una variable canónica y las variables generales para el AAC-AFM se define como:

$$L(F_s, F_s^t) = \frac{1}{T} \sum_{t=1}^T \left(\frac{\overline{p_j^t}}{\lambda_t^1} \right) r^2 \left(F_s, \left(\frac{p_{ij}^t}{\overline{p_j^t}} - 1 \right) \right) \quad (3.7)$$

de este modo, dentro de un contexto del análisis multicanónico, los factores globales del AFM, notados como F_s , se consideran las variables generales, mientras que los factores parciales F_s^t , son las variables canónicas. Aquí, el primer factor del AFM maximiza la inercia proyectada de todos los periodos del tiempo, es decir, maximiza $\sum L(F_s, F_s^t)$ (Escofier & Pagès 1992).

3.1.5. Relación entre periodos de tiempo

Poder realizar comparaciones entre periodos de tiempo es uno de los objetivos primarios del AAC-AFM. W_t definido en la sección 2.5 es apropiado para representar las categorías del periodo t . Cada W_t puede tomarse como un punto en el espacio R^{n^2} (Espacio que contiene la nube de los T periodos de tiempo, notada N_T).

N_T se proyecta sobre el subespacio inducido por las variables generales F_s . La coordenada de proyección del periodo t a lo largo de F_s es igual a la medida de relación $L(F_s, F_s^t)$, y al mismo tiempo, la contribución del periodo t a la inercia del eje s (Pagès et al. 2004).

Sean t y h dos periodos de tiempo de tiempo distintos, para el caso del AAC-AFM la medida de relación entre ellos, según la expresión 2.14, es

$$L_g(t, h) = \langle W_t D, W_h D \rangle = \langle P_t P'_t D, P_h P'_h D \rangle = \sum_t \left(\frac{\overline{p_j^t}}{T \lambda_t^1} \right) \sum_h \left(\frac{\overline{p_j^h}}{T \lambda_h^1} \right) \quad (3.8)$$

El coeficiente L_g será mayor en la medida que las categorías de un periodo estén más relacionadas con las categorías de otro periodo.

Aplicación del AAC-AFM

4.1. Descripción de trayectorias de progresión de la ERC para una población de pacientes en un programa de prevención secundaria

El principal objetivo de aplicar la combinación del *análisis armónico cualitativo* y el *análisis factorial múltiple* (AAC-AFM) es realizar la búsqueda de una tipología de las trayectorias de progresión a través de los distintos estadios de la enfermedad renal crónica (ERC), de un conjunto de individuos incluidos en un programa de prevención secundaria de dicha enfermedad. En este contexto, la *variable longitudinal cualitativa* (VLC) activa la constituye los cambios de estadio (y las duraciones en cada uno de ellos) observados para cada individuo durante el tiempo que duró el estudio.

El estudio donde se propone realizar el AAC-AFM, es un estudio observacional descriptivo longitudinal. El tiempo de estudio comienza el 1 de Enero del 2009 y finaliza el 31 de Diciembre del 2014 (duración de 2190 días, o equivalentemente, 73 meses). Durante tal tiempo los individuos observados constituyeron una cohorte dinámica, ya que podían ser admitidos al programa en cualquier momento, siempre y cuando, para su inclusión se asegurara al menos un año de permanencia activa dentro del programa de prevención. Como criterios de inclusión al estudio, además del anterior, se tenían que el individuo debía ser mayor de 18 años, haber ingresado al *programa de clínica de salud renal* (CSR) en estadios de la ERC 3a, 3b o 4 y tener al menos cuatro estimaciones de la tasa de filtrado glomerular

(TFG). Como único criterio de exclusión se definió que el individuo no podía ingresar al estudio si era diagnosticado inicialmente con *cáncer avanzado*.

4.2. Los datos de progresión a través de la ERC

La base de datos fue suministrada a través del doctor Rafael Mauricio Sanabria, co-director de este trabajo y gerente de asuntos científicos de *Renal Therapy Service Latinoamérica* (RTS), y se consolidó a partir de información presente en la historia clínica de RTS (*Renir*[®]) y en bases de datos del *Consortio Salud EPS*. Esta base se caracteriza por recoger información clínica y socio económica relacionada con la atención a 3048 individuos dentro del programa de prevención secundaria de la ERC y estar construida y mantenida bajo estrictos sistemas de control de calidad y seguridad, lo que la convierte en una fuente de información confiable. Por otro lado, su estudio está sujeto a rigurosos criterios de confidencialidad de los individuos, buenas prácticas investigativas y aspectos éticos clínicos que se describen en el apéndice A.

4.2.1. La variable longitudinal cualitativa (VLC) activa

La progresión de un individuo en cada uno de los estadios de su ERC (la VLC activa dentro del análisis) está determinada a través de medidas repetidas de la *TFG* estimada y/o el nivel de creatinina sérica, según el rango de valores que se presentan en la tabla 1.1 de la sección 1.1. De manera más concreta, el estadio de la ERC que asume un individuo en algún instante del tiempo de estudio, es definido por un especialista durante la consulta del individuo a la CSR, después de estimar su *TFG* o su nivel de creatinina. En la base original esta información es presentada en la tabla 4.1 (*E1*: estadio 1, *E2*: estadio 2, *E3a*: estadio 3*a*, *E3b*: estadio 3*b*, *E4*: estadio 4 y *E5*: estadio 5).

Para cada uno de los 3048 individuos que se observaron, el cambio de un estadio a otro define un registro de su trayectoria de progresión en su ERC. Los datos del registro son: fecha de consulta y el estadio en el que fue diagnosticado el individuo. Por ejemplo, para el primer individuo se observa que el 13 de julio del 2010 fue diagnosticado en estadio 3*b*, en la siguiente consulta (el 13 de agosto del mismo año) no hubo cambio de estadio, y así sucesivamente, hasta llegar a su última consulta, el 10 de septiembre del 2012, donde se estableció que estaba en estadio 4. Para efectos prácticos, se asume que la permanencia de

Individuo	Fecha Consulta	Creatinina	TFG	Estadio
1	13/07/2010	1.90	36.66	E3b
1	13/08/2010	1.90	36.21	E3b
1	14/09/2010	2.00	34.31	E3b
1	14/10/2010	2.10	32.05	E3b
1	17/11/2010	2.20	29.74	E4
1	14/12/2010	1.90	35.53	E3b
1	14/01/2011	1.90	35.31	E3b
1	22/02/2011	2.00	34.88	E3b
1	22/03/2011	2.00	34.88	E3b
1	11/04/2011	1.80	38.74	E3b
1	11/05/2011	1.86	36.83	E3b
1	13/06/2011	2.14	31.31	E3b
1	14/07/2011	2.10	31.30	E3b
1	16/08/2011	2.24	29.70	E4
1	16/09/2011	2.10	32.16	E3b
1	20/10/2011	2.10	31.97	E3b
1	26/11/2011	2.30	28.78	E4
1	07/02/2012	2.20	29.97	E4
1	22/02/2012	2.13	31.44	E3b
1	08/03/2012	2.20	29.65	E4
1	10/04/2012	2.11	31.77	E3b
1	07/06/2012	1.90	35.84	E3b
1	10/09/2012	2.60	24.93	E4
2	26/09/2011	1.90	37.16	E3a
2	26/12/2011	1.60	45.11	E3a
2	22/03/2012	1.50	48.61	E3a
2	21/06/2012	1.60	45.27	E3a

TABLA 4.1. VLC activa: cambios de estadio de la ERC, *TFG* estimada y creatinina sérica (primeros dos individuos)

un individuo en un estadio específico de la ERC va desde la fecha donde se determinó este estadio hasta un día antes de la siguiente consulta.

4.2.2. Variables transversales y longitudinales ilustrativas de la ERC

Adicional a la VLC activa, se presenta en la base de datos para cada paciente un conjunto de variables clínicas y socioeconómicas, que en términos del análisis a realizar y a los objetivos propuestos, constituirán un conjunto de variables no activas, denominadas *variables ilustrativas* (VI). Se consideran no activas en el sentido de que no son utilizadas para realizar los cálculos dentro del análisis factorial y la clasificación automática, sino que tendrán como propósito ayudar a caracterizar las distintas clases de trayectorias, una vez se definan, a través de los estadios de la ERC.

Las VI se clasifican en *variables transversales ilustrativas* (VTI) y en *variables longitudinales ilustrativas* (VLI). Las VTI (TFG inicial, TFG media, TFG final, edad, días aportados al estudio, género, nivel de escolaridad, diagnóstico de cáncer, antecedente de enfermedad cardiovascular, causa de la ERC, diagnóstico de tensión arterial alta, diag-

nóstico de diabetes mellitus y causa de egreso del estudio) son observadas una única vez dentro del estudio y se consideran constantes a lo largo de todo el análisis. Las primeras cinco VTI son de naturaleza cuantitativa (TFG inicial, TFG media, TFG final, edad y días aportados al estudio), mientras que las restantes ocho (género, nivel de escolaridad, diagnóstico de cáncer, antecedente de enfermedad cardiovascular, causa de la ERC, diagnóstico de tensión arterial alta, diagnóstico de diabetes mellitus y causa de egreso del estudio) son variables cualitativas. La estructura de la tabla que las describe se presenta en la tabla B.1 del apéndice B.

Por otro lado, las VLI (TFG, nitrógeno ureico en la sangre, hemoglobina, tensión arterial sistólica, tensión arterial diastólica, creatinina sérica y el índice de masa corporal) son variables que cambian en el tiempo de estudio, constituidas por medidas repetidas de las mismas para cada uno de los individuos en cada una de las consultas realizadas. A diferencia de las VTI, pero al igual que la VLC activa, las VLI definen trayectorias en el tiempo, salvo que en lugar de ser cualitativas, todas son de naturaleza cuantitativa. La estructura de la tabla que describe las VLI se presenta en la tabla B.2 del apéndice B.

Finalmente, la definición operativa (ver tabla B.3) y un análisis descriptivo de cada una de las VI se realiza en el apéndice B.

4.3. Etapa de preprocesamiento de los datos

Con el propósito de alcanzar los objetivos propuestos y desarrollar el análisis armónico cualitativo para luego combinarlo con el análisis factorial múltiple, es necesario realizar primero un conjunto de adaptaciones a los datos originales, sobre todo, a los que describen la VLC. Esta fase constituye la etapa de pre procesamiento, ajuste de los datos y decisiones de orden metodológico.

4.3.1. Ajustes de información y reestructuración de la tabla VLC

Uno de los criterios de inclusión al estudio era que el individuo debía ingresar al inicio del mismo en estadio *3a*, *3b* o 4 de la ERC, sin embargo, a lo largo de su progresión era posible que éste asumiera otros estadios, por ejemplo, estadio 1 o 2 si su condición mejoraba o estadio 5 si ocurría lo contrario. Tras revisar la proporción de individuos que asumieron,

en algún momento del estudio, el estadio 1 se constata que dicha proporción es muy baja en relación a los otros estadios asumidos (aproximadamente 2.2%), por lo cual se decide unir dicho estadio al estadio 2 consolidando uno solo (notado como $E2$). Otra razón por la cual se toma esta decisión es que, desde el punto de vista clínico, el interés gira en torno a la progresión de los individuos que en su ERC presentan TFG estimadas menores, esto es, estadios $3a$, $3b$, 4, 5, son remitidos a terapia de reemplazo renal TRR o a trasplante renal.

Por otro lado, la información suministrada inicialmente en la tabla 4.1 da cuenta de la fecha (día, mes, año) en que un individuo asistió a una consulta y fue incluido en un estadio particular de la ERC. Sin embargo, esta información resulta redundante en el sentido de que podía suceder que dicho individuo no evolucionara en su ERC en un intervalo de tiempo determinado (por ejemplo, el individuo 1 permaneció en estadio $3b$ entre la primera y cuarta consulta), por tal razón, se decide realizar un primer ajuste a la tabla VLC de acuerdo al siguiente criterio: estadios en un mismo nivel contiguos en el tiempo, para un mismo individuo, se colapsan en uno solo, tomando como tiempo inicial la fecha de la primera consulta y como tiempo final la fecha del día inmediatamente anterior a la consulta donde se dictaminó un estadio distinto al estadio descrito inicialmente. De esta manera, a partir de este criterio y volviendo a la tabla 4.1, el individuo 1 permaneció estable en el estadio $3b$ desde el 13/07/2010 hasta el 16/11/2010, por lo cual se consolida un sólo estadio ($E3b$) para el intervalo de tiempo descrito. Un extracto de la tabla de la VLC con este primer ajuste se muestra en la tabla 4.2. Se excluyen la estimación de la creatinina y de la TFG por no considerarse necesarias dentro del análisis.

Individuo	Fecha.Inicio	Fecha.Final	Estadio
1	13/07/2010	16/11/2010	E3b
1	17/11/2010	13/12/2010	E4
1	14/12/2010	15/08/2011	E3b
1	16/08/2011	15/09/2011	E4
1	16/09/2011	25/11/2011	E3b
1	26/11/2011	05/01/2012	E4
1	06/01/2012	06/02/2012	E3b
1	07/02/2012	09/04/2012	E4
1	10/04/2012	08/07/2012	E3b
1	09/07/2012	10/01/2013	E4
2	26/09/2011	22/10/2012	E3a
3	04/05/2010	17/03/2011	E3a
3	18/03/2011	02/04/2012	E4

TABLA 4.2. VLC activa: cambios de estadio de la ERC, primer ajuste (primeros tres individuos)

Una primera decisión metodológica consiste en dejar a un lado el tiempo calendario y considerar un tiempo propio del estudio, en términos de días. De esta forma, todos los individuos comienzan el estudio en un tiempo inicial común que se denominó *día cero*, y corresponde al día en que ingresan al mismo (fecha de la primera consulta). A partir de este punto común se cuentan los días (*duración*) en que un individuo particular permanece en un estadio determinado hasta cambiar a otro estadio, lo mismo que el número días que el individuo permaneció dentro del estudio.

El hecho de trasladar todos los tiempos a un día cero genera censuras a derecha. Dada la naturaleza dinámica de la cohorte en cuestión (sección 4.1) muchos individuos no totalizaron los 2190 días que duró el estudio, lo cual llevó a que tuvieran un tiempo donde no se les observó (antes de su primera consulta y/o después de haber salido del estudio). De acuerdo a esto, se tienen cuatro tipos de individuos: aquellos que completaron los 2190 días, es decir, permanecieron todo el tiempo del estudio; aquellos individuos que llegaron tiempo después del inicio del estudio y salieron al finalizar el mismo; otros que llegaron al inicio del estudio pero salieron prematuramente por diversas razones (según la variable *causa de egreso de estudio* descrita en la tabla B.3); y finalmente, aquellos que llegaron tiempo después del inicio del estudio pero no lo finalizaron. A los dos primeros tipos de individuos se les asigna una nueva categoría, denominada *categoría de censura A* (CA), con el propósito de describir su situación particular, mientras que a los individuos del segundo tipo se les asigna la categoría *categoría de censura B* (CB).

Adicional a las categorías de censura CA y CB, se asigna una nueva categoría, notada como TRR, a aquellos individuos que salieron del estudio porque fueron remitidos a alguna terapia de reemplazo renal o a trasplante. Si bien, estos individuos se pueden categorizar dentro de CB (se censuran por salir tempranamente del estudio), el interés clínico que ellos suscitan es mayor que la de las otras causas de egreso del estudio, de ahí la necesidad de generar una nueva categoría para ellos.

Al igual que con los estadios de la ERC, para las categorías de censura CA y CB y para TRR, para cada individuo se calculan sus duraciones en días dentro del estudio. Tras este nuevo ajuste la tabla de la VLC activa queda reestructurada tal como se muestra en 4.3. Cabe anotar que si se suman las duraciones, tanto de los estadios como la de las censuras para cada individuo, se obtiene el mismo resultado: 2190 días.

Individuo	Estadio	Dia.Inicio	Dia.Final	Duracion
1	E3b	0	7	7
1	E4	7	34	27
1	E3b	34	279	245
1	E4	279	310	31
1	E3b	310	381	71
1	E4	381	422	41
1	E3b	422	454	32
1	E4	454	517	63
1	E3b	517	607	90
1	E4	607	792	185
1	CB	792	2190	1398
2	E3a	0	272	272
2	CB	272	2190	1918
3	E3b	0	198	198
3	E4	198	579	381
3	CB	579	2190	1611

TABLA 4.3. VLC activa: duración y transición entre estadios de la ERC. Segundo ajuste (primeros tres individuos)

La inclusión de estas tres nuevas categorías a la VLC activa puede ser criticada, en el sentido de que se está introduciendo una homogeneidad a derecha que no es propiamente la que interesa (la progresión de los individuos a lo largo de sus estadios de la ERC), sino debida precisamente a la ausencia de una trayectoria observable a lo largo del periodo de estudio. Obviamente esto incidirá directamente en los resultados del análisis tipológico de las clases de progresión de la ERC. Una primera solución a esta situación es tomar como variables ilustrativas estas censuras, de tal modo que no influyan en los cálculos del análisis factorial; una segunda solución es retirar del análisis aquellos individuos con CA o CB de muy larga duración y una tercera, recortar a derecha el tiempo de estudio.

Para efectos del estudio, se decide mantener CA, CB y TRR y recortar a derecha el tiempo de estudio. Dos razones para incluir CA, CB y TRR en la VLC activa son las siguientes: primero, si el análisis propuesto fuera de tipo inferencial y predictivo resultaría evidente que la inclusión de estas tres categorías vendrían a ser “parásitas” y claramente añadirían ruidos no deseados al análisis, sin embargo, cabe recordar que el objetivo que se persigue es de orden descriptivo y tipológico; una segunda razón, parte precisamente del propósito tipológico del que se hace mención, dado que resulta deseable poder diferenciar a aquellos individuos que por alguna razón salieron del estudio antes tiempo (o progresaron muy positivamente o progresaron muy negativamente) de aquellos que se mantuvieron “estables” y terminaron el estudio, pues se supone que sus trayectorias y los factores que las motivaron son precisamente distintas.

4.3.2. La tabla P: proporción de tiempo de permanencia en un estadio de la ERC

Tal como se indicó en la sección 2.2.2, Deville y Saporta (1982, 1983) demuestran que si el tiempo de estudio se discretiza adecuadamente, el análisis armónico es equivalente a un análisis de correspondencias sobre una tabla específica. En el caso particular, la tabla 4.3 descrita anteriormente se reestructura nuevamente con el ánimo de transformarla en una tabla que tuviese la forma de la tabla 2.1. Esto no es otra cosa que una codificación de la VLC activa, que permite describir la proporción de tiempo de permanencia de cada individuo en cada estadio de su ERC (y la proporción de tiempo no observado, CA, CB o TRR según sea su caso) en un periodo de corte del tiempo de estudio determinado. De este modo, se asegura aplicar el AAC sobre los datos como un análisis de correspondencias difuso (ACD).

En primer lugar, para lograr la transformación deseada, se deben definir unos periodos de corte del periodo de estudio, de acuerdo a los criterios establecidos en la sección 2.2.2. Uno de estos criterios consiste en construir los intervalos de tal manera que en cada uno de ellos se recoja más o menos la misma frecuencia de transiciones entre estadios y censuras. La decisión del número y extensión de dichos periodos se basa en dos gráficas: un histograma de frecuencias absolutas y otro de frecuencias acumuladas de transiciones entre estadios y censuras, que se muestran en la figura 4.1.

Al observar los histogramas se puede constatar que los primeros 1200 días del estudio recogen aproximadamente el 27% del total de transiciones presentadas a lo largo del estudio, mientras que en los últimos 245 días el porcentaje acumulado de transiciones registradas es apenas cercano al 0.98%. Estas dos situaciones llevan a tomar una segunda decisión metodológica: recortar el tiempo de estudio a 1820 días (120 a izquierda y 250 a derecha), es decir, pasar de seis años a cinco años.

La alta frecuencia de transiciones entre estadios observada durante los primeros 120 días del estudio tiene una explicación de orden clínico: existe una importante variabilidad entre las primeras estimaciones de la *TFG*, y en consecuencia, en la asignación a un estadio determinado de la ERC. La razón principal de esta variabilidad es el conocimiento casi nulo de la evolución del individuo en su ERC con la que viene antes de entrar al programa de prevención y de los factores clínicos y socio demográficos que inciden en dicha evolución. A

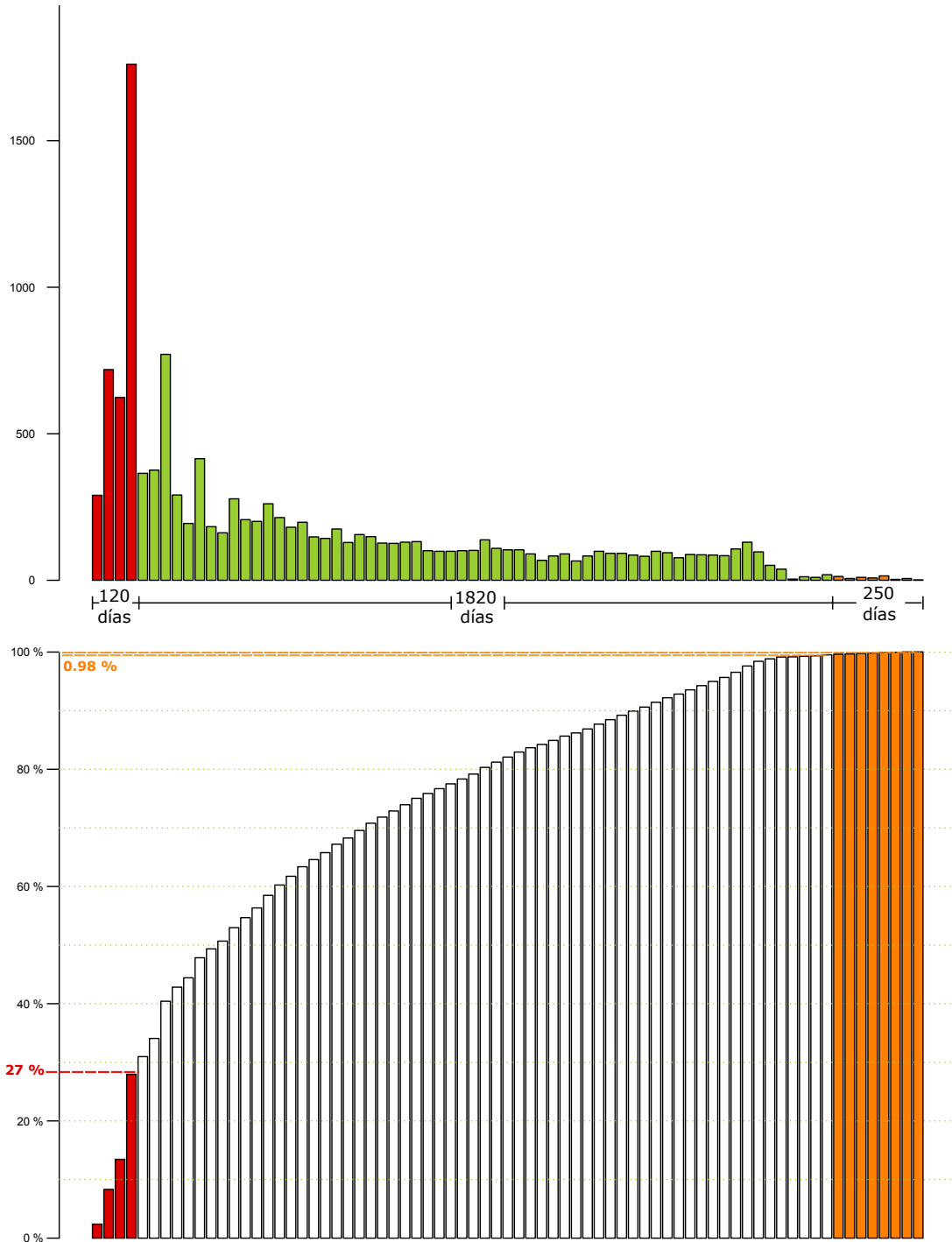


FIGURA 4.1. Histograma de frecuencias absolutas de transiciones entre estadios de la ERC y censuras (parte superior) e histograma de frecuencias acumuladas de transiciones entre estadios de la ERC y censuras (parte inferior).

medida que aumenta la permanencia del individuo en el estudio, aumenta el conocimiento de su progresión en la ERC, y por ende, crece el control sobre la misma por parte de los

especialistas, lo que relativamente permite estabilizar hasta cierta medida sus cambios de estadio (Wouters et al. 2015).

Para explicar la baja frecuencia de transiciones entre estadios de la ERC y las censuras al final del estudio, vale la pena realizar las siguientes observaciones: durante todo el periodo de estudio (sobre todo al inicio y a mediados) se presentan frecuentes transiciones entre los distintos estadios de la ERC (de *E3a* a *E2* o de *E5* a *E4*, por ejemplo) y transiciones de estadios a categorías de censura (por ejemplo de *E2* a *CA* o de *E5* a *TRR*). Si bien, la transición de estadio a censura es posible, la transición de censura a estadio no lo es (no tiene sentido pasar de *CA*, *CB* o *TRR* a cualquier estadio de la ERC). Durante los últimos meses del estudio, muchos de los individuos se censuraron, razón por la cual dejaron de hacer transición a otros estadios, mientras que los restantes individuos siguieron progresando entre estadios, sólo que en una mucho menor proporción, lo que llevó a una menor variabilidad.

Recortar el tiempo de estudio implicó reducir también el número de individuos observados de 3048 a 2445. Este recorte se realizó con el propósito de cumplir el criterio de inclusión que dictaminaba que al inicio del estudio los individuos debían estar en estadios *3a*, *3b* o *4*. De este modo, en el nuevo punto cero del tiempo de estudio aquellos individuos que no cumplieron este criterio fueron retirados.

Después de los ajustes descritos, se opta por construir catorce periodos de corte, todos ellos de distinta longitud. Cada periodo de corte (*P1*, *P2*, *P3*, *P4*, etc.) recoge aproximadamente el 7 por ciento del total de transiciones presentes (Ver tabla 4.4).

Periodos	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	Total
Días por corte	92	107	102	93	88	93	101	115	120	123	149	156	209	252	1825
% de transiciones	7.14	7.14	7.15	7.13	7.16	7.14	7.10	7.16	7.15	7.14	7.16	7.12	7.14	7.15	100

TABLA 4.4. Longitud de cortes del periodo de estudio y proporción de cambios de estadio retenidos

En total se registran 9745 transiciones entre estadios y entre estadios y censuras a lo largo de todo el periodo de estudio (ver tabla C.2 del apéndice C). La tabla 4.4 con los periodos de corte permite observar que los periodos iniciales son los más cortos y que a medida que el tiempo avanza, dichos periodos de corte se hacen más largos.

La tabla 4.5 presenta los tiempos de permanencia (en días) de cada uno de los individuos en los estadios de la ERC y las censuras en cada uno de los catorce periodos de corte definidos anteriormente. Esta tabla corresponde a la denominada tabla *A* definida en la

sección 2.1 y es construida utilizando el programa estadístico R (R Core Team 2016), a través de la librería `qha` (Corrales & Pardo 2016).

individuo	P1.E3a	P1.E3b	P1.E4	P1.E5	P1.TRR	P2.E3a	P2.E3b	P2.E4	P2.E5	P2.TRR	P3.E3a	P3.E3b	P3.E4
1	0	0	65	27	0	0	0	107	0	0	0	0	80
3	0	0	92	0	0	0	0	106	1	0	0	0	0
4	0	0	92	0	0	0	0	107	0	0	0	78	24
5	0	27	0	65	0	0	107	0	0	0	0	102	0
6	0	5	87	0	0	0	107	0	0	0	0	102	0
7	0	92	0	0	0	35	72	0	0	0	102	0	0
8	0	23	69	0	0	54	53	0	0	0	37	65	0
9	0	92	0	0	0	0	107	0	0	0	0	102	0
10	0	92	0	0	0	0	107	0	0	0	0	51	51

TABLA 4.5. Tabla A: duración en días de cada individuo en un estadio de la ERC o en censura para cada periodo de corte (primeros diez individuos y tres periodos de corte)

Finalmente, para aplicar el AAC sobre los tiempos de permanencia se transforma la tabla 4.5 en una tabla de código difuso, tabla que representa la proporción de tiempo de duración de los individuos en los estadios de la ERC y las censuras por cada uno de los catorce periodos de corte. Esta nueva tabla, que se denomina tabla P , se construye utilizando la librería `ade4` (Dray et al. 2007) de R, se presenta a continuación:

individuo	P1.E3a	P1.E3b	P1.E4	P1.E5	P1.TRR	P2.E3a	P2.E3b	P2.E4	P2.E5	P2.TRR	P3.E3a	P3.E3b	P3.E4
1	0.00	0.00	0.71	0.29	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.78
3	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.99	0.01	0.00	0.00	0.00	0.00
4	0.00	0.00	1.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.76	0.24
5	0.00	0.29	0.00	0.71	0.00	0.00	1.00	0.00	0.00	0.00	0.00	1.00	0.00
6	0.00	0.05	0.95	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	1.00	0.00
7	0.00	1.00	0.00	0.00	0.00	0.33	0.67	0.00	0.00	0.00	1.00	0.00	0.00
8	0.00	0.25	0.75	0.00	0.00	0.50	0.50	0.00	0.00	0.00	0.36	0.64	0.00
9	0.00	1.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	1.00	0.00
10	0.00	1.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.50	0.50

TABLA 4.6. Tabla P: proporción de días de permanencia de cada individuo en un estadio de la ERC o en censura para cada periodo de corte (primeros diez individuos y tres periodos de corte)

La tabla 4.6, originalmente constituida por 2445 filas (individuos) y 112 columnas (14 periodos de corte por 8 categorías), se reduce a 102 columnas, dada la presencia de 10 columnas nulas (estadios o censuras con duración cero en un periodo de corte determinado) no informativas.

4.4. Análisis descriptivo de perfiles de permanencia global en estadios de la ERC

La tabla P conserva la información relativa a las duraciones individuales en cada uno de los estadios de la ERC y las censuras definidas, pero no del orden cronológico de las transiciones entre estadios y censuras (las trayectorias) dentro de cada periodo de corte,

sin embargo, resulta una importante herramienta si el propósito es explorar y describir la proporción de tiempo de permanencia en los estadios y las censuras en cada uno de los periodos de corte del tiempo de estudio de todos los individuos bajo observación. Con tal propósito se elabora un diagrama de perfiles global basado en la tabla P , que se presenta en la figura 4.2. El perfil se elabora utilizando la librería **FactoClass** (Pardo & Del Campo 2007) de R. El diagrama de perfil se complementa con un diagrama de niveles (librería **ggplot2** de R (Wickham 2016)) que presenta gráficamente la distribución global de las transiciones entre estadios y censuras a lo largo de los periodos de corte (figura 4.3) y se basa en la tabla C.2 del apéndice C.

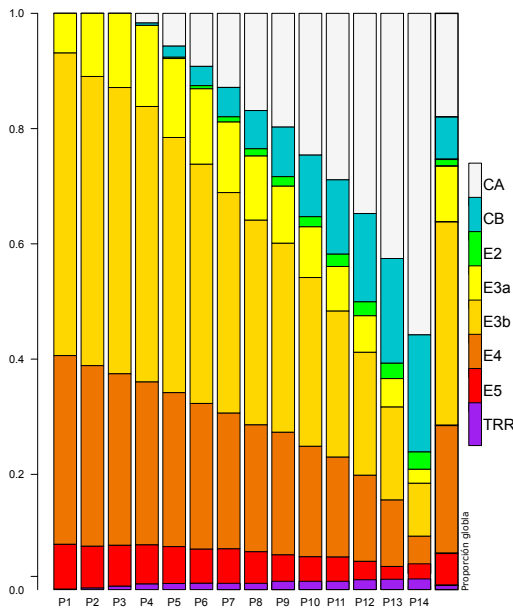


FIGURA 4.2. Perfil global de tiempo de permanencia en los estadios de la ERC y censuras por periodos de corte

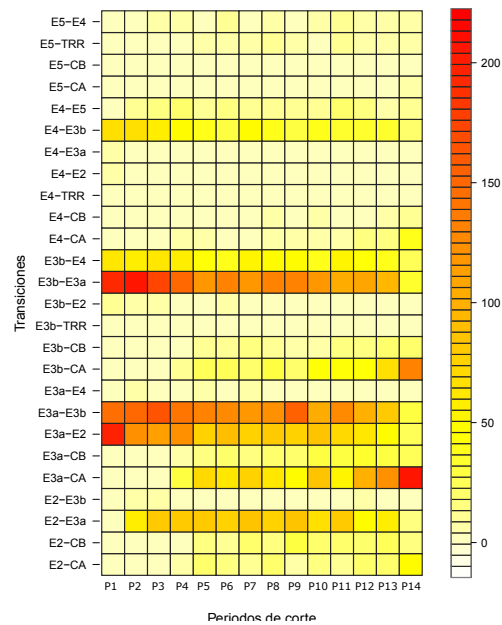


FIGURA 4.3. Transiciones entre estadios de la ERC y censuras por periodos de corte

La última barra a la derecha del perfil presenta la proporción global de permanencia en cada uno de los estadios y censuras a lo largo de todo el estudio de la población observada. Las mayores proporciones de permanencia se presentan en los estadios $3b$ (32.11 %) y 4 (22.43 %). Con menor frecuencia se presentan, en su orden, permanencias en la categoría de censura CA (17.15 %), los estadios 2 (10.29 %) y $3a$ (7.53 %), censura CB (5.15 %) y finalmente, estadio 5 (4.11 %) y TRR (1.23 %) (ver tabla C.1 del apéndice C). A partir de esta información también puede reportarse que 1724 de los 2445 individuos

observados finalizaron el estudio, 628 salieron prematuramente del mismo (por muerte, remisión a consulta externa, pérdida de seguimiento, et.) y 93 fueron incluidos en terapia de reemplazo renal (dos de ellos a trasplante de riñón). Debe aclararse, por ejemplo, que 1314 registrado para $E2$ hace referencia al número de veces que este estadio fue asumido en algún momento del periodo de estudio, y no el número de individuos que comenzaron o finalizaron en $E2$.

De manera general, el perfil global muestra que a lo largo de los catorce periodos de corte los estadios con mayor permanencia son $E3b$ y $E4$, seguidos de permanencias en $E3a$ y $E5$ respectivamente. A medida que el tiempo de estudio avanza, estas proporciones disminuyen paulatinamente, mientras que la proporción de individuos en CA y CB van en aumento. La razón de esta situación puede deducirse de la figura 4.3: a medida que el tiempo de estudio aumenta, individuos en estadios $E3a$, $E3b$ y $E4$ (que son los más frecuentes) van finalizando su tiempo estudio (CA) o se censuran de manera prematura (CB). Por ejemplo, la transición $E3a-CA$ verifica un aumento progresivo en su frecuencia a partir de $P4$, alcanzando un pico en $P14$, donde el número de individuos que estando en $E3a$, y que se censuran por finalizar su tiempo de estudio, es mayor a 200. Situaciones similares se observan con las transiciones $E3b-CA$, $E2-CA$, $E3a-CB$, $E3b-CB$, entre otras.

Si bien se observa una disminución progresiva de las permanencias en cada uno de los estadios, debido al proceso natural de censura ya explicado, se puede afirmar que la progresión de la población general a través de la ERC es relativamente estable. Las razones que sustentan esta afirmación, nuevamente se pueden encontrar en la gráfica 4.3, donde se puede observar que las transiciones complementarias (por ejemplo, la transición $E3a$ a $E4$ es complementaria de la transición $E4$ a $E3a$) tienen más o menos la misma distribución de frecuencias a lo largo de todo el tiempo de estudio. De este modo, el número transiciones del estadio $E3a$ a $E3b$ en cada uno de los periodos de corte es aproximadamente el mismo que las transiciones de $E3b$ a $E3a$; situaciones similares se presentan con las parejas $E4-E3b$ y $E3b-E4$ y $E3a-E2$ y $E2-E3a$ respectivamente.

Pese a que los tiempos de permanencia en el estadio 5 y el número de remitidos a terapia de reemplazo renal (TRR) resultan mínimos en comparación con la permanencia en los otros estadios y categorías de censura (lo cual en términos del programa de prevención resulta ideal), su importancia clínica obliga a mirarlos con mayor detalle.

Para $E5$, como ya se indicó, se observa una disminución progresiva en el tiempo del número de individuos que estuvieron en este estadio. La mayor cantidad de transiciones hacia este estadio se hacen desde $E4$, aunque su transición complementaria ($E5 - E4$) presenta aproximadamente la misma distribución de frecuencias a través de todos los periodos de corte, lo cual muestra estabilidad entre estas transiciones.

En cuanto a la TRR , las primeras remisiones a estas terapias se presentan a partir de $P2$ y se prolongan hasta final de estudio con una frecuencia aproximadamente constante. En términos de transiciones, se puede afirmar que los individuos que fueron remitidos a terapia de reemplazo renal era porque estaban en $E5$, $E4$ y $E3b$, ya que transiciones como $E3a - TRR$ y $E2 - TRR$ no están presentes en la figura 4.3.

A modo de conclusión, en cuanto a la progresión a través de los estadios de la ERC de la totalidad individuos bajo estudio, se puede afirmar que esta es positiva, en el sentido que la progresión se mantuvo estable en cada uno de los estadios a lo largo de cada uno de los periodos de corte (no se mostró tendencia de progresar de un estadio particular a un estadio de mayor gravedad) y que la proporción de individuos en las categorías más críticas ($E5$ y TRR) es muy baja. Esta situación en general, habla bien de la pertinencia del programa de prevención y seguimiento de la ERC a la que fueron incluidos los sujetos bajo estudio, pues la evidencia confirma el cumplimiento de uno de sus objetivos principales: retrasar en mayor medida la progresión de los individuos a estadios terminales de la ERC.

4.5. Análisis factoriales aplicados

Construida la tabla P de proporción de permanencias en los estadios de la ERC y definidas las variables transversales y longitudinales ilustrativas, la siguiente etapa es aplicar análisis factoriales sobre dicha tabla. Cabe recordar que el orden establecido para este análisis es primero realizar un AAC sobre P como si fuera un análisis de correspondencias difuso (ACD), para luego combinarlo con un AFM (capítulos 2 y 3 respectivamente), y a partir de las coordenadas factoriales resultantes, realizar una clasificación automática sobre ellas. El objetivo final es encontrar las distintas clases de progresión a través de la ERC, en términos de trayectorias individuales semejantes de los sujetos bajo estudio, para luego caracterizar dichas clases por medio de las VTI y las VLI.

4.5.1. Aplicación del análisis armónico cualitativo

La tabla P (que describe la VLC activa codificada) es sometida a un AAC a través de un análisis de correspondencias difuso (ACD) en concordancia a las características inherentes a dicha tabla, descritas en la sección 2.3. Este AC particular se realizó utilizando la función *fuzzy {gha}* implementada en R. Como ya se indicó, la tabla P es una tabla de 2445×102 y no fue objeto de más modificaciones.

La figura 4.4 presenta histograma de valores propios del AAC, mientras que la figura 4.5 muestra el primer plano factorial donde se proyectan simultáneamente individuos y estadios y censuras en los distintos periodos de corte. Estos primeros resultados del AAC se complementan con el porcentaje de inercia proyectada sobre los primeros ejes factoriales (tabla 4.7).

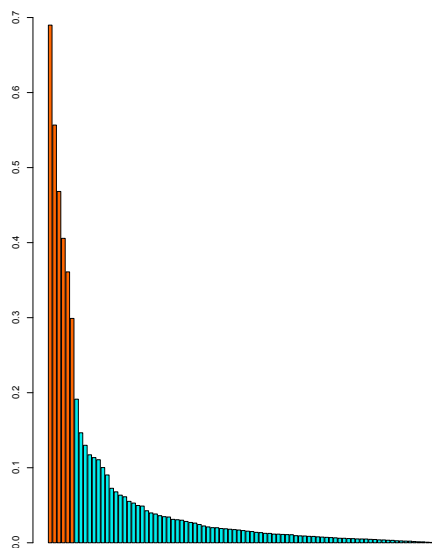


FIGURA 4.4. Histograma de valores propios del AAC de la tabla P

Eje	V. propio	Inercia acum.	Prop. inercia
1	0.690	0.690	13.4
2	0.557	1.247	24.1
3	0.468	1.715	33.2
4	0.406	2.121	41.1
5	0.361	2.482	48.1
6	0.299	2.781	53.9
7	0.191	2.972	57.6
8	0.147	3.118	60.4
9	0.130	3.248	62.9
10	0.117	3.366	65.2
11	0.114	3.479	67.4
12	0.111	3.590	69.5
13	0.100	3.690	71.5
14	0.090	3.780	73.2
15	0.073	3.853	74.6
16	0.068	3.921	75.9
17	0.063	3.984	77.2
18	0.061	4.046	78.4
19	0.055	4.101	79.4
20	0.053	4.154	80.4

TABLA 4.7. Proporción de inercia proyectada sobre los primeros 20 ejes del AAC

El histograma de valores propios sugiere retener los primeros seis ejes factoriales para el análisis (existe un salto considerable entre la sexta y séptima barra). Estos primeros seis factores recogen un poco más de la mitad de la inercia de la nube (53.9%), sin embargo, hay indicios de que el análisis se puede priorizar sobre los dos primeros ejes (primer plano factorial), pues la proporción de inercia retenida es relativamente alta ($13.4\% + 10.8\% =$

24.1%). Lo anterior no implica dejar a un lado el análisis de los demás ejes y planos factoriales que se generan (ver figura D.1 del apéndice D).

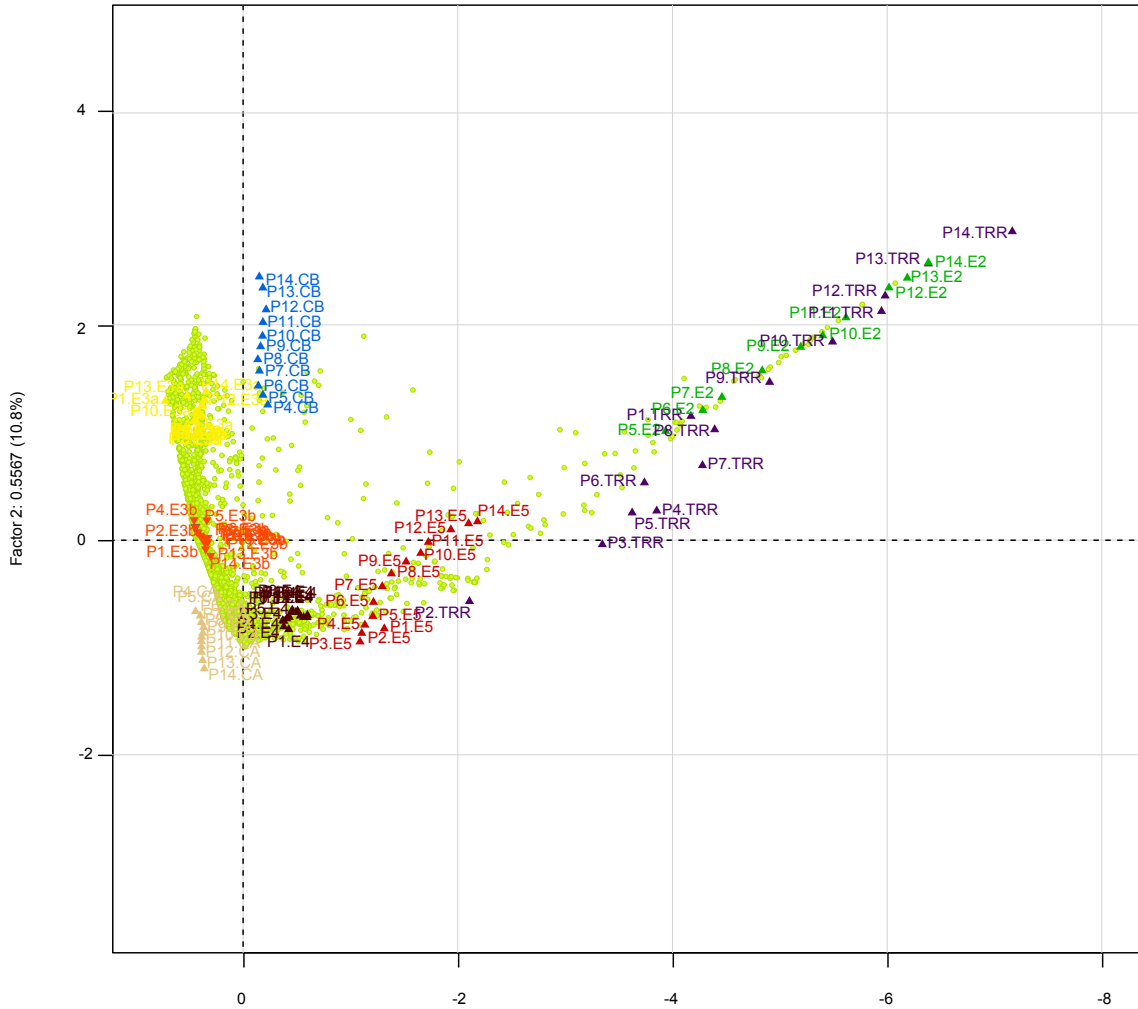


FIGURA 4.5. Primer plano factorial del AAC con individuos y estadios y censuras por periodos proyectados

Cabe recordar que el ACD es un análisis de correspondencias (AC) particular. En un AC los centroides de las categorías asumidas con mayor frecuencia por los individuos se ubican cerca al origen del plano factorial, mientras que categorías con baja frecuencia tienen centroides alejados de dicho origen; para la interpretación de los dos primeros ejes factoriales del ACD se debe tener en cuenta esta propiedad y la ubicación de las categorías de la VLC activa sobre el primer plano factorial (figura 4.5): el primer eje se determina esencialmente por la permanencia en los estadios *E2*, *E4*, *E5* y *TRR*, mientras que el segundo eje se caracteriza por las permanencias en los estadios *3a*, *3b* y las categorías de

censura CA y CB . Ambos ejes agrupan y ordenan cronológicamente estadios y censuras a través de los periodos de corte: los primeros periodos del estudio se ubican cerca del origen del plano, mientras que los últimos periodos tienden a ubicarse lejos del mismo.

En términos de los individuos (figura 4.5), el primer eje separa a los individuos que permanecieron mayor tiempo en estadios 4 y 5 en su progresión de la ERC y se censuraron por finalizar el estudio (CA) de los que estuvieron mayor tiempo en estadio $E3a$, $E2$, salieron de manera temprana del estudio (categoría de censura CB) y fueron remitidos a TRR . El segundo eje, por otro lado, separa a los individuos que permanecieron mayor tiempo en estadios $E3a$ y $E3b$ y finalizaron el estudio (CA) de aquellos que permanecieron la mayor parte del estudio en estadios $E2$, $E4$ y $E5$, y se censuraron prematuramente (categoría CB y TRR). Esta situación indica que el segundo eje separa a los individuos que progresaron positivamente en su ERC (semiplano izquierdo) de aquellos individuos cuya progresión fue menos favorable (semiplano derecho), salvo el caso de los individuos en $E2$, que están muy próximos a los remitidos a TRR (esta situación quizás obedezca a la baja frecuencia con que están asumidas estas categorías). Finalmente, se aprecia una alta frecuencia de individuos con mayor permanencia en $E3b$ y CA (por estar cercanos al origen del plano), mientras que existe baja frecuencia de aquellos individuos que fueron remitidos a TRR (ubicados muy lejos del origen).

4.5.1.1. Clases de progresión de la ERC a partir del AAC

Con el propósito de realizar un primer acercamiento a las posibles clases de progresión de la ERC y su caracterización tipológica a través sus perfiles, se realiza una clasificación automática en el espacio de los primeros factores del AAC. Para decidir el número de ejes factoriales a utilizar en la clasificación, se tuvo en cuenta el procedimiento descrito en la sección 2.2.3, esto es, variar el número de ejes y escoger aquella partición cuya clase mayoritaria fuese la de menor tamaño, en relación a las restantes particiones. De este modo, se ensayan clasificaciones de 2 a 88 ejes (número de valores propios no nulos que se asocian a dichos ejes factoriales), donde al final se seleccionan 73 ejes para la clasificación, puesto que la partición definida en este espacio tiene como clase mayoritaria un grupo de 765 individuos (clase de tamaño menor en relación a las restantes clases mayoritarias de las demás particiones).

El método de clasificación de las trayectorias individuales que se emplea es un *análisis de clasificación jerárquica a través del criterio de agregación de Ward* (Lebart et al. 2006). Este método permite determinar el número óptimo de clases a través de un histograma de índices de nivel (figura 4.6)

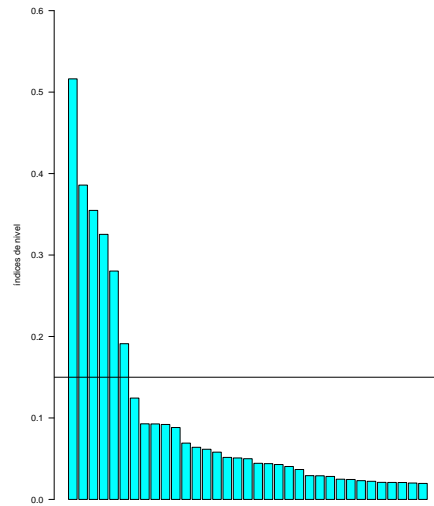


FIGURA 4.6. Histograma de índices de nivel para clasificación en el AAC. Se consideran 7 clases de trayectorias.

En el histograma de índices de nivel se observa un alisamiento después de la sexta barra, razón por la cual se decide considerar siete clases de trayectorias. Una vez definido el número de clases dentro de la partición, éstas se consolidan mediante del método de *K-means* (Lebart et al. 2006), implementado en **R** a través de la función `kmeans{stats}`. La figura 4.7 presenta el primer plano factorial del AAC con las siete clases de progresión proyectadas. No se realizará una descripción detallada de estas clases, por cuanto el interés se centrará en la obtención y caracterización de las clases de progresión de la ERC a partir de la combinación del AAC con el AFM. Sin embargo, cabe importante señalar que la clase mayoritaria es la clase 3, que en tamaño es seguida por las clases 2, 5 y 4 respectivamente, mientras que las clases más pequeñas son, en orden decreciente, la clase 1, la clase 6, y finalmente, la clase 7.

Si se hace corresponder este plano con el primer plano del AAC donde se proyectan conjuntamente estadios y censuras por periodos (figura 4.5), se puede deducir que la clase 1 corresponde a los individuos que salieron anticipadamente del estudio (censura *CB*), la clase 2 corresponde al conjunto de individuos que finalizaron el estudio (censura *CA*), la clase 3 progresó mayoritariamente en *E3b*, la clase 4 corresponde a los individuos que

progresaron en *E3a*, mientras que los individuos de la clase 5 lo hicieron en *E4*, la clase 6 corresponde a individuos con mayor permanencia en *E5* y finalmente, la clase 7, muy seguramente, fueron remitidos en su mayoría a *TRR*.

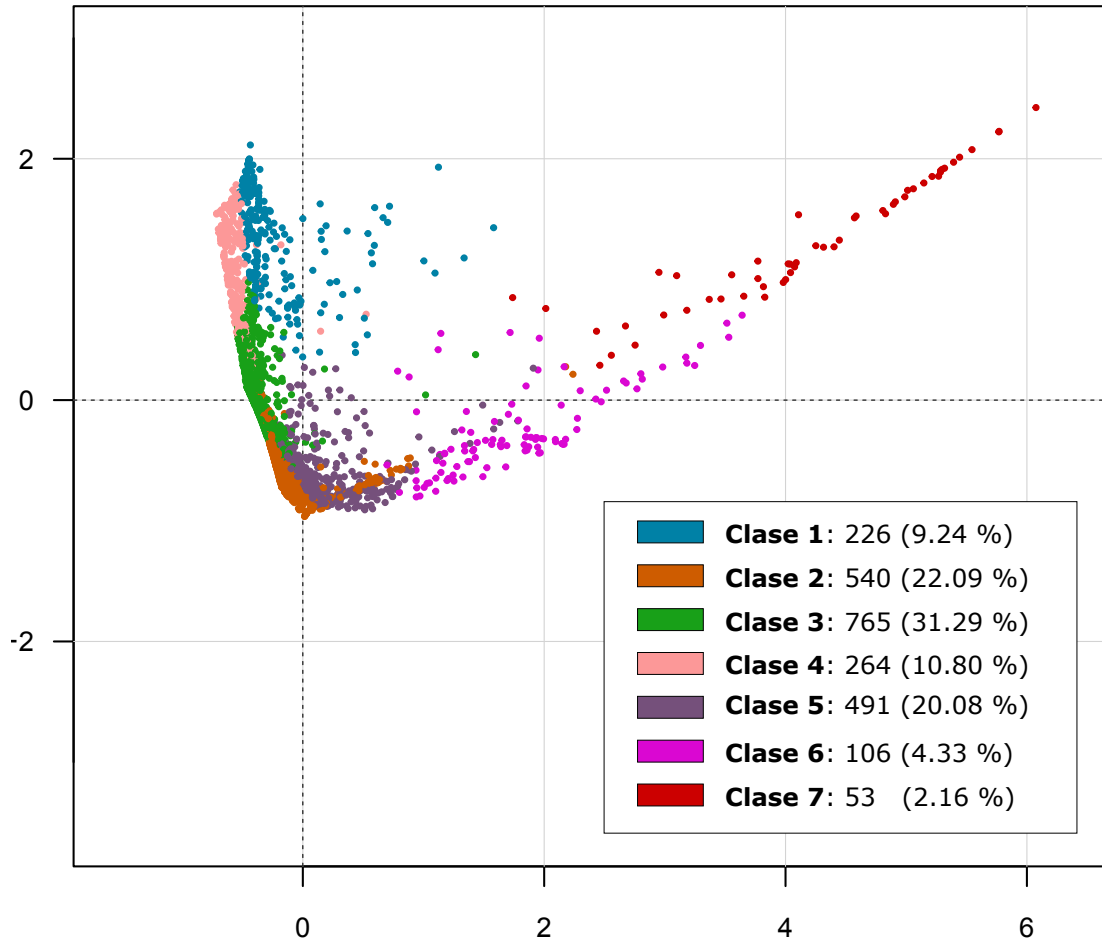


FIGURA 4.7. Proyección de las siete clases de trayectorias de la ERC en el primer plano factorial del AAC

4.5.2. Aplicación del análisis factorial múltiple sobre el AAC

Realizar la combinación del AAC y del AFM (AAC-AFM) tiene múltiples propósitos, que no se alcanzan aplicando únicamente el AAC: a) obtener una síntesis global y una síntesis parcial (por periodos de corte del tiempo de estudio) de toda la información disponible; b) equilibrar la influencia de cada uno de los periodos de corte en la combinación; c) obtener trayectorias parciales de los individuos en el tiempo y, d) asegurar medidas de asociación y semejanza entre los periodos.

Para ejecutar el AAC-AFM se utiliza nuevamente la librería `qha`. Esta librería permite tomar las coordenadas del AAC, en términos de los periodos de corte, y sobre ellos aplicar el AFM, esto es, realiza los análisis separados y el análisis global del AFM descrito en el capítulo 2. También es posible realizar el AAC-AFM a través de la librería `FactoMineR` (Husson et al. 2008) de `R`, mediante el comando `MFA`.

Una primera justificación de aplicar el AFM sobre el AAC, es la disparidad en la influencia que ejerce cada uno de los periodos de corte sobre el análisis global. Esto se aprecia mejor si se observan los primeros valores propios de cada uno de los catorce AAC separados:

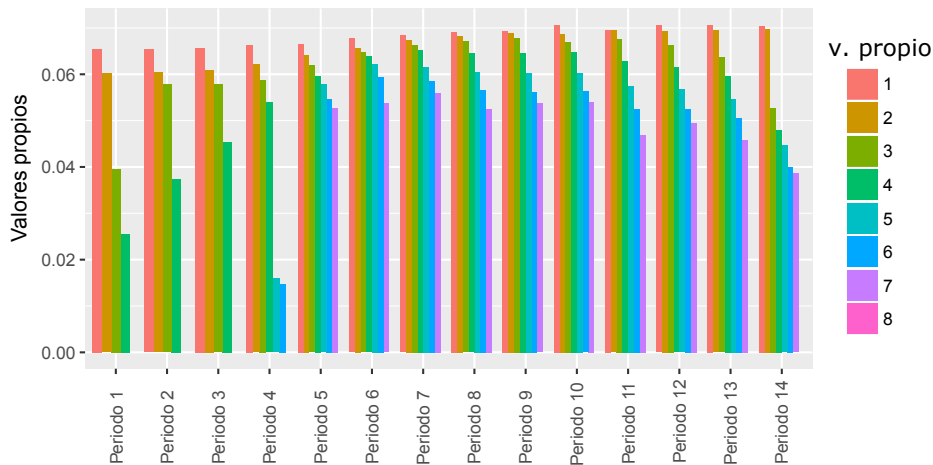


FIGURA 4.8. Valores propios no nulos de los catorce AAC separados del AFM

la influencia de un periodo depende esencialmente de su estructura interna, es decir, en el modo en que su inercia queda distribuida sobre las nubes proyectadas. De esta manera, si uno de los periodos presenta una inercia considerable (representada por sus valores propios) en una dirección específica, tal dirección influenciará fuertemente el primer eje del análisis global (Bécue-Bertaut & Pagès 2008). Por ejemplo, en la figura 4.8 los últimos periodos, que son los periodos con los primeros valores propios más grandes, seguramente determinarán la dirección del primer eje del análisis global, mientras que los primeros periodos (hasta P_5) ejercerán poca influencia en la dirección de dicho eje. La solución desde el AFM, como ya se había indicado, es estandarizar esos valores propios a 1 para el análisis global, ponderando por los pesos (inversos de los primeros valores propios de cada análisis separado), que se describen en la tabla 4.8.

Periodo	Primer v. propio	Peso
1	0.065	15.385
2	0.065	15.385
3	0.066	15.152
4	0.066	15.152
5	0.067	14.925
6	0.068	14.706
7	0.068	14.706
8	0.069	14.493
9	0.069	14.493
10	0.071	14.085
11	0.070	14.286
12	0.071	14.085
13	0.071	14.085
14	0.070	14.286

TABLA 4.8. Primeros valores propios y pesos de los AAC separados

El histograma de valores propios y el porcentaje de inercia proyectada sobre los primeros ejes factoriales del AAC-AFM se presentan en la figura 4.9 y en la tabla 4.9 respectivamente. El histograma sugiere retener seis ejes factores para el análisis, donde se recoge aproximadamente el 53.77% de la inercia acumulada. El porcentaje de inercia proyectada para los dos primeros ejes es del 13.38 y 10.75 respectivamente (muy similar con el AAC), por lo cual resulta válido priorizar el análisis sobre el primer plano factorial.

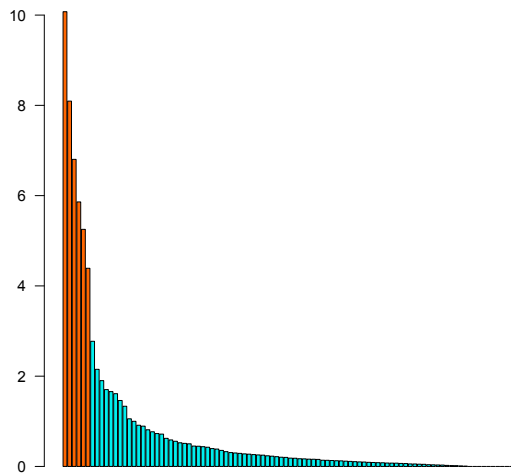


FIGURA 4.9. Histograma de valores propios del AAC-AFM

Eje	V. propio	Prop. inercia	Inercia acum.
1	10.07	13.38	13.38
2	8.09	10.75	24.14
3	6.80	9.04	33.17
4	5.86	7.79	40.96
5	5.25	6.98	47.94
6	4.39	5.83	53.77
7	2.77	3.68	57.45
8	2.15	2.86	60.31
9	1.90	2.53	62.84
10	1.70	2.26	65.10
11	1.66	2.20	67.30
12	1.61	2.14	69.44
13	1.46	1.94	71.39
14	1.33	1.77	73.16
15	1.05	1.40	74.56
16	1.00	1.33	75.89
17	0.91	1.21	77.10
18	0.89	1.19	78.29
19	0.81	1.08	79.37
20	0.77	1.02	80.39

TABLA 4.9. Proporción de inercia proyectada sobre los primeros 20 ejes del AAC-AFM

Por otro lado, el primer plano factorial del AAC-AFM con individuos y centroides de las categorías de las VTI cualitativas se presenta en la figura 4.10.

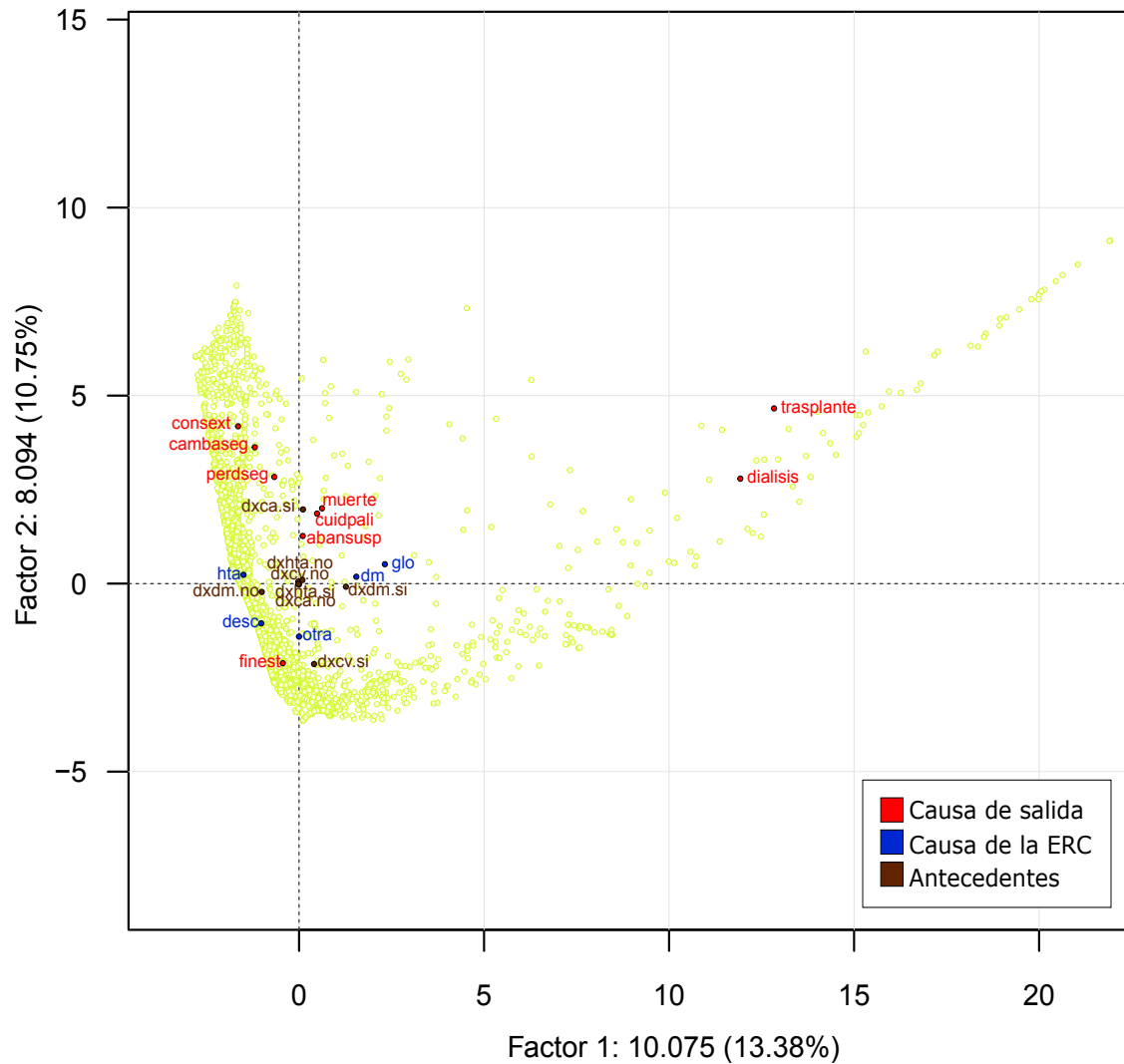


FIGURA 4.10. Primer plano factorial del AAC-AFM con individuos y centroides de las categorías de las VTI cualitativas proyectados

Después de realizar un AFM, el primer trabajo es caracterizar los primeros ejes factoriales a través de las primeras componentes principales del análisis global. Para el caso particular, esto se logra describiendo la manera como los ejes separan a los individuos en términos de las categorías más importantes de las VTI, y se determina, en un círculo de correlaciones (figura 4.11), cuáles periodos y estadios más se relacionan con los dos primeros factores del AAC-AFM.

En el primer plano del AAC-AFM de la figura 4.10, se observa que el primer eje separa a los individuos que finalizaron el estudio de los individuos que salieron anticipadamente del mismo por diversas razones (consulta externa, cambio de aseguradora, pérdida de seguimiento, muerte, cuidados paliativos, abandono o suspensión del tratamiento, diálisis y

trasplante). Este eje separa también a los individuos cuya causa de su ERC es la diabetes, la glomerulonefritis y la HTA (tensión arterial alta) de aquellos cuya causa es otra enfermedad o es desconocida. Por otro lado, el segundo eje separa a los diabéticos de los no diabéticos. Las categorías ubicadas sobre el origen del plano corresponden a categorías cuya influencia de la construcción de los ejes no es determinante, por ejemplo, diagnostico positivo o negativo de HTA y antecedente negativo de enfermedad cardiovascular.

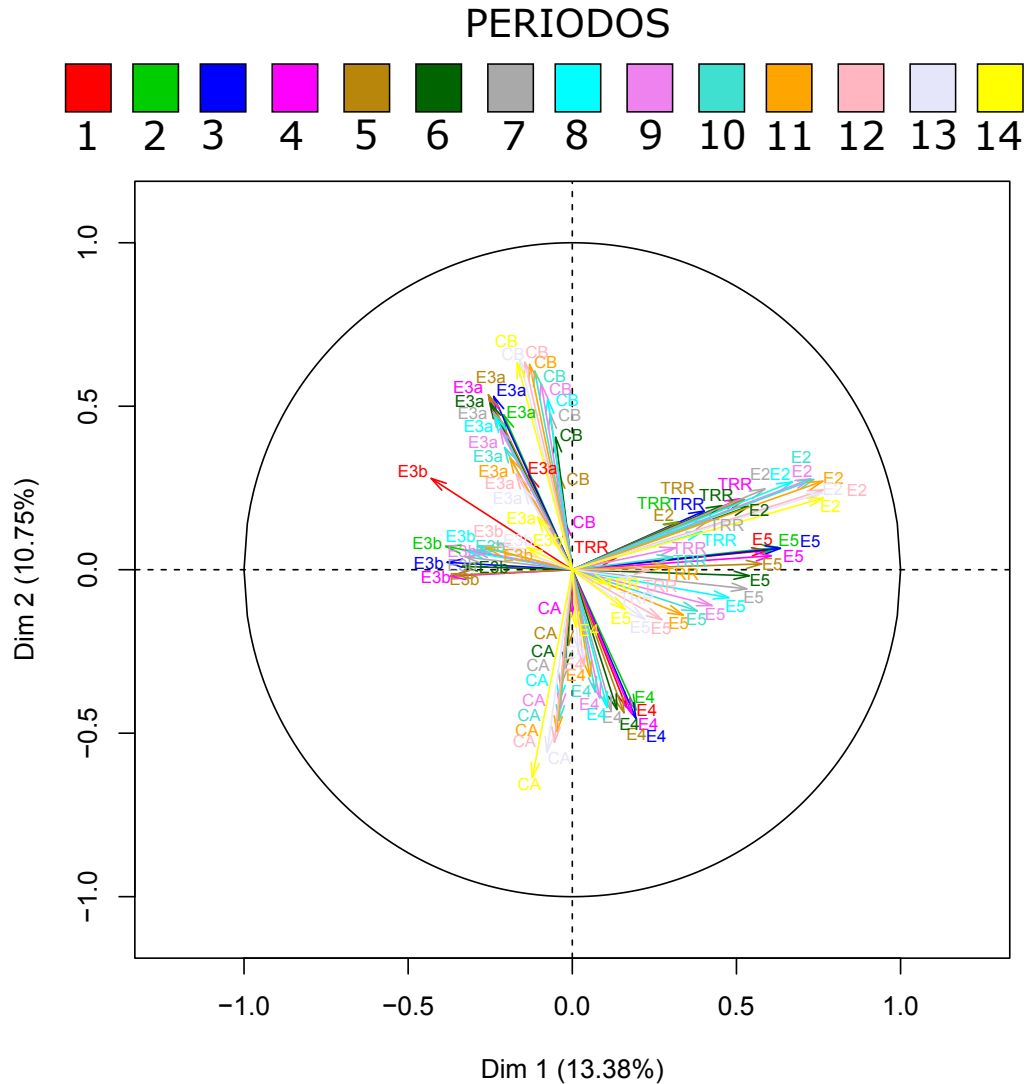


FIGURA 4.11. Círculo de correlación de periodos de corte y estadios del AAC-AFM

De otro lado, en el círculo de correlaciones (figura 4.11) se observa que las componentes más relacionadas positivamente con el primer factor del AAC-AFM (aquellas semirrectas de mayor longitud en dirección al eje) son: las permanencias en los estadios *E2* y *E5* en los últimos periodos del estudio, remisiones a *TRR* en los primeros periodos de corte y

permanencia den $E3b$ durante el primer periodo. Con el segundo factor de AAC-AFM las componentes más correlacionadas positivamente son: las permanencias en $E3a$ y $E4$ durante los primeros periodos de corte y las categorías de censura CA y CB en los últimos periodos. Resulta interesante observar la relación directa entre las parejas de componentes $E3a$ y CB y CA y $E4$ respectivamente, pero de relación inversa entre este par de parejas.

4.5.2.1. Análisis de los periodos de corte del AAC-AFM y grupos de VTI

A partir de los resultados del análisis global del AAC-AFM es posible estudiar diversos aspectos de los periodos de corte del tiempo y de los grupos de variables ilustrativas, en total tres, que se definen de la siguiente manera: El primer grupo de VTI se conforma por las cinco variables cuantitativas (TFG inicial, TFG media, TFG final, edad y el número de días aportados al estudio); el segundo grupo por variables cualitativas socio demográficas (nivel de escolaridad y género), y el tercer grupo, por variables clínicas cualitativas (diagnóstico de cáncer, diagnóstico de HTA, diagnóstico de diabetes mellitus, causa de la ERC, antecedentes de enfermedad cardiovascular y causa de egreso del estudio).

Un primer aspecto que se analiza es la correlación de los periodos y los grupos de VTI con los primeros ejes del AAC-AFM. La tabla 4.10, que presenta los coeficientes de correlación entre los periodos de corte y los primeros tres ejes del AAC-AFM, indica alta correlación de todos los 14 periodos al menos con los dos primeros ejes; mientras que con el tercer eje, la relación es alta con los periodos $P5$ a $P13$.

Grupo	Eje 1	Eje 2	Eje 3
Periodo 1	0.65	0.47	0.34
Periodo 2	0.74	0.59	0.46
Periodo 3	0.78	0.65	0.51
Periodo 4	0.83	0.66	0.59
Periodo 5	0.88	0.76	0.72
Periodo 6	0.90	0.81	0.80
Periodo 7	0.92	0.86	0.85
Periodo 8	0.93	0.89	0.88
Periodo 9	0.93	0.89	0.87
Periodo 10	0.92	0.89	0.86
Periodo 11	0.90	0.88	0.83
Periodo 12	0.87	0.85	0.77
Periodo 13	0.83	0.81	0.71
Periodo 14	0.80	0.75	0.60

TABLA 4.10. Correlación de los periodos de corte con los primeros tres ejes del AAC-AFM

La relación entre los periodos de corte y grupos de VTI con los dos primeros ejes del AAC-AFM se puede estudiar en el plano de la figura 4.12. En este plano, los periodos y grupos de VTI más alejados del origen conservan mayor relación con los ejes. De este

modo, la mayor correlación con los dos primeros ejes la presentan los periodos $P5$ a $P14$, aunque la relación de dichos ejes con los primeros periodos (de $P1$ a $P4$) no es despreciable. En cuanto a los grupos de VTI, los grupos 1 y 3 ($Ilust1$ e $Ilust3$ respectivamente) son los que más se relacionan con los ejes, mientras que el grupo 2 ($Ilust2$, el de las variables socioeconómicas), por estar ubicado cerca al origen, presenta una relación prácticamente nula.

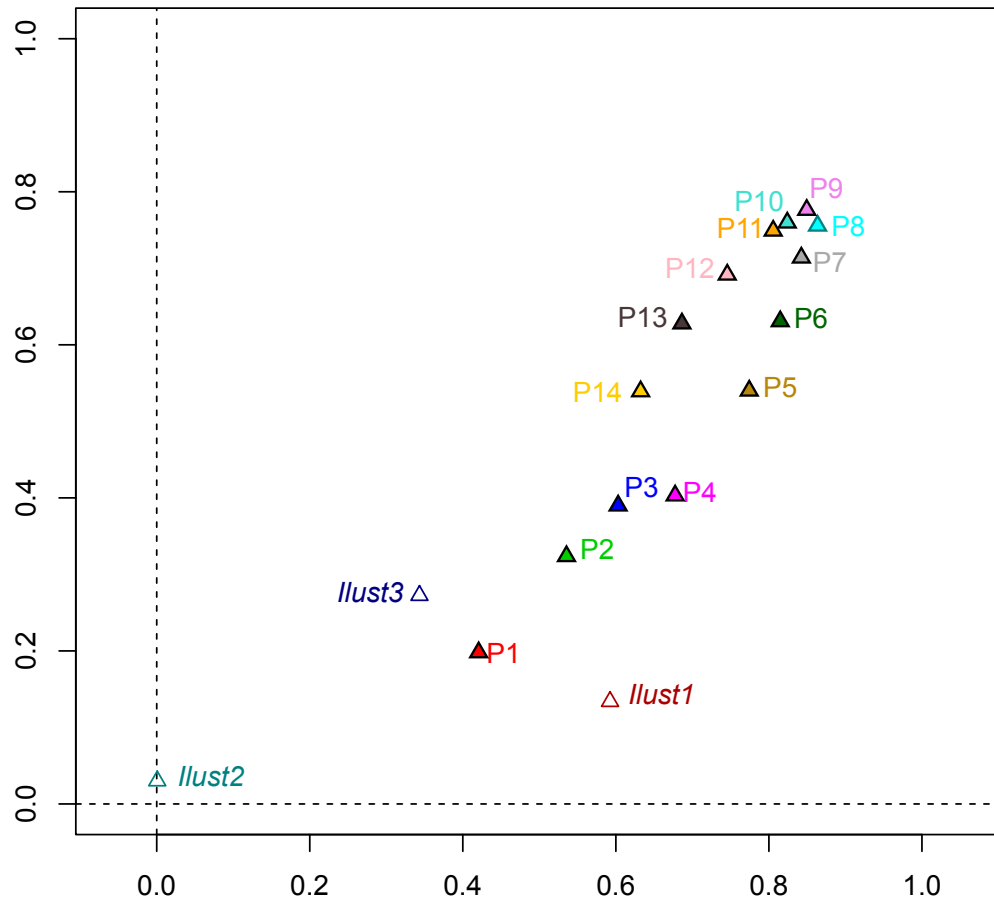


FIGURA 4.12. Proyección de los periodos de corte y grupos de variables ilustrativas en el primer plano factorial del AAC-AFM

Otro aspecto que se puede estudiar, en relación a los periodos de corte y los grupos de VTI, es la contribución de cada uno de ellos a la construcción de los primeros ejes factoriales y la calidad de sus representaciones en los mismos. Coordenadas factoriales de los periodos y grupos (que determinan sus ubicaciones sobre el plano de la figura 4.12), contribuciones y cosenos cuadrados sobre los tres primeros ejes factoriales se describen en la tabla 4.11.

Grupo	Coordenadas			Contrib. absolutas			Cosenos cuadrados		
	1	2	3	1	2	3	1	2	3
Periodo 1	0.42	0.20	0.12	4.18	2.45	1.70	0.07	0.02	0.01
Periodo 2	0.54	0.32	0.19	5.32	4.00	2.74	0.10	0.04	0.01
Periodo 3	0.60	0.39	0.24	5.98	4.82	3.46	0.12	0.05	0.02
Periodo 4	0.68	0.40	0.31	6.72	4.98	4.61	0.13	0.05	0.03
Periodo 5	0.77	0.54	0.48	7.69	6.67	7.07	0.11	0.05	0.04
Periodo 6	0.81	0.63	0.60	8.09	7.79	8.80	0.11	0.07	0.06
Periodo 7	0.84	0.71	0.69	8.36	8.82	10.12	0.12	0.08	0.08
Periodo 8	0.86	0.76	0.73	8.57	9.34	10.79	0.13	0.10	0.09
Periodo 9	0.85	0.77	0.73	8.43	9.53	10.68	0.12	0.10	0.09
Periodo 10	0.82	0.76	0.71	8.18	9.39	10.39	0.12	0.10	0.09
Periodo 11	0.81	0.75	0.66	8.00	9.25	9.67	0.12	0.10	0.08
Periodo 12	0.75	0.69	0.56	7.40	8.55	8.16	0.10	0.09	0.06
Periodo 13	0.69	0.63	0.47	6.81	7.76	6.91	0.09	0.08	0.04
Periodo 14	0.63	0.54	0.33	6.28	6.66	4.89	0.10	0.07	0.03
Ilustrativas 1	0.59	0.13	0.10				0.30	0.02	0.01
Ilustrativas 2	0.00	0.03	0.03				0.00	0.00	0.00
Ilustrativas 3	0.34	0.27	0.14				0.02	0.01	0.00

TABLA 4.11. Ayudas para la interpretación de los periodos de corte y los grupos de variables ilustrativas para el AAC-AFM. Primeros tres ejes factoriales

Existe una alta contribución de todos los periodos de corte del tiempo de estudio a la construcción del primer eje; una situación similar se presenta con el segundo eje, salvo los dos primeros periodos, cuya contribución es baja en relación a los demás. En cuanto al tercer eje, los periodos que más contribuyen a su construcción son del 5 al 14. Si se observa la tabla, las contribuciones absolutas de los grupos de VTI no son reportadas, pues al ser grupos de naturaleza ilustrativa no aportan a la construcción de los ejes dentro del análisis.

En la tabla 4.11, los cosenos cuadrados que se reportan para el primer eje indican que todos los periodos están bien representados sobre este eje (valores mayores a $1/14 \approx 0.07$ sugieren buena representación), mientras que los periodos 6 a 14 y 7 a 11 se consideran bien representados en el segundo y tercer eje respectivamente. En relación a los grupos de VTI, sólo el primer grupo de variables ilustrativas (el grupo de variables cuantitativas) está bien representado sobre el primer eje.

Por otro lado, la matriz RV (tabla 4.12) permite examinar la relación entre los distintos periodos de corte del periodo de estudio, los grupos de variables ilustrativas y la componente común del AFM (ver sección 2.5). Como habría de esperarse, existe una alta correlación entre periodos de corte consecutivos en el tiempo (diagonal inmediatamente superior a la diagonal principal de la matriz RV). Se destaca la dependencia entre los periodos 8 a 11. A su vez, la relación entre los periodos de corte y los grupos de variables transversales es muy baja, situación explicada por el hecho de que los periodos se derivan de una estructura temporal, mientras que las VTI se consideran independientes del tiempo. Finalmente, la

dependencia de los periodos y los grupos de VTI con la componente común del AFM (última columna de la tabla), es relativamente alta en la mayoría de los casos, a excepción de los grupos de ilustrativas 2 y 3.

Grupo	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	Ilust1	Ilust2	Ilust3	AFM	
P1	1.00	0.52	0.35	0.28	0.19	0.15	0.14	0.13	0.11	0.09	0.08	0.07	0.05	0.04	0.35	0.00	0.02	0.33	
P2		1.00	0.62	0.41	0.28	0.21	0.19	0.17	0.15	0.13	0.10	0.09	0.07	0.06	0.35	0.00	0.04	0.43	
P3			1.00	0.62	0.35	0.27	0.23	0.21	0.18	0.15	0.13	0.11	0.09	0.07	0.37	0.00	0.05	0.48	
P4				1.00	0.56	0.35	0.29	0.24	0.21	0.17	0.15	0.12	0.10	0.09	0.37	0.00	0.05	0.52	
P5					1.00	0.64	0.42	0.32	0.26	0.22	0.19	0.15	0.13	0.11	0.28	0.01	0.06	0.58	
P6						1.00	0.67	0.46	0.37	0.30	0.25	0.20	0.17	0.14	0.25	0.01	0.08	0.64	
P7							1.00	0.68	0.49	0.39	0.33	0.26	0.22	0.18	0.24	0.02	0.09	0.69	
P8								1.00	0.72	0.54	0.43	0.34	0.28	0.23	0.23	0.02	0.11	0.72	
P9									1.00	0.72	0.54	0.42	0.32	0.27	0.22	0.02	0.12	0.72	
P10										1.00	0.72	0.52	0.39	0.32	0.21	0.03	0.13	0.71	
P11											1.00	0.70	0.50	0.39	0.19	0.03	0.16	0.69	
P12												1.00	0.71	0.49	0.17	0.04	0.17	0.64	
P13													1.00	0.66	0.15	0.05	0.19	0.58	
P14														1.00	0.14	0.03	0.23	0.49	
Ilust1															1.00	0.01	0.10	0.41	
Ilust2																1.00	0.02	0.03	
Ilust3																	1.00	0.19	
AFM																			1.00

TABLA 4.12. Coeficientes RV de los periodos de corte y de los grupos de variables ilustrativas del AAC-AFM

Finalmente, se analizan las dimensionalidades de los periodos y los grupos de VTI a través de la matriz de coeficientes Lg (ver sección 2.5). En esta matriz, donde también se reporta la componente común del AFM, la atención se centra en los valores de su diagonal principal (tabla 4.13).

Grupo	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	Ilust1	Ilust2	Ilust3	AFM	
P1	2.36	1.39	0.95	0.79	0.71	0.58	0.53	0.47	0.41	0.33	0.29	0.23	0.17	0.13	0.59	0.01	0.09	0.93	
P2		2.96	1.87	1.31	1.14	0.90	0.80	0.71	0.62	0.51	0.42	0.35	0.27	0.20	0.65	0.01	0.15	1.34	
P3			3.11	2.04	1.47	1.17	1.00	0.87	0.75	0.65	0.55	0.44	0.35	0.26	0.71	0.01	0.19	1.54	
P4				3.43	2.47	1.60	1.30	1.08	0.93	0.77	0.66	0.53	0.43	0.33	0.73	0.02	0.22	1.75	
P5					5.65	3.74	2.48	1.84	1.50	1.25	1.04	0.83	0.67	0.52	0.73	0.03	0.33	2.51	
P6						5.99	4.01	2.70	2.16	1.75	1.44	1.12	0.90	0.70	0.67	0.05	0.42	2.85	
P7							6.03	4.05	2.88	2.29	1.87	1.47	1.19	0.89	0.65	0.07	0.52	3.06	
P8								5.80	4.17	3.07	2.44	1.91	1.50	1.11	0.61	0.10	0.61	3.15	
P9									5.83	4.11	3.03	2.33	1.76	1.30	0.57	0.11	0.67	3.16	
P10										5.64	4.00	2.82	2.10	1.51	0.53	0.13	0.73	3.06	
P11											5.47	3.75	2.61	1.82	0.49	0.16	0.83	2.92	
P12												5.30	3.67	2.25	0.43	0.20	0.91	2.68	
P13													5.02	2.95	0.37	0.23	0.98	2.34	
P14														4.03	0.29	0.13	1.07	1.78	
Ilust1															1.18	0.03	0.25	0.80	
Ilust2																3.96	0.07	0.12	
Ilust3																	5.21	0.77	
AFM																			3.28

TABLA 4.13. Coeficientes Lg de los periodos de corte y de los grupos de variables ilustrativas del AAC-AFM

Periodos y grupos de VTI con coeficientes Lg similares sobre la diagonal principal indican la presencia de una estructura común interna similar. De esta manera, un primer conjunto de periodos los constituyen aquellos con dimensionalidad cercana a tres: $P1$, $P2$, $P3$ y $P4$; otro conjunto con dimensionalidad cercana a cuatro: $P14$ e $Ilust2$; un tercer conjunto con dimensionalidad cercana a cinco: $P11$, $P12$, $P13$ e $Ilust3$, y finalmente, un

último conjunto con dimensionalidad seis: $P6$ a $P10$. Esta heterogeneidad lleva a pensar en una inestabilidad de los resultados del AAC-AFM (Abascal Fernández & Landaluce Calvo 2002), sin embargo, al observar los coeficientes Lg de los periodos y grupos con la componente común del AFM (última columna) se evidencian valores relativamente cercanos, lo cual justifica la suposición de que existe al menos un factor común dentro del análisis a los distintos periodos de corte. Además, si se observa la diagonal inmediatamente superior a la diagonal principal, donde se describen los coeficientes Lg de los periodos de corte contiguos en el tiempo, se evidencian valores grandes distintos de cero (si $Lg = 0$ no existe relación entre las variables de un periodo con la componente principal que determina el otro periodo), por tanto, se puede deducir alta relación entre los periodos de corte.

A modo de conclusión, los distintos resultados obtenidos del análisis de los periodos de corte y grupos de variables ilustrativas sobre los dos primeros ejes del AAC-AFM, obligan a pensar que el primer plano factorial se constituye en un escenario adecuado para la obtención y caracterización de las distintas clase de progresión de la ERC, trabajo que se describe en las siguientes secciones. En todo caso, otros planos factoriales del AAC-AFM (los cuatro primeros) se presentan (figura D.2) y analizan en el apéndice D.

4.6. Clases de progresión de la ERC a partir del AAC-AFM

Realizado el AAC-AFM, se efectúa un análisis de clasificación jerárquica bajo el criterio de Ward a partir de la coordenadas de dicho análisis factorial, con el propósito de determinar las clases de progresión a través de los distintos estadios de la ERC. El procedimiento que se utiliza es prácticamente el mismo que se ejecutó para hallar la clases de progresión desde el AAC (sección 4.5.1.1): *a*) se decide el número de ejes factoriales para la clasificación, variando los mismos hasta encontrar aquella partición con la clase mayoritaria de menor tamaño en relación a las demás particiones; *b*) se escoge el número óptimo de clases y, *c*) se consolidan las clases encontradas a través de un K-means. El método completo se desarrolla mediante la función $hqa\{qha\}$ del programa R.

Para realizar la clasificación son necesarios 16 ejes, número notablemente inferior a los 73 ejes para la clasificación que se utilizaron en el AAC. Se obtienen, al igual que en análisis armónico cualitativo, siete clases de progresión. El número de clases se determina a partir del histograma de índices de nivel de la clasificación jerárquica (figura 4.13)

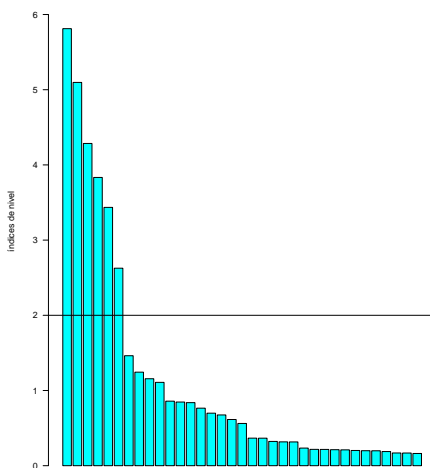


FIGURA 4.13. Histograma de índices de nivel para clasificación en el AAC-AFM. Se consideran 7 clases de progresiones.

En la tabla 4.14 se muestra el número de individuos en cada una de las siete clases de progresión y sus tamaños porcentuales en relación al número total de individuos bajo estudio:

Partición	Clase 1	Clase 2	Clase 3	Clase 4	Clase 5	Clase 6	Clase 7
Tamaño de clase	232	553	761	261	480	107	51
Porcentaje	9.48	22.61	31.12	10.67	19.63	4.38	2.08

TABLA 4.14. Clases de progresiones de la ERC a partir del AAC-AFM

La clase mayoritaria la constituye la clase 3 que representa el 31.12% del total de individuos, le siguen en tamaño las clases 2, 5, 4, 1, 6 y 7 respectivamente, ésta última con 51 individuos (2.08% de la población). Comparada ésta partición con la obtenida a partir del AAC (ver figura 4.7), se aprecian pequeñas diferencias en cuanto a los tamaños de las clases que las determinan (incluso la clase 7 se mantiene). Esta situación permite afirmar que las clases encontradas a través de los dos análisis factoriales tienden a permanecer estables, al menos entre los dos análisis.

Las siete clases de progresión de la ERC son proyectadas conjuntamente en el primer plano factorial del AAC-AFM, plano que se muestra en la figura 4.14. En el plano también se presentan los tamaños y las proporciones relativas de cada una de las clases.

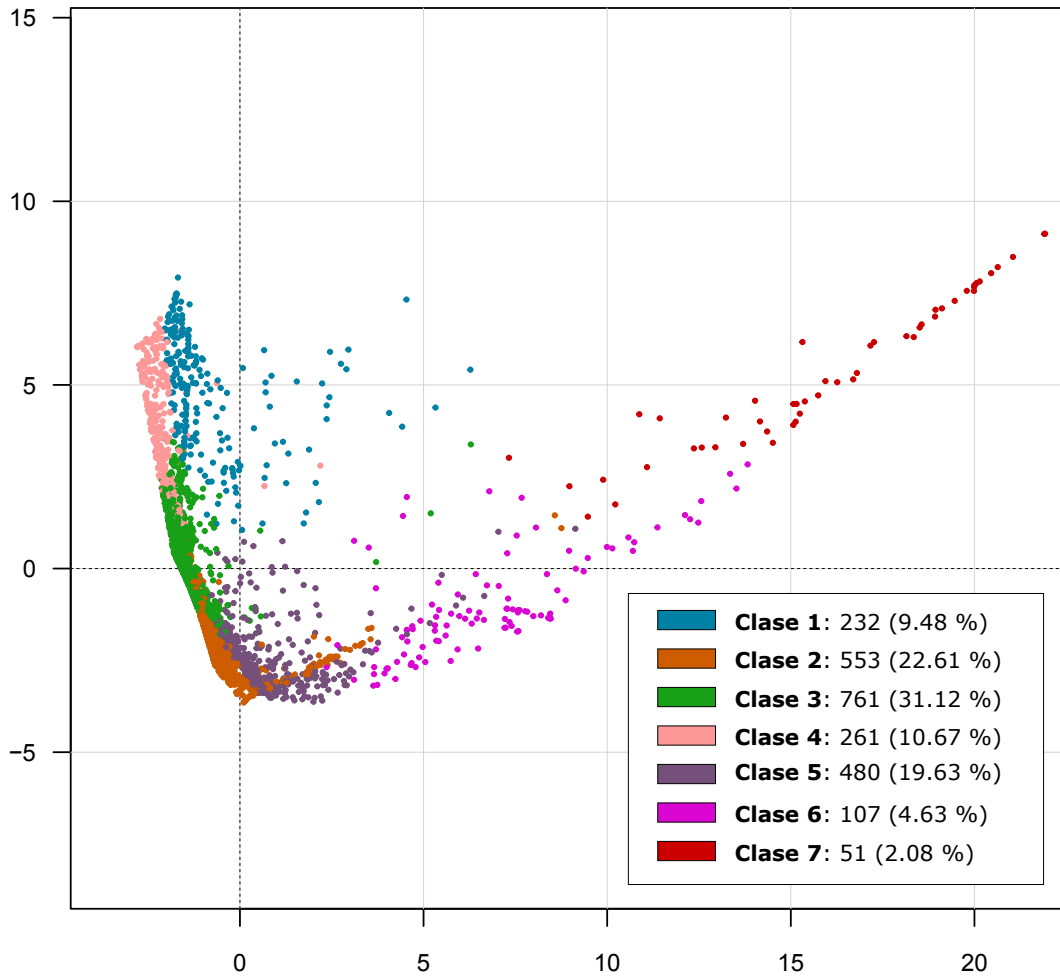


FIGURA 4.14. Proyección de las siete clases de progresión en el primer plano factorial del AAC-AFM.

La caracterización de los dos primeros ejes factoriales del AAC-AFM realizada en la sección 4.5.2, a través del círculo de correlaciones de periodos-estadios (figura 4.11) y el primer plano factorial con los centroides proyectados de las VTI cualitativas (figura 4.10), dan una primera idea de la naturaleza de las clases de progresión de la ERC: la **clase 1** se caracteriza por individuos que son excluidos prematuramente del estudio (CB) por ser remitidos a consulta externa, cambio de aseguradora, pérdida de seguimiento, muerte, remisión a cuidados paliativos o abandono voluntario del estudio, además de padecer cáncer; la **clase 2** la componen principalmente los individuos que sobrevivieron al estudio (CA) y la causa de su ERC es una enfermedad desconocida; la **clase 3** se caracteriza por progresar principalmente a través de *E3a* y *E3b* y no ser diabéticos pero sí hipertensos; la **clase 4** es la clase que progresa mayoritariamente en *E3a*, y posiblemente, sus individuos se cen-

suran del estudio anticipadamente; la **clase 5** corresponde a los individuos que progresan principalmente en $E4$, tienen antecedentes de enfermedad cardiovascular, son diabéticos y la causa de su ERC es otra distinta a la hipertensión y la glomerulonefritis; la **clase 6** está compuesta por sujetos con mayor permanencia en $E5$, muy posiblemente diabéticos y cuya causa de su ERC es precisamente esta enfermedad y la glomerulonefritis; por último, la **clase 7** se caracteriza por individuos remitidos a TRR (diálisis y trasplante), y quizás, diabéticos.

Nuevamente, la descripción realizada en el párrafo anterior es apenas un primer acercamiento al estudio tipológico de las clases encontradas. Para una caracterización más confiable y acertada de las mismas es necesario utilizar distintas medidas y criterios, a través de las cuales es posible determinar qué VTI's son las que, para cada una de las clases, se constituyen en las variables que las explican. Mediante las VTI, en unión con la VLC, es posible construir perfiles de progresión de las clases que facilitan su caracterización. Además, es importante señalar que como complemento a este trabajo con los perfiles, se añada la descripción de las clases a través de las variables longitudinales ilustrativas (VLI).

4.6.1. Trayectorias promedio de las clases de progresión de la ERC

El AFM proporciona un medio útil para el estudio de la evolución en el tiempo de un individuo a través de los estadios de su ERC. Esto se logra proyectando sobre un plano factorial particular, los puntos parciales del individuo en cada uno de los periodos de corte (sección 2.4.2). La unión de los puntos parciales determinan una *trayectoria promedio*, la que a su vez describe su progresión a través de la ERC.

Es posible proyectar sobre el plano factorial los puntos parciales de todos los individuos con el ánimo de trazar sus trayectorias individuales, y a partir de estas, definir clases con trayectorias similares. Sin embargo, debido al gran número de individuos, la identificación de las posibles clases en el plano se hace engorrosa, y si se quiere, imposible.

Dado que las clases de progresión ya están identificadas, se procede a construir individuos artificiales que harán las veces de representantes de cada una de ellas. La construcción de dichos individuos (siete, uno por cada clase) se realiza a partir de las correspondientes subtablas P (proporción de permanencia en los estadios de la ERC por periodo de corte) de cada una de las clases. Las trayectorias a través de los periodos de corte en cada uno

de los estadios o censuras la ERC de cada individuo artificial corresponden simplemente a la marginal columna de cada subtabla. Definidos estos individuos, se introducen al AAC-AFM como individuos ilustrativos, que del mismo modo que las VTI, no tienen peso al interior del análisis. Estando dentro del análisis factorial, lo que queda es proyectar sus puntos parciales sobre el primer plano factorial y unir los mismos para definir sus trayectorias. Las trayectorias promedio de los primeros cuatro clase se presentan en la figura 4.15, mientras que las trayectorias de las tres clases restantes se muestran en la figura 4.16.

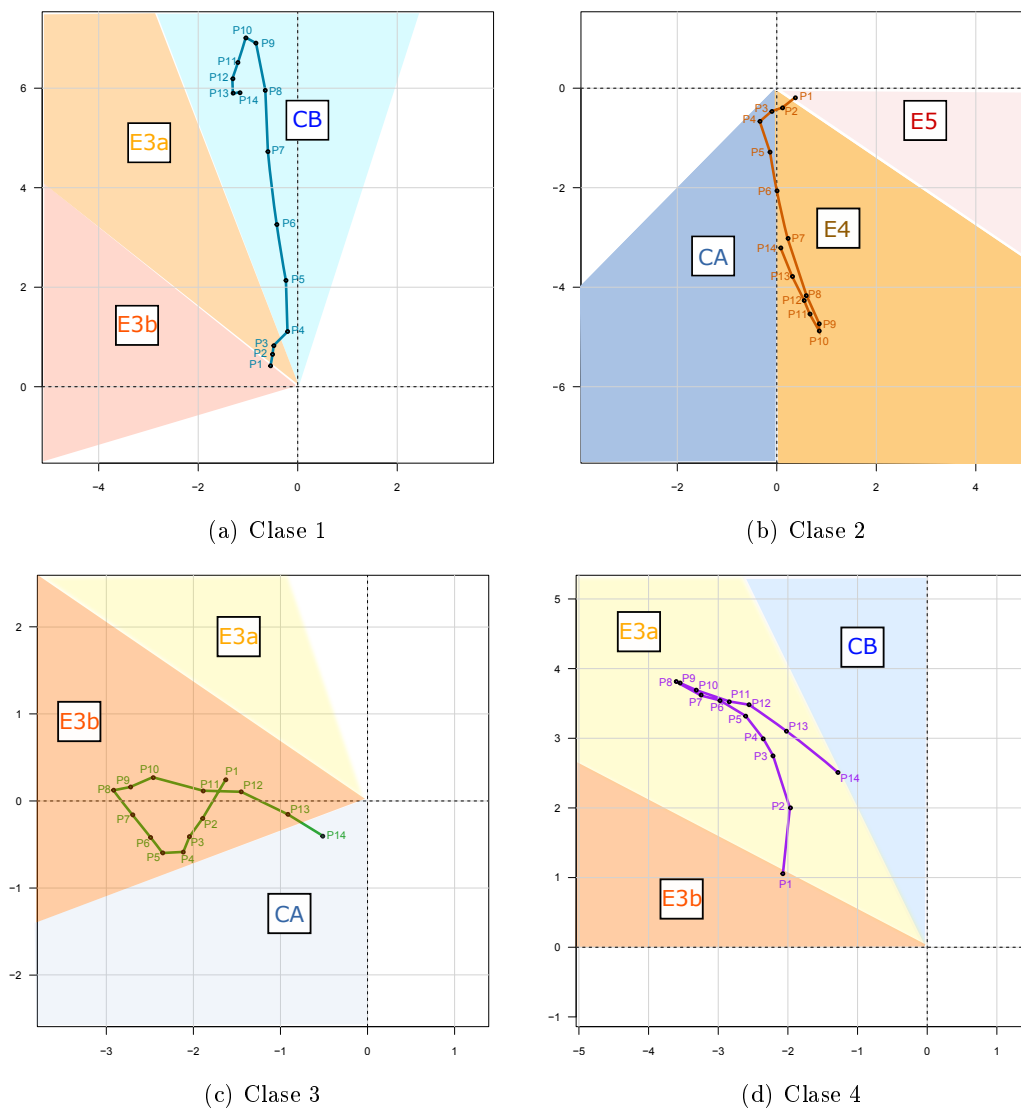


FIGURA 4.15. Proyección en el primer plano factorial del AAC-AFM de las trayectorias promedio de las primeras cuatro clases de progresión de la ERC. Las regiones sombreadas corresponden a las zonas de influencia de estadios y/o censuras

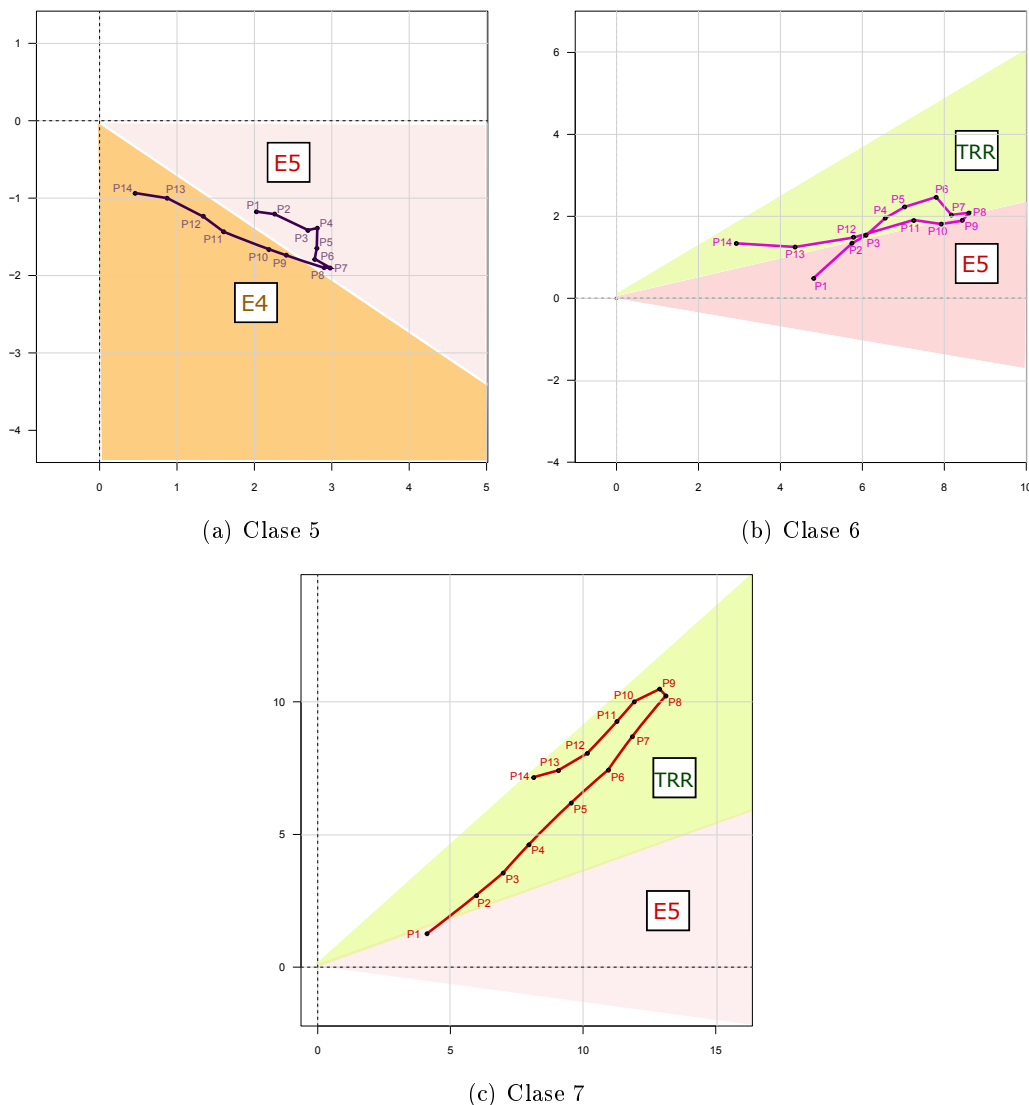


FIGURA 4.16. Proyección en el primer plano factorial del AAC-AFM de las trayectorias promedio de las tres últimas clases de progresión de la ERC. Las regiones sombreadas corresponden a las zonas de influencia de estadios y/o censuras

4.6.1.1. Caracterización de las trayectorias promedio de las clases a partir de la VLC activa

El círculo correlaciones entre los periodos de corte y estadios y los dos primeros ejes del AAC-AFM que se presentaron en la figura 4.11, permite describir las trayectorias promedio de cada una de las clases. Para facilitar la interpretación de los planos se incluyen etiquetas y regiones sombreadas con las zonas de influencia de los estadios y/o las censuras. La caracterización de las trayectorias promedio de cada una de las clases a partir de los puntos parciales de la VLC activa es la siguiente:

- Clase 1:** Del periodo $P1$ al periodo $P3$ permanencias altas en los estadios $E3a$ y $E3b$. A partir de $P4$ hasta $P14$ las permanencias predominan en la categoría de censura CB .
- Clase 2:** Permanencia mayoritaria en estadio $E4$, y en menor proporción en la categoría de censura CA , hasta el periodo $P10$. A partir de $P11$ la tendencia es hacia la categoría de censura CA .
- Clase 3:** Permanencia exclusiva en el estadio $E3b$ a través de los primeros 8 periodos. A partir $P9$ existe una tendencia de progresar hacia la categoría de censura CA , sin dejar de progresar en menor proporción, en $E3b$.
- Clase 4:** Permanencia exclusiva en estadio $E3a$, salvo en el primer periodo que muestra permanencia en $E3b$. A partir de $P8$ se muestra una tendencia marcada de progresar hacia la categoría de censura CB , sin dejar de progresar, en menor proporción, en $E3a$.
- Clase 5:** En la primera mitad del tiempo de estudio (de $P1$ a $P7$) progresión a través del estadio $E5$. A partir de $P8$, y de ahí en adelante, permanencia exclusiva en $E4$.
- Clase 6:** Permanencia alternada a través de los distintos periodos entre el estadio $E5$ y la categoría TRR . En los últimos periodos (a partir de $P11$) la transición es casi exclusiva a TRR .
- Clase 7:** Permanencia casi exclusiva en la categoría TRR a través de cada uno de los periodos de corte, salvo algunas permanencias al principio del estudio en estadio $E5$.

4.6.2. Los perfiles de progresión de las clases de la ERC

Para completar la descripción de las clases de progresión de la ERC, se construyen para cada una ellas perfiles de progresión, que resultan del cruce de la VLC activa (condicionada a cada clase) con los periodos de corte del tiempo de estudio. Los perfiles gráficos de las clases se elaboran mediante la función *plotct{FactoClass}*, también de R. La descripción de cada perfil se complementa con las transiciones entre estadios y censuras de mayor frecuencia por periodos de corte, resumidas en correspondientes diagramas de nivel.

La caracterización final de los perfiles de las clases se realiza a partir de la propia VLC activa, las VTI cualitativas y las VTI cuantitativas, haciendo uso de *valores test* (Lebart et al. 1995). Los valores test son índices descriptivos, cuya ordenación dentro de cada clase, para el caso de las VTI cualitativas, permite determinar aquellas categorías cuya proporción dentro de la clase se diferencian lo suficiente de la proporción global (proporción de la categoría dentro de la población). En el caso de la VLC activa y las VTI cuantitativas, los valores test determinan aquellas variables cuya media es suficientemente diferente de la media de la variable pero a nivel global. Tal vez resulta importante aclarar, que si bien inicialmente la VLC activa se define como una realización de un proceso longitudinal cualitativo, tras las codificaciones realizadas, puede ser tomada como una variable cuantitativa, de ahí que los valores test sobre las que ella actúa no comparen proporciones sino medias (tiempo medio de permanencia en un estadio/censura en un periodo corte determinado).

Las tablas que permiten describir las clases a través de la VLC activa y las VTI cualitativas y cuantitativas, con sus respectivos valores test, se obtienen a través de la función *cluster.carac{FactoClass}* de R (Pardo & Del Campo 2007) y se presentan en el apéndice E, mientras que las tablas con las permanencias y las transiciones entre estadios y/o censuras de las clases, por periodos de corte, se presentan en el apéndice F.

Finalmente, una síntesis operativa de cada clase, donde se resumen las características más relevantes de cada una de ellas, se presenta en el apéndice G.

4.6.2.1. Perfil de progresión de la ERC: clase 1 (9.48%)

Está constituida por individuos que progresaron de $P1$ a $P5$ principalmente en $E3a$ (la permanencia media dentro de la clase en este estadio es mayor a la media global). De $P1$ a $P10$ las permanencias medias en $E3b$ y $E4$ están por debajo de la permanencia media global, lo cual implica una progresión positiva, al menos en relación a la totalidad de la población. Desde $P4$ hasta $P10$ los individuos se censuran del estudio de manera prematura, al punto que después de $P10$ todos los individuos de esta clase están en la categoría de censura CA (ver figura 4.17 y tabla E.1). Las transiciones más frecuentes, por periodos de tiempo, que se tienen para esta clase son: $E3a - E2$ (de $P1$ a $P3$); $E3a - E3b$ (en $P4$); $E2 - CB$, $E3a - CB$ y $E3b - CB$ (de $P5$ a $P10$), lo cual lleva pensar que la progresión de esta clase a través de la ERC es positiva, en el sentido que, la mayoría de transiciones operadas son a estadios de menor gravedad (ver figura 4.18 y tabla F.2).

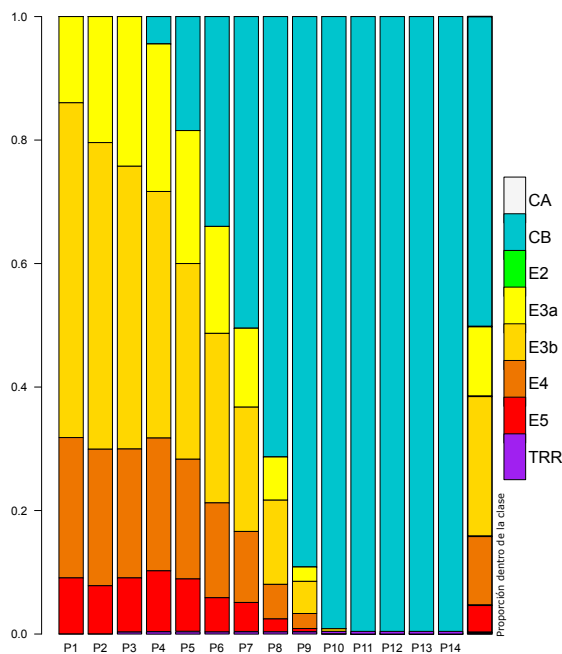


FIGURA 4.17. Perfil de progresión de la ERC de la clase 1 según la VLC

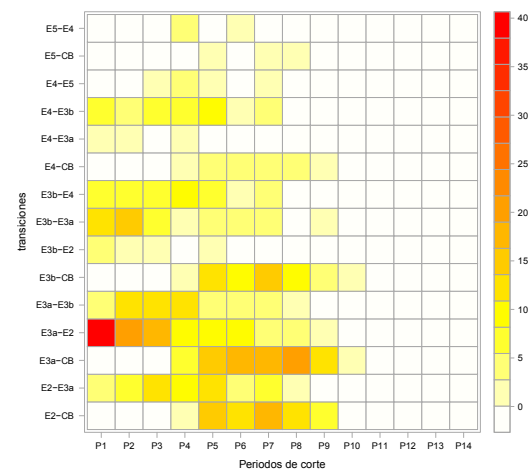


FIGURA 4.18. Transiciones entre estadios y/o censuras por periodos de corte de la clase 1

A nivel clínico los individuos de esta clase se caracterizan por una TFG inicial de 47.36, una TFG media de 49.75 y una TFG final de 50.9, todas ellas por encima de las TFG medias globales respectivas (tabla E.3). Además, están diagnosticados con cáncer

(3% dentro de la clase, proporción baja dentro de la clase pero suficientemente diferente de la proporción de esta categoría a nivel global), no tienen antecedentes de enfermedad cardiovascular (98.3%) y egresaron prematuramente del estudio por las siguientes razones: muerte (13.8%), remisión a consulta externa (38.4%) o cuidados paliativos (2.2%), haber cambiado de aseguradora (15.5%), abandonar el estudio (7.8%) o por pérdida de seguimiento (22.4%) (tabla E.2). Finalmente, el promedio de días aportados por la clase al estudio es de 352.65, número de días por debajo de la media global (tabla E.3), mientras que el nivel educativo alcanzado por sus integrantes es la básica primaria (tabla E.2).

4.6.2.2. Perfil de progresión de la ERC: clase 2 (22.61%)

Esta clase se caracteriza por permanencia mayoritaria en *E4* durante los dos primeros periodos y permanencia en *E3b* en los dos siguientes periodos (las permanencias medias en estos estadios son mayores a la media global). A partir de *P4* los individuos comienzan a ser censurados por haber cumplido el tiempo de estudio (en estos periodos la permanencia media en *CA* está por encima de la media global), de tal modo, que después de *P10*, la gran mayoría de los miembros de esta clase han salido ya del estudio. (ver figura 4.19 y tabla E.4).

Las transiciones más frecuentes, por periodos de corte, para la clase 2 son: *E3b* – *E3a* (de *P1* a *P4*) y *E3a* – *CA* (de *P5* a *P10*) (ver figura 4.20 y tabla F.3), lo que lleva a pensar en una progresión relativamente positiva, dado que se presentan con mayor frecuencia transiciones a estadios de menor gravedad.

En términos clínicos, los integrantes de esta clase presentan TFG inicial de 44.38 (por debajo de la media global) y TFG final de 49.74 (por encima de la media global). La TFG promedio no es determinante en esta clase, pues no se diferencia mucho de la TFG promedio global (tabla E.6). También se caracterizan por ser mujeres (57.1%) que finalizaron la educación secundaria (89.7%), sin cáncer (100% dentro de la clase), con antecedentes de enfermedad cardiovascular (12.7%), cuya causa de su ERC es distinta de la HTA, la diabetes o la glomerulonefritis (27.3%), o es una enfermedad desconocida (3.1%) y concluyeron el periodo de estudio (99.1% de la clase) (tabla E.5).

Finalmente, el promedio de días aportados por la clase al estudio es de 333.38, valor muy por debajo del promedio de días aportados por la población general (tabla E.6).

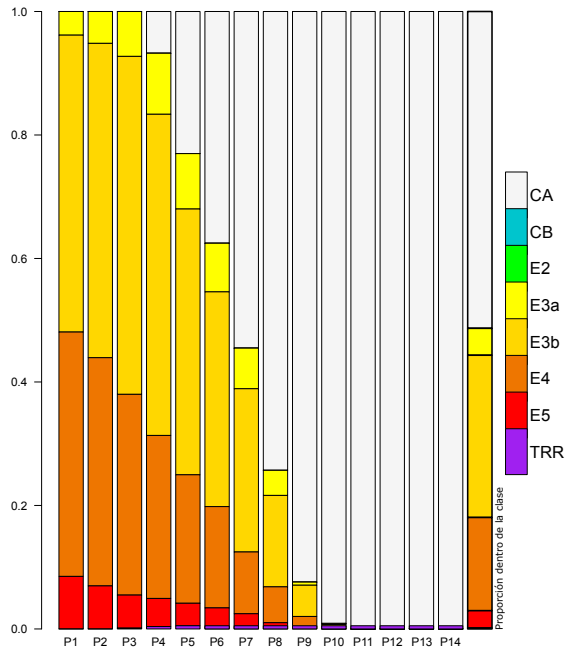


FIGURA 4.19. Perfil de progresión de la ERC de la clase 2 según la VLC

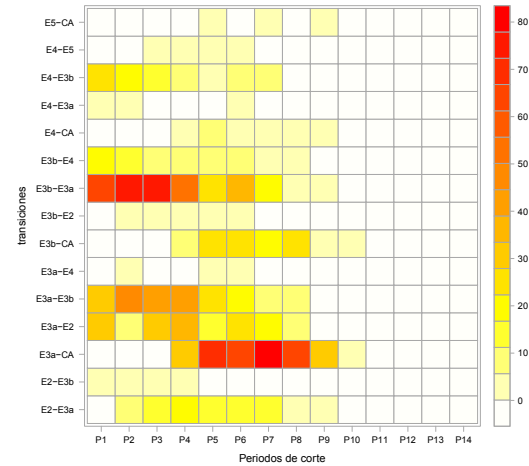


FIGURA 4.20. Transiciones entre estadios y/o censuras por periodos de corte de la clase 2

4.6.2.3. Perfil de progresión de la ERC: clase 3 (31.12%)

Esta clase se caracteriza por una progresión permanente en estadio *E3b* durante todo el estudio (la permanencia media en este estadio está por encima de la media global), en tanto que su permanencia media en el estadio *E4* está por debajo de la media de la población, lo que representa una progresión relativamente favorable en la ERC. A partir de *P11* y *P12* hay censuras por finalización del tiempo de estudio (*CA*) y por egreso anticipado (*CB*) respectivamente, las permanencias en estas categorías están por debajo de las medias globales (ver figura 4.21 y tabla E.7).

Dentro de la clase las transiciones más frecuentes, por periodos de corte, que se reportan son: *E3b – E3a* y *E3a – E3b* (de *P1* a *P11*) y *E3a – CA* (de *P12* a *P14*) (ver figura 4.22 y tabla F.4), lo cual indica una progresión estable a lo largo de su ERC pues se presentan transiciones complementarias con más o menos la misma distribución en el tiempo.

A nivel clínico, los individuos que pertenecen a esta clase se caracterizan por tener una TFG inicial de 50.16, una TFG promedio de 50.87 y una TFG final de 51.4, cada una

de ellas mayores a las respectivas TFG medias globales (tabla E.9). Son individuos sin antecedentes de enfermedad cardiovascular (98.4%) y para quienes la causa principal de su ERC es la tensión arterial alta (64.8%). También se caracterizan porque la mayoría de ellos culminaron el periodo de estudio (77.9%) (tabla E.8).

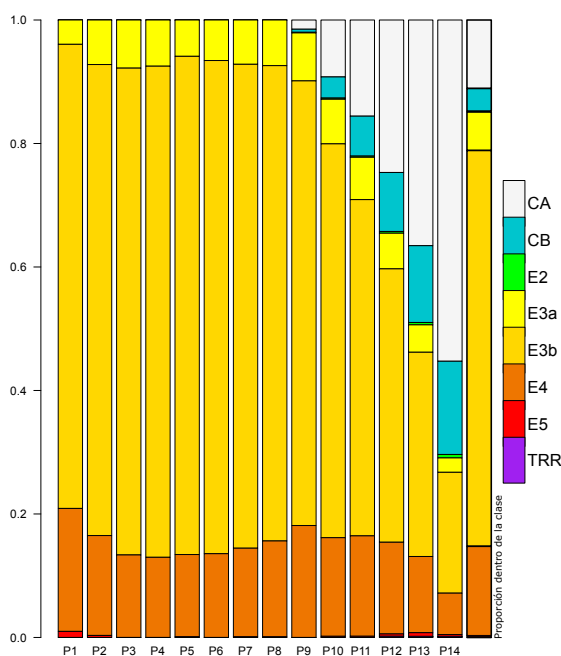


FIGURA 4.21. Perfil de progresión de la ERC de la clase 3 según la VLC

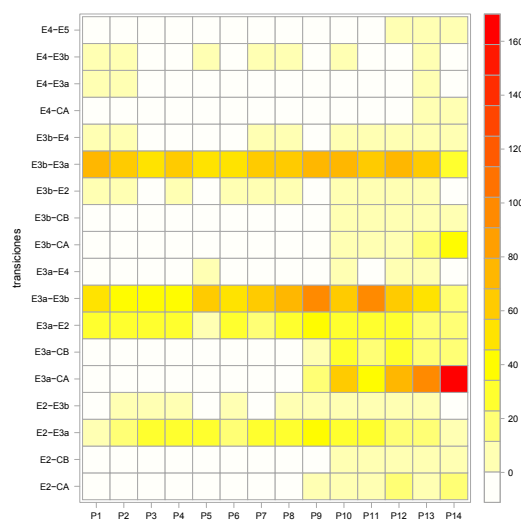


FIGURA 4.22. Transiciones entre estadios y/o censuras por periodos de corte de la clase 3

Por otro lado, el promedio de días aportados al estudio por los individuos de esta clase está muy por encima de la media global (1139.15 días) (tabla E.9), y sus integrantes se caracterizan porque una proporción importante de ellos culminaron la educación básica (26.7%) y otra proporción tienen estudios en educación superior (12.7%) (tabla E.8) .

4.6.2.4. Perfil de progresión de la ERC: clase 4 (10.67%)

Los individuos que componen esta clase progresan mayoritariamente en *E2* y *E3a* a través de todos los periodos de corte del tiempo de estudio, salvo en *P14*, donde la categoría dominante es la de censura *CB* (las permanencias medias en estos estadios es mayor que las medias globales respectivas). A partir de *P11* los individuos de esta clase empiezan a egresar del estudio prematuramente, mientras que la censura por cumplimiento del tiempo

de estudio está por debajo de la permanencia media global en esta categoría, por lo cual esta categoría no caracteriza la clase (ver figura 4.23 y tabla E.10).

Las transiciones más frecuentes, por periodos de corte, que se tienen para esta clase son: $E3a - E2$ (de $P1$ a $P5$); $E3a - E2$ y $E2 - E3a$ (de $P6$ a $P13$); $E3a - CA$ y $E2 - CA$ (en $P14$), lo que lleva a concluir que esta clase tiene una progresión favorable en su ERC pues se reportan con mayor frecuencia transiciones a estadios de menor gravedad en cada uno de los periodos de corte (ver figura 4.24 y tabla F.5).

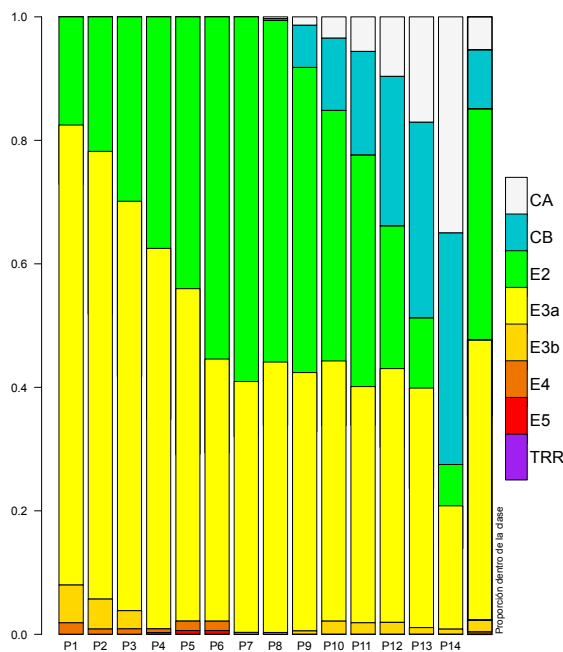


FIGURA 4.23. Perfil de progresión de la ERC de la clase 4 según la VLC

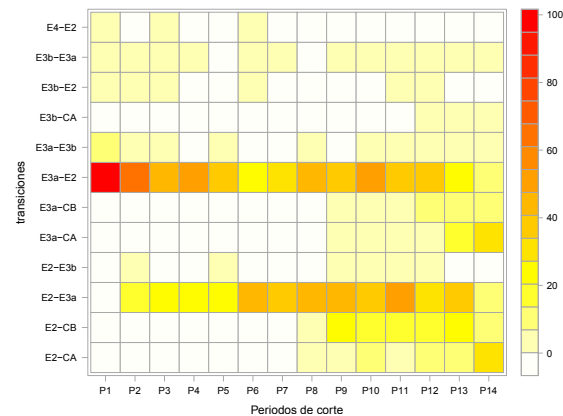


FIGURA 4.24. Transiciones entre estadios y/o censuras por periodos de corte de la clase 4

Por otro lado, esta clase se caracteriza por una TFG inicial de 53.28, una TFG promedio de 60.87 y una TFG final de 62.63, todas ellas por encima de las respectivas medias globales (tabla E.12); la causa principal de su ERC es la tensión arterial alta (70.5 %) y como causas más frecuentes de egreso temprano del programa están la remisión a consulta externa (39.5 %) y el cambio de aseguradora (6.5 %)(tabla E.11).

Esta clase se destaca también por ser en su mayoría hombres (55.2 %), que en algunos casos no terminaron la primaria (9.6 %), en otros alcanzaron el nivel básico (30.7 %) o que tuvieron algún tipo de educación superior (14.9 %)(tabla E.11). Finalmente, el promedio

de días aportados por esta clase al estudio es de 1121 días, promedio por encima de la media aportada por el total de la población (tabla E.12).

4.6.2.5. Perfil de progresión de la ERC: clase 5 (19.63%)

Esta clase progresa mayoritariamente en $E4$ a través de todo el tiempo de estudio (permanencia media en esta categoría por encima de la media global), mientras que la permanencia media en $E3b$ (estadio de menor gravedad que $E4$) está por debajo de la media presentada a nivel de la población, lo que supone una progresión relativamente negativa a través de la ERC. En el último periodo, las categorías con mayor permanencia (cuya media es mayor a la media global) son precisamente $E4$ y la categoría de censura CA . Llama la atención que el perfil de la figura 4.25 evidencia la presencia, a lo largo de todo el estudio, de individuos en estadio $E5$, sin embargo, la permanencia media en este estadio no es lo suficientemente diferente de la permanencia media, pero a nivel global, según valores test (ver tabla E.13)

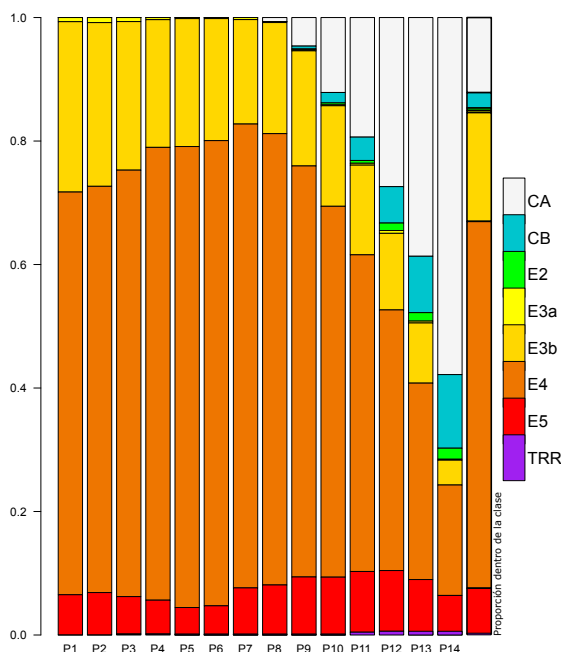


FIGURA 4.25. Perfil de progresión de la ERC de la clase 5 según la VLC

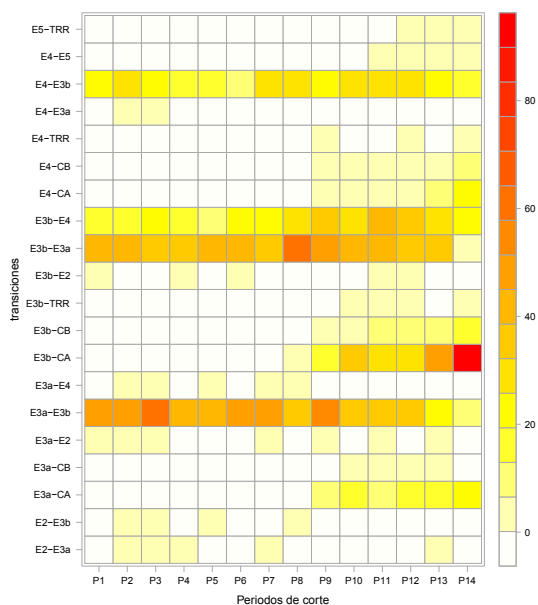


FIGURA 4.26. Transiciones entre estadios y/o censuras por periodos de corte de la clase 5

Aunque a partir de $P10$ y $P13$ respectivamente, los individuos de la clase se censuran por finalizar el tiempo de estudio y por salir del mismo de manera anticipada, la permanencias medias en estas categorías están por debajo de la media global (ver figura 4.25 y tabla E.13).

Para esta clase las transiciones más frecuentes, por periodos de tiempo, que se reportan son: $E3a - E3b$ y $E3b - E3a$ (de $P1$ a $P12$); $E3a - CA$ (de $P13$ a $P14$) (ver figura 4.26 y tabla F.6), lo que implica una progresión estable a lo largo de la ERC debido a que las frecuencias de transiciones complementarias son más o menos las mismas a lo largo de casi todo el periodo de estudio.

Esta clase reporta una TFG inicial de 41.2, una TFG promedio de 38.44 y una TFG final de 38.75, cada una de ellas por debajo de las TFG medias respectivas (tabla E.15). En su mayoría son mujeres (56.2%) con primaria no terminada por un lado (9.4%), y con educación básica culminada (26.9%) por otro; sin antecedentes de enfermedad cardiovascular (99.2%), pero diabéticas (30.2%), cuya causa de su ERC es precisamente esta enfermedad (22.1%). La mayoría de ellas terminaron el periodo de estudio (80.8%), aunque también hay una importante proporción de ellas que egresaron del mismo por haber fallecido (5%) (tabla E.14). El número de días aportados por la clase en promedio es de 1132 días, valor por encima de la media global (tabla E.15).

4.6.2.6. Perfil de progresión de la ERC: clase 6 (4.38%)

Los individuos de esta clase presentan permanencias predominantes en estadio $E5$ en todo el tiempo de estudio, salvo en $P1$ donde la permanencia media predominante está en $E4$ (las permanencias medias en estos dos estadios están por encima de la media global). De $P10$ a $P12$ los individuos de esta clase se caracterizan por su remisión a terapia de reemplazo renal (sólo a partir de $P10$ la permanencia media en este estadio es considerablemente mayor que la media global de permanencia). Si bien, a partir de $P9$ empiezan a censurarse los sujetos por culminar su periodo de estudio o por salir anticipadamente del mismo (por razones distintas a la TRR), las permanencias medias en estas categorías están por debajo de la media global (ver figura 4.27 y tabla E.16).

Las transiciones más frecuentes, por periodos de tiempo, que se tienen para esta clase son: $E4 - E3b$ y $E3b - E4$ (de $P1$ a $P12$); $E3a - CA$ (de $P13$ a $P14$), lo cual indica que la

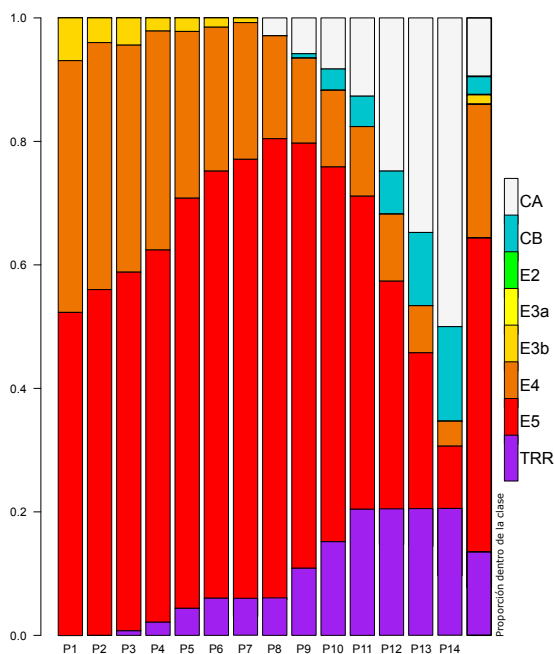


FIGURA 4.27. Perfil de progresión de la ERC de la clase 6 según la VLC

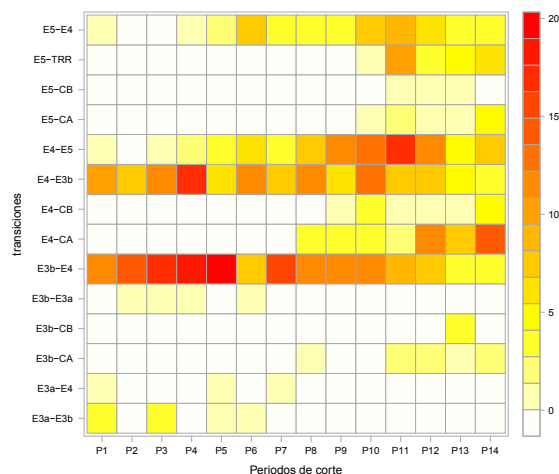


FIGURA 4.28. Transiciones entre estadios y/o censuras por periodos de corte de la clase 6

progresión de esta clase a través de la ERC es estable (transiciones complementarias con más o menos la misma frecuencia en la mayoría de periodos de corte), pero de riesgo alto, dada la gravedad de los estadios sobre los que varía (ver figura 4.28 y tabla F.7).

Las TFG inicial, promedio y final reportadas para esta clase (31.17, 23.97 y 19.21 respectivamente) están todas por debajo de las respectivas TFG medias globales (tabla E.18). Son individuos diabéticos (42.1%), cuya causa de la ERC es precisamente esta enfermedad (35.5%), que se caracterizan por tener 65.3 años en promedio, edad por debajo de la media de la población. La causa principal de egreso del periodo de estudio de estos individuos es la inclusión a diálisis (24.3%) (tabla E.17). El número de días aportados al estudio por la clase no se diferencia de la media global, por tanto este aspecto no es representativo del mismo.

4.6.2.7. Perfil de progresión de la ERC: clase 7 (2.08%)

Esta clase corresponde a los individuos que progresan de manera mayoritaria en *E5* de *P1* a *P9*, mientras que de *P2* hasta *P14*, los integrantes de esta clase se caracterizan

por ser incluidos en terapia de reemplazo renal (*TRR*), a tal punto, que después de *P9* la totalidad de ellos se reportan en esta categoría. Otras permanencias medias en *E3b* y *E4* (en *P1*, *P2*, *P3*, *P4* y *P6*) están por debajo de la media global (ver figura 4.29 y tabla E.19). En esta clase las transiciones más frecuentes, por periodos de corte, que se reportan son: *E3a* – *E3b* y *E3a* – *E3b* (en *P1*); *E4* – *E5* (de *P2* a *P4*); *E5* – *TRR* (de *P5* a *P9*), situación que permite concluir para esta clase una progresión negativa y de altísimo riesgo a través de la ERC (ver figura 4.30 y tabla F.8).

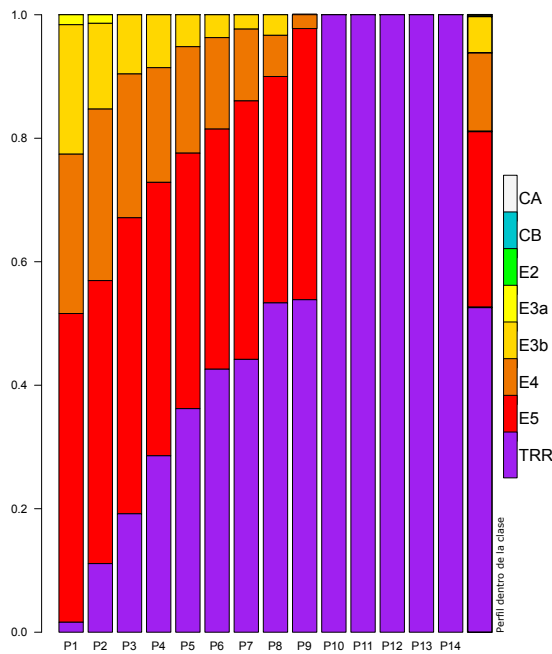


FIGURA 4.29. Perfil de progresión de la ERC de la clase 7 según la VLC

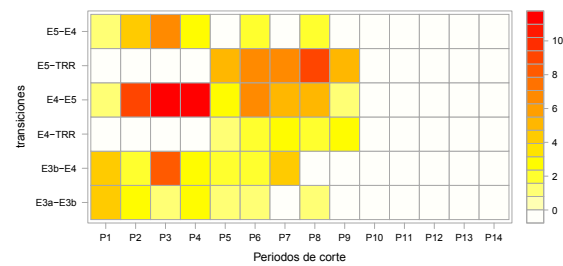


FIGURA 4.30. Transiciones entre estadios y/o censuras por periodos de corte de la clase 7

Las TFG inicial (31.82), promedio (23.36) y final (15.36) reportadas para esta clase están todas notablemente por debajo de las TFG medias globales respectivas. Los integrantes de esta clase tienen una edad promedio de 58.3 años, edad por debajo de la edad media de la población observada (tabla E.21). Son individuos diabéticos (49%), cuya causa de la ERC es precisamente esta última enfermedad (41.2%) y la glomerulonefritis (13.7%). La mayoría de ellos entraron a diálisis (92.2%), mientras que otra proporción (más pequeña) a trasplante (3.9%) (tabla E.20). Los días aportados al periodo de estudio por esta clase es en promedio de 413.86 días, muy por debajo del promedio de días aportados por la población general (tabla E.21).

4.6.3. Comparación de las clases de la ERC mediante las VTI

En esta sección se comparan descriptivamente las siete clases de progresión de la ERC identificadas y caracterizadas en las secciones inmediatamente anteriores. Primero se comparan las clases según el modo como las distintas VTI cuantitativas se distribuyen a través de cada una de ellas; para realizar esto se utilizan boxplots. La segunda comparación, mediante perfiles, se realiza describiendo la manera como las clases se distribuyen a través de las distintas categorías de las VTI cualitativas.

4.6.3.1. Distribución de las TFG inicial, promedio y final en las clases de la ERC

Las figuras 4.31, 4.32 y 4.33 presentan boxplots que permiten comparar las distribuciones de las TFG iniciales, promedios y finales a través de cada una de las clases de progresión de la ERC. Adicionalmente, sobre cada uno de los boxplots, se ubican puntos que representan cada uno de los individuos, en cada una de las clases, cuyas posiciones indican aproximadamente el valor de las tasas que registraron en cada uno de los tres momentos.

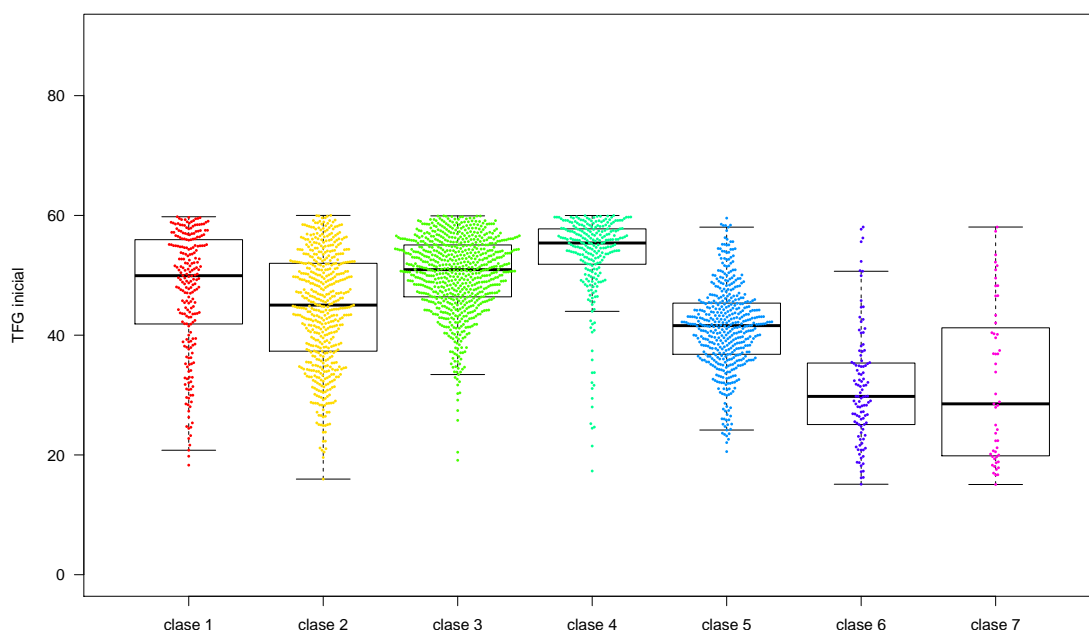


FIGURA 4.31. Distribución de las TFG iniciales para las siete clases de progresión de la ERC

En relación a las tasas de filtrado iniciales (figura 4.31), la clase 4 es la que ingresa al estudio con tasas mayores (mediana de TFG inicial aproximada a 55), mientras que la clase 7 (que es la clase en mayor riesgo de todas las clases) presenta las menores tasas de ingreso, con una mediana menor a 30. Si se ordenan las otras clases según sus tasas de ingreso, de mayor a menor, se tiene en primer lugar las clases 1 y 3 (medianas de TFG de 50), seguidas de la clase 2 y de la clase 5 con medianas de TFG de 45 y 41 respectivamente, y finalmente, la clase 6 con una mediana ligeramente menor a 30. Por otro lado, la clase 2 presenta tasas de ingreso, que si no es por el corte en $TFG = 60$ (debido al criterio de inclusión que permite el ingreso a estudio sólo de individuos con TFG menor o igual a 60), con una distribución aproximadamente normal; lo mismo puede afirmarse para las clases 3 y 5. La clase 4 presenta un buen número de individuos con tasas iniciales menores a 40, que son considerados atípicos. Por su parte, la clase 7, además de presentar las tasas de ingreso más bajas, también es la que presenta la mayor variabilidad, con un claro sesgo positivo (alta frecuencia de tasas de ingreso menores a 30).

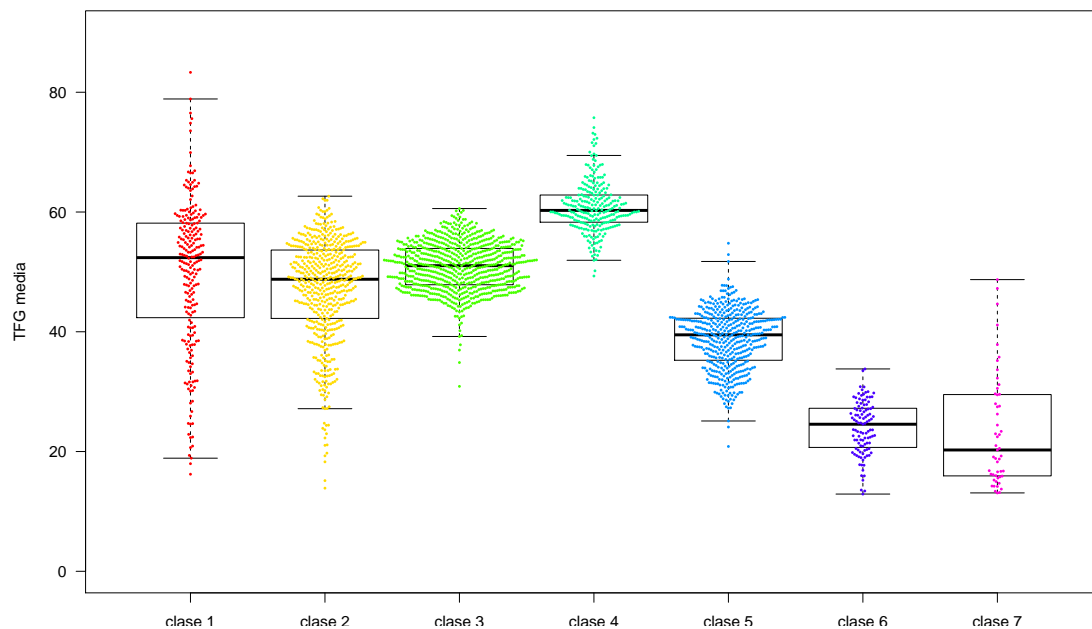


FIGURA 4.32. Distribución de las TFG promedios para las siete clases de progresión de la ERC

En cuanto a las TFG promedio de las clases (figura 4.32), se puede apreciar la misma ordenación, en términos de sus medianas, observada en las tasas iniciales, sin embargo, las clases 5, 6 y 7 muestran un descenso leve en sus tasas promedio (comparadas con las medianas de las tasas de ingreso), mientras que las primeras cuatro clases se mantuvieron

estables. De otro lado, las clases 3 y 4 dejan de presentar alargamientos hacia abajo (como se observaba en las tasas iniciales) lo cual evidencia una mejora general en estas clases, debido a que individuos que ingresaron con tasas iniciales bajas registraron posteriormente tasas promedios más altas. El caso contrario lo muestra la clase 6, cuyo alargamiento hacia arriba desaparece en las tasas promedio, esto significa que los pocos individuos con tasas iniciales altas, después de cierto tiempo dentro del estudio, empezaron a mostrar TFG bajas. El caso más preocupante lo presenta la clase 7, pues la mayoría de sus individuos presentan tasas promedios inferiores a 30, aunque una buena cantidad de ellos tenían tasas de ingreso relativamente altas.

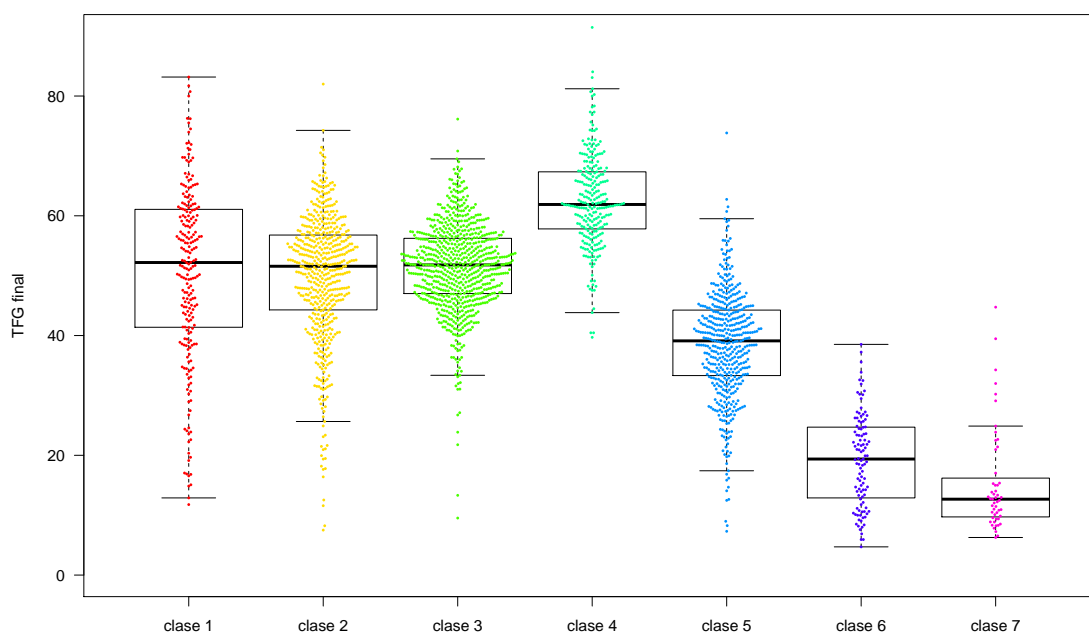


FIGURA 4.33. Distribución de las TFG finales para las siete clases de progresión de la ERC

Por último, la configuración de los boxplots de las TFG finales (TFG reportada en la última consulta) difiere muy poco de los boxplots que representan las distribuciones de las tasas promedio en cada una de las clases. Sin embargo, las clases 6 y 7 finalizan el estudio con tasas menores a las tasas con las que ingresaron y a las tasas con que progresaron en promedio, a tal punto, que la clase 6 finaliza con una TFG cercana a 20, mientras que la clase 7, lo hace con una tasa próxima a 15; las otras clases, en contraste, permanecen estables, en el sentido que: las medianas de sus correspondientes TFG finales son muy cercanas a las medianas iniciales y promedios registradas y, que si se observan

sus distribuciones, tienden a aproximarse a la normalidad, lo que supone un control sobre ellas.

4.6.3.2. Distribución de la edad y los días aportados al estudio en las clases de la ERC

En relación a las edades (ver figura 4.34), se observa que las clases con individuos más jóvenes la presentan las clases 6 y 7 (con medianas aproximadas a los 65 y 55 años respectivamente), y que debido al tipo de progreso en la ERC observadas en estas clases, pueden llevar a pensar que la edad temprana es un factor de riesgo para desarrollar ERC terminal. Las medianas de las restantes clases son aproximadamente iguales, rondando los 70 años, y para cada una de ellas, se observa la presencia de individuos atípicos con edades menores de 45 años.

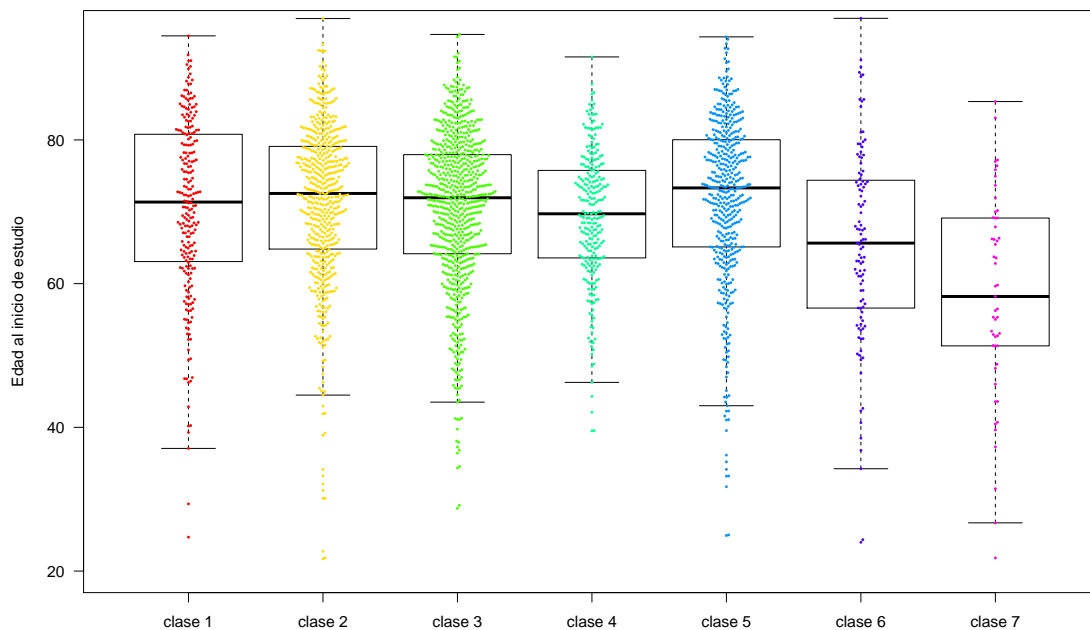


FIGURA 4.34. Distribución de las edades para las siete clases de progresión de la ERC

Finalmente, en la figura 4.35 la distribución de los días aportados al estudio de cada una de las clases, ratifica la caracterización que se hizo de cada una de ellas con respecto a esta variable: las clases 1, 2 y 7 con la menor cantidad de días en estudio (medianas inferiores a 500 días), mientras que las clases 3 a 6 reportaron medianas mayores a los 1000 días.

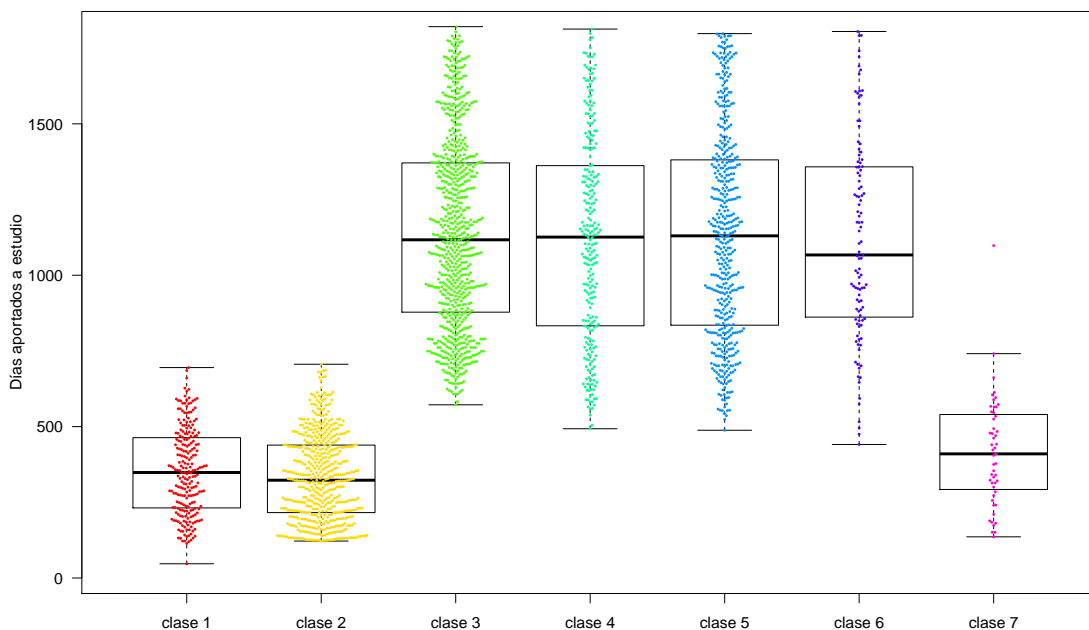


FIGURA 4.35. Distribución de las edades para las siete clases de progresión de la ERC

4.6.3.3. Distribución de las categorías de las VTI cualitativas en las clases de la ERC

Para finalizar la comparación descriptiva de las clases a través de las VTI, se comentan los gráficos de perfiles que describen la manera como las distintas categorías de las VTI cualitativas se distribuyen a través de estas clases.

El perfil por género (figura 4.36) muestra que hombres y mujeres se distribuyen a lo largo de cada una de las clases más o menos en la misma proporción, y en general, todos ellos son muy similares al perfil global. La clase con mayor cantidad de hombres fue la clase 7, aunque con distribución muy similar al de las clases 4 y 6. Por último, las clases con mayor proporción de mujeres son las clases 2 y 5.

El perfil por diagnóstico de cáncer (figura 4.38) pone en evidencia la baja proporción de individuos con cáncer dentro del estudio. La mayor cantidad de individuos con esta enfermedad (aunque muy baja en relación a la categoría complementaria) la presentan las clase 7 y 1. En menor proporción también se presentan individuos con esta condición en las clases 4 y 5 (con distribución similar) y la clase 3. Para concluir, es posible ver que en las clases 2 y 6 ningún individuo padece esta enfermedad.

Por el nivel de escolaridad reportado (figura 4.37), se puede apreciar que, salvo la clase 2, todas las clases muestran un perfil muy similar al perfil global: mayoría de individuos con educación secundaria culminada, seguida de una apreciable proporción de individuos con educación básica y con educación superior, un bajo porcentaje de personas que no culminaron la primaria, y en menor proporción que todas, no reportaron ningún tipo de formación. En cuanto a la clase 2, ésta se puede catalogar como una clase donde casi todos sus miembros (al menos el 95 %) terminaron la educación secundaria.

En cuanto a si se reportó, sí o no, antecedentes de enfermedades cardiovasculares, el perfil de la figura 4.39 evidencia que la mayoría de individuos de la población no mostraban este antecedente. La clase 2 describe la mayor cantidad de individuos con antecedentes cardiovasculares, seguida de la clase 7. En mucho menor proporción las demás clases también tuvieron dentro de sus miembros personas con estos antecedentes, pero que comparadas con el perfil global, resultan todavía escasos.

Según la causa de egreso de estudio (figura 4.40), la clase 1 presenta como mayor causa de egreso la consulta externa, seguida de los individuos que se perdieron o abandonaron el estudio, de los individuos que cambiaron de aseguradora o fallecieron. La clase 2 muestra que todos sus individuos fueron censurados por finalizar el estudio. Por otro lado, la clase 4 muestra que menos de la mitad de sus miembros finalizan el estudio, pero también egresan por consulta externa. Finalmente, la clase 7 tuvo como mayor causa de egreso de estudio la remisión a diálisis.

Con relación al diagnóstico de diabetes mellitus, el perfil de la figura 4.42 presenta a nivel global una baja proporción de diabéticos (aproximadamente el 22 % de la población), lo mismo que en cada una de las clases, sin embargo, se puede apreciar que las clases 5 a 7 son las que muestran mayor cantidad de individuos con esta enfermedad, tanto así, que la clase 7, muestra un perfil con la misma proporción de diabéticos y no diabéticos. Las otras clases, presentan perfiles muy similares al perfil global.

De otro lado, el perfil que describe las clases según las causas de la ERC (ver figura 4.41) permite observar que la causa más frecuente de la ERC, a nivel global y dentro de las clases, es la tensión arterial alta. Otras causas también reportadas en menor proporción, pero comunes a todas las clases, son la diabetes y otras enfermedades (distintas a la HTA, la diabetes y la glomerulonefritis). El perfil de la clase 3 es el más parecido al perfil global, mientras que los perfiles de las clases 6 y 7 son los que más divergen. En estas últimas

clases se presenta un numero importante de individuos cuya causa de la ERC es la diabetes y la glomerulonefritis, que los diferencia de las demás clases.

Por último, el perfil que describe el diagnóstico de tensión arterial alta (HTA) se presenta en la figura 4.43. Es evidente la importante proporción de hipertensos dentro de cada una de las clases y a nivel población. Si bien, todos los perfiles son muy similares entre sí, se destaca que la clase 6 es la que presenta la mayor cantidad de no hipertensos, mientras que en la clase 4, ocurre lo contrario.

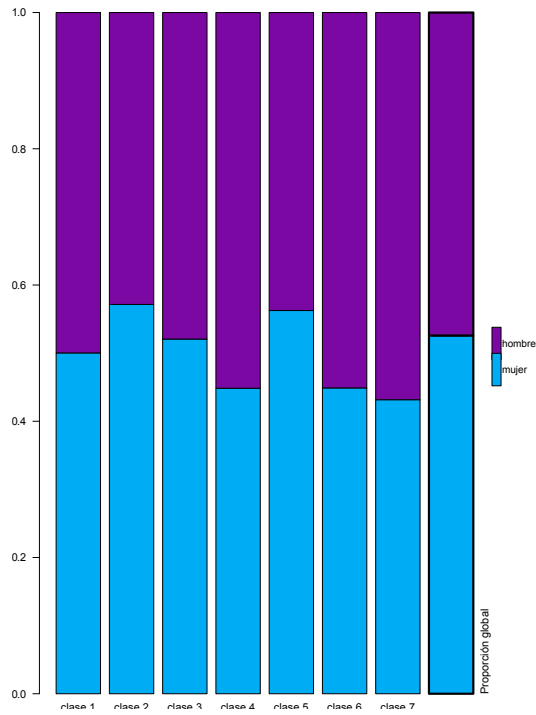


FIGURA 4.36. Perfil de clases de progresión según género

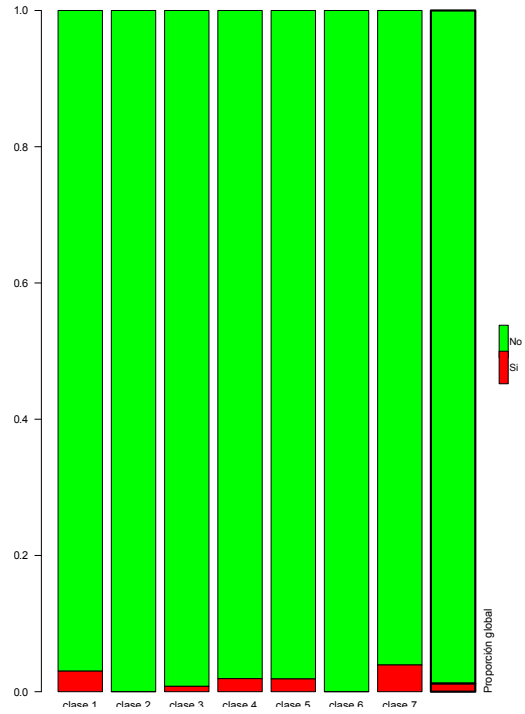


FIGURA 4.38. Perfil de clases de progresión según diagnóstico de cáncer

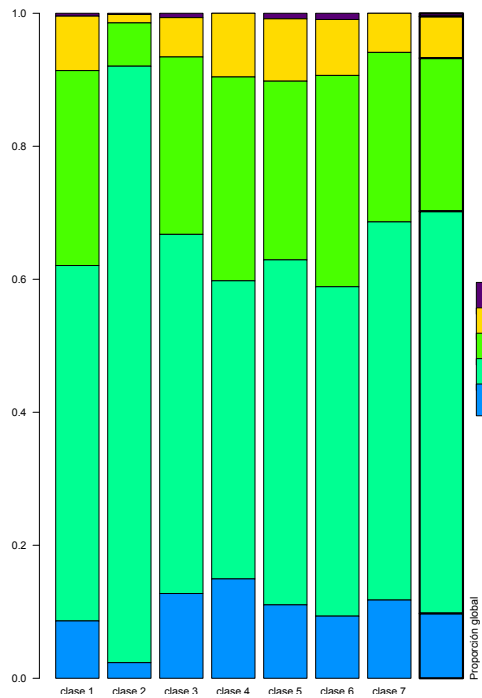


FIGURA 4.37. Perfil de clases de progresión según nivel de escolaridad

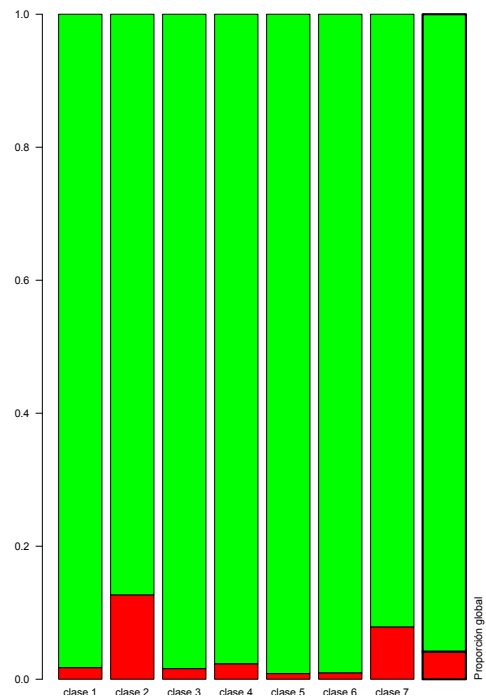


FIGURA 4.39. Perfil de clases de progresión según antecedente de enfermedad cardiovascular

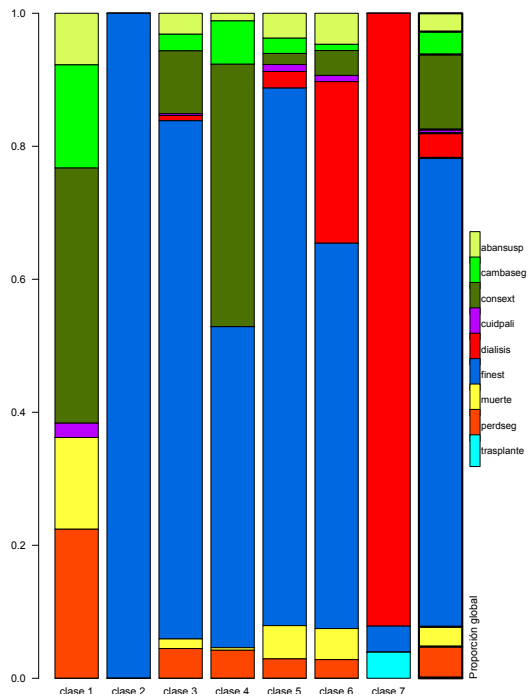


FIGURA 4.40. Perfil de clases de progresión según causa de egreso del estudio

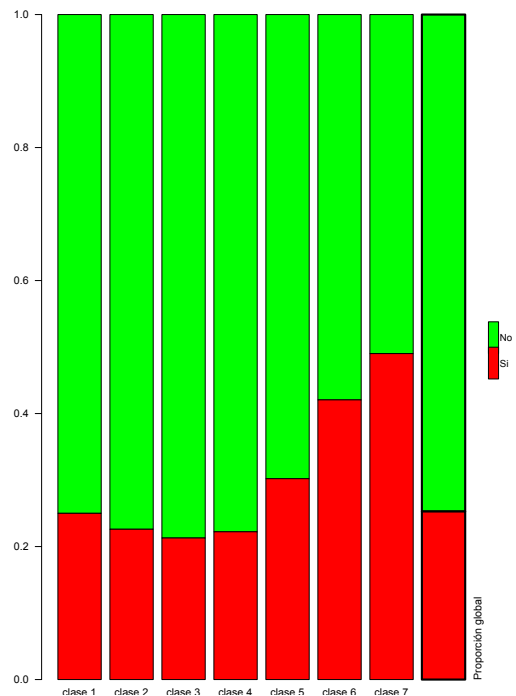


FIGURA 4.42. Perfil de clases de progresión según diagnóstico de diabetes mellitus

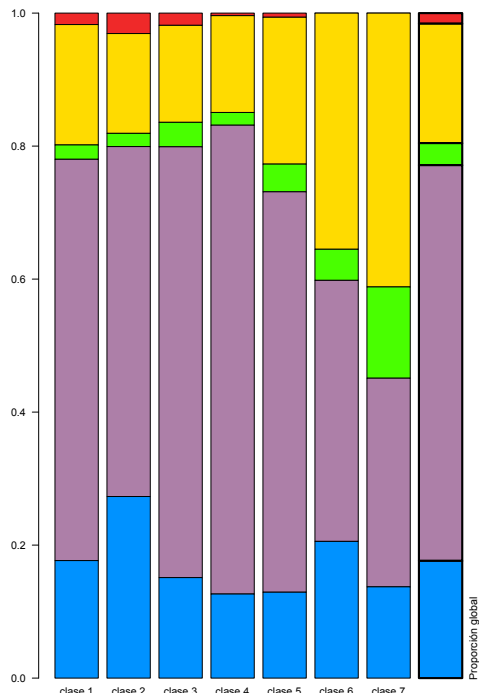


FIGURA 4.41. Perfil de clases de progresión según causa de la ERC

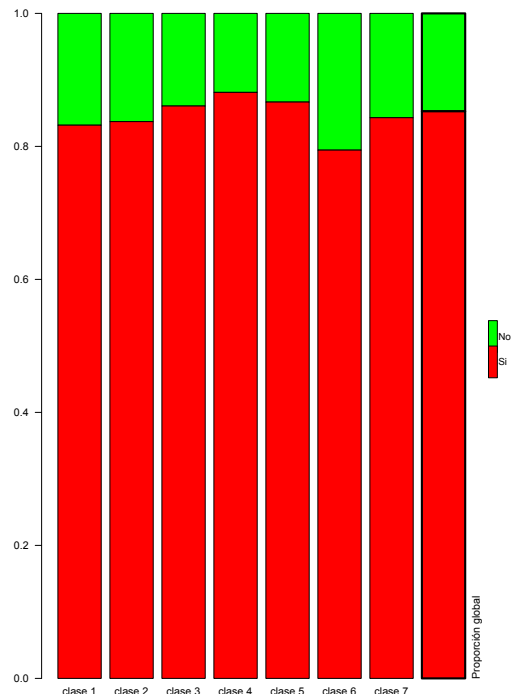


FIGURA 4.43. Perfil de clases de progresión según diagnóstico de HTA

4.6.4. Clases de progresión de la ERC a través de trayectorias de las VLI

Tal como se indicó en la sección 4.2.2, las *variables longitudinales ilustrativas* (VLI), en oposición a las VTI, están constituidas por distintas observaciones registradas en un mismo individuo en distintos instantes del periodo de estudio, es decir, son dependientes del tiempo. Para el caso particular, estas medidas repetidas (denominadas así para referirse a que se realizan sobre un mismo individuo) varían en número (dependiendo de la cantidad de consultas a la que asistió el individuo) y son asincrónicas (intervalos de tiempo donde se realizaron las observaciones varían de individuo a individuo). Se consideran VLI: la *tasa de filtrado glomerular estimada*¹ (TFG), el *nitrógeno ureico en la sangre* (bun), la *creatinina sérica estimada* (creatin), la *hemoglobina* (hgb), la *tensión arterial sistólica* (tasis), la *tensión arterial diastólica* (tadius) y el *índice de masa corporal* (imc).

La progresión longitudinal de las VLI describen trayectorias, que del mismo modo que la TFG, pueden dar indicios del tipo de progresión que tiene un individuo en su ERC. Por ejemplo, trayectorias del nitrógeno ureico, la creatinina, la tensión sistólica, la tensión diastólica y el índice de masa corporal con pendientes positivas, es decir, con un crecimiento sostenido, pueden indicar una pérdida progresiva de la función renal, mientras que una trayectoria también creciente, pero para el caso de la hemoglobina, puede ser indicio de una evolución favorable de la ERC. La relación de estas variables con la función renal, y por ende, con la progresión de esta enfermedad, se explicó en la sección 1.4.

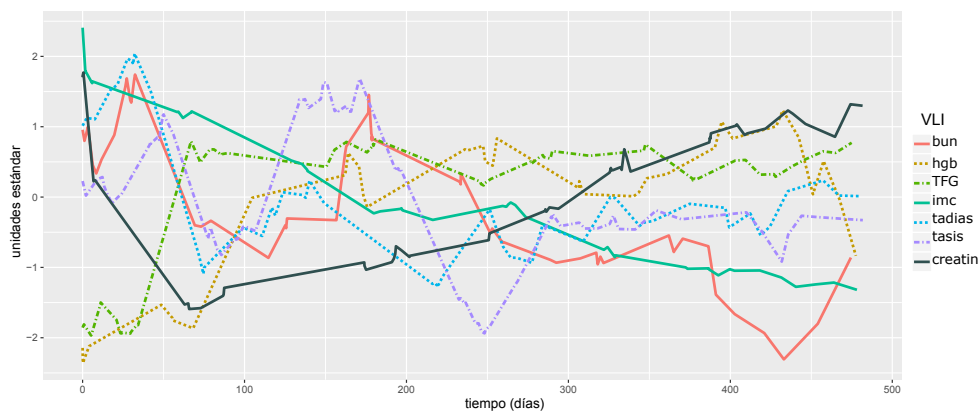


FIGURA 4.44. Trayectorias medias crudas de las VLI de la clase 1

¹Notada simplemente TFG para diferenciarla de las TFG inicial, promedio y final que son de naturaleza transversal

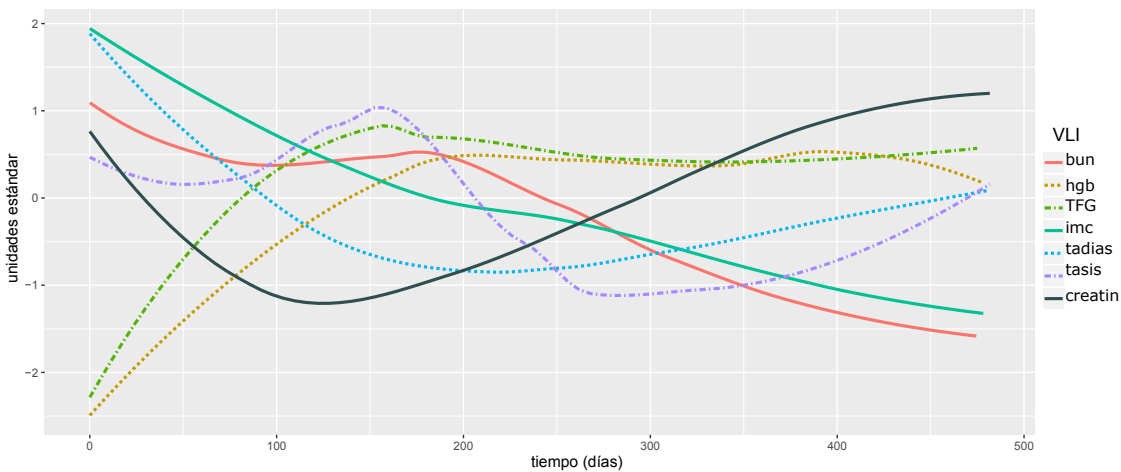
Al promediar las trayectorias individuales (cuyos datos faltantes fueron imputados utilizando las medias de dichas trayectorias) en cada una de las VLI, se obtienen trayectorias medias correspondientes, que van a permitir caracterizar y comparar cada una de las clases de progresión de la ERC a las que cada uno de dichos individuos pertenecen. Las trayectorias medias son calculadas mediante el paquete *kmlShape* (Genolini et al. 2016) de *R*. Por ejemplo, en la figura 4.46 presentan las trayectorias medias de cada una de las VLI para la clase 1.

Las trayectorias medias de las VLI de la clase 1 son estandarizadas (debido a que cada una está expresada en distintas unidades de medida) para poder facilitar su comparación. Es posible observar entonces, que por ejemplo, para esta clase la trayectoria del *imc* decrece de una manera permanente en el tiempo, mientras que la *TFG* crece aceleradamente de los 40 a los 60 días, para luego mostrar una tendencia estable hasta final del tiempo de estudio. Sin embargo, trayectorias “crudas” de variables como el *bun* y la *tasis* presentan saltos muy pronunciados que dificultan observar alguna tendencia global en dichas trayectorias. Esta alta variabilidad (denominada *variabilidad biológica inherente al individuo*), sumada a la ausencia de algún ciclo rítmico o tendencia marcada, es típica de realizaciones de procesos biológicos, o combinación de los mismos, que operan dentro individuos se observan en el tiempo (Fitzmaurice et al. 2012).

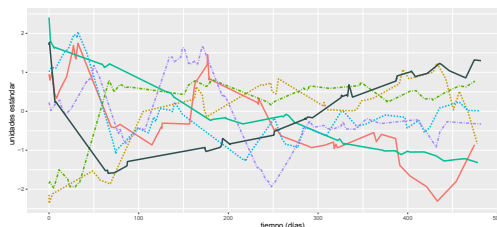
Para mejorar en alguna medida la visualización de las trayectorias, y poder observar tendencias en ellas con mayor facilidad, los datos se ajustan mediante una *regresión polinómica ponderada local* (*loess*) ((Cleveland, 1996) en Diggle (2002, pág. 44)). En este tipo de regresiones el ajuste es local, es decir, el ajuste de un punto en el instante t se realiza, vía mínimos cuadrados ponderados, utilizando puntos en la vecindad de dicho instante. El resultado es una curva ajustada suavizada donde los polinomios que las describen localmente son generalmente de grado dos. Adicional a este suavizado, y con fines meramente exploratorios y descriptivos, los datos se ajustan mediante *modelos lineales generalizados* (*glm*) (McCullagh 1984) que obligan a las trayectorias a describir un comportamiento lineal a través del tiempo, permitiendo observar para cada clase una especie de perfil que las caracteriza una a una. Las trayectorias medias suavizadas por *loess* y por *glm* de las VLI de cada clase son realizadas a través de la función `geom_smooth{ggplot2}` de *R* mediante las opciones `loess` y `glm` respectivamente.

4.6.4.1. Trayectorias medias de las VLI: clase 1

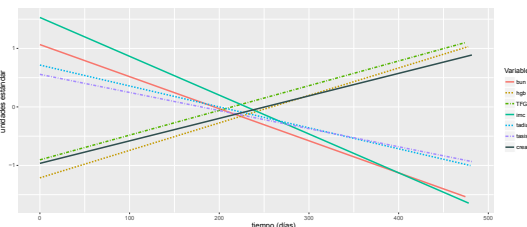
Las trayectorias suavizadas por glm de las VLI de la clase 1 (figura 4.45(c)) inducen a pensar en una progresión positiva general en su ERC, ya que la trayectorias de la tasa de filtrado glomerular, la hemoglobina y la creatinina (con menor pendiente que la TFG) crecen positivamente, mientras que las trayectorias del nitrógeno ureico, la tensión diastólica, el índice de masa corporal y la tensión sistólica decrecen en el tiempo. Al realizar un análisis más detallado sobre las trayectorias loess y crudas (figuras 4.45(a) y 4.45(b) respectivamente), se puede apreciar que efectivamente el bun, el imc, y la tadius (con menor pendiente que las anteriores) decrecen en el tiempo, mientras que la TFG y la hgb crecen rápidamente durante los primeros 150 días, para luego estabilizarse hasta finalizar el tiempo de observación. El comportamiento de la creatinina obedece a la relación recíproca que tiene con la TFG, es por tal razón que mientras la TFG crece en un intervalo específico, la creatinina decrece (más o menos en la misma proporción) en dicho intervalo. En cuanto a la tensión sistólica, ésta presenta la mayor variación en el tiempo de todas las VTI y su tendencia no es concluyente.



(a) Suavizadas por loess



(b) Crudas

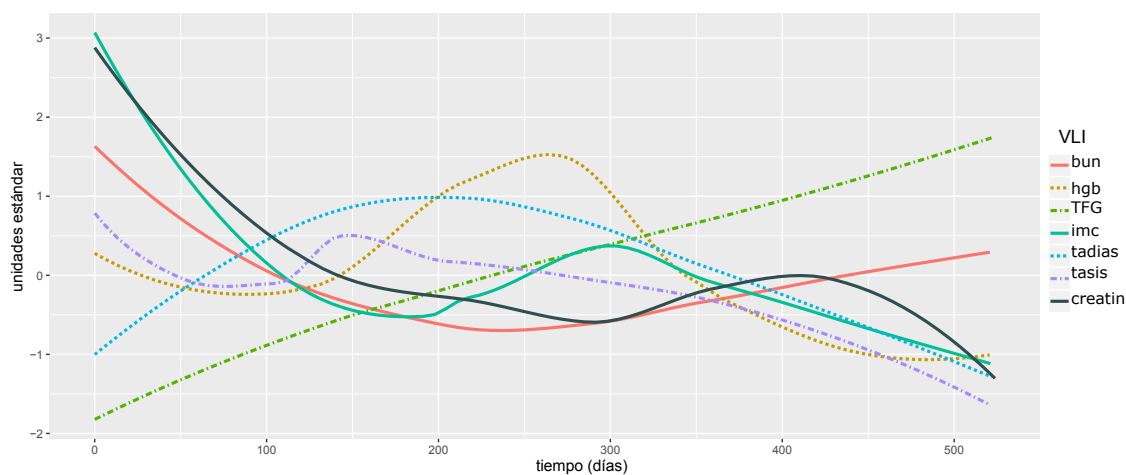


(c) Suavizadas por glm

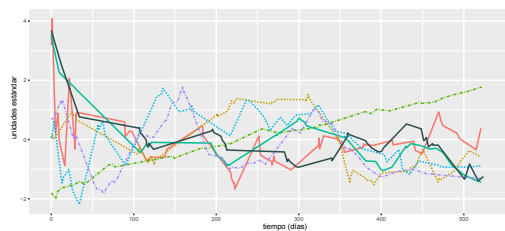
FIGURA 4.45. Trayectorias de las VLI de la clase 1

4.6.4.2. Trayectorias medias de las VLI: clase 2

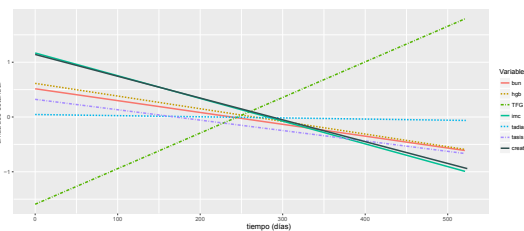
Las trayectorias gml (figura 4.46(c)) de esta clase muestran un progresión de la TFG positiva con una pendiente muy marcada, lo cual por sí misma indica una progresión positiva de esta clase en la ERC. Esta suposición se refuerza al observar la progresión negativa en el tiempo del imc, la creatinina, el bun y la tensión sistólica. Si bien la hemoglobina también decae progresivamente, lo hace muy lentamente, además la tensión diastólica parece mantenerse constante en el tiempo, lo cual es un buen indicio. En las trayectorias loess y crudas (figuras 4.46(a) y 4.46(b) respectivamente) se puede apreciar el marcado crecimiento (casi lineal) de la TFG en el tiempo observado en la figura 4.46(c), mientras que el bun, la creatinina y el imc decaen (rápidamente durante los primeros 250 días) para luego variar más o menos estable hasta después de los 500 días. Nuevamente se puede ver la alta variabilidad de la tensión sistólica, aunque en general, tiende a decrecer muy lentamente en el tiempo, mientras que la tensión diastólica aumenta aceleradamente durante los primeros 200 días, para luego decrecer paulatinamente en el tiempo.



(a) Suavizadas por loess



(b) Crudas

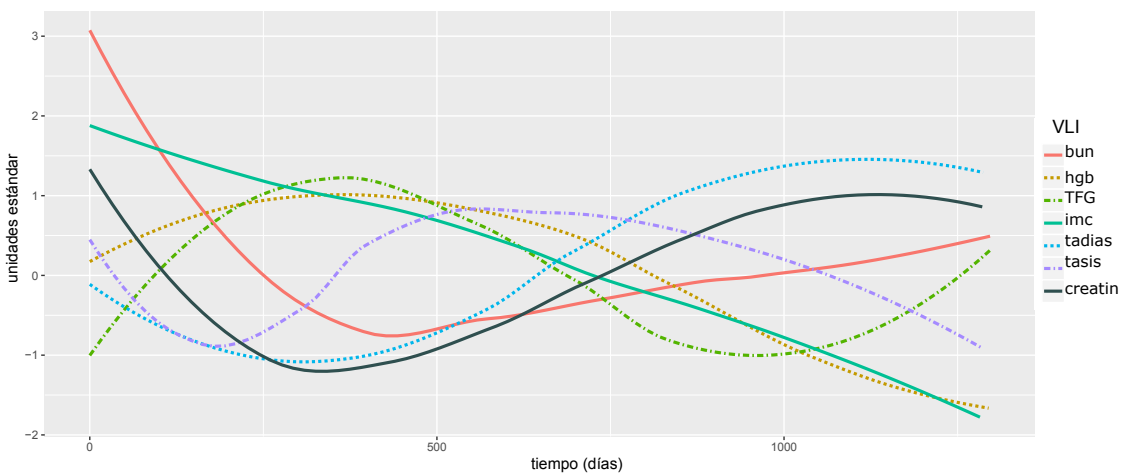


(c) Suavizadas por glm

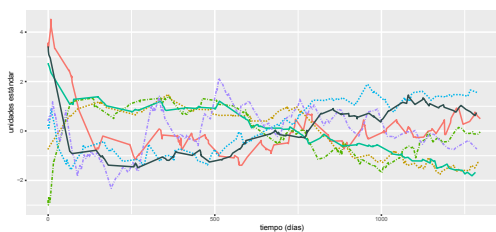
FIGURA 4.46. Trayectorias de las VLI de la clase 2

4.6.4.3. Trayectorias medias de las VLI: clase 3

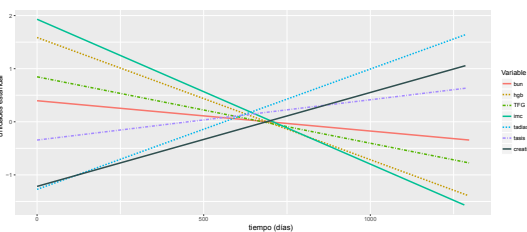
Esta clase se caracteriza por caídas progresivas de la TFG y del bun, aunque no muy marcadas 4.47. El imc y la hemoglobina también decrecen pero con mayor velocidad que las anteriores variables, mientras que la tensión sistólica y diastólica aumentan paulatinamente en el tiempo (esto concuerda con el hecho de que esta clase se caracteriza porque sus individuos son hipertensos, sección 4.6.2). Por otro lado, las trayectoria suavizada por loess de la TFG (y la cruda), presenta una alta variabilidad en el tiempo, con un crecimiento importante durante el primer año de estudio, seguida de una caída marcada hasta antes de los primeros 1000 días, para luego volver a crecer hasta finales del estudio. Esto puede llevar a pensar en una progresión razonablemente estable de la ERC, ya que los valores de la TFG varían más o menos en el mismo rango a través del tiempo. También llaman la atención la caída pronunciada de la trayectoria loess del bun en los primeros 500 días y la trayectoria casi lineal (también con pendiente negativa) del imc.



(a) Suavizadas por loess



(b) Crudas

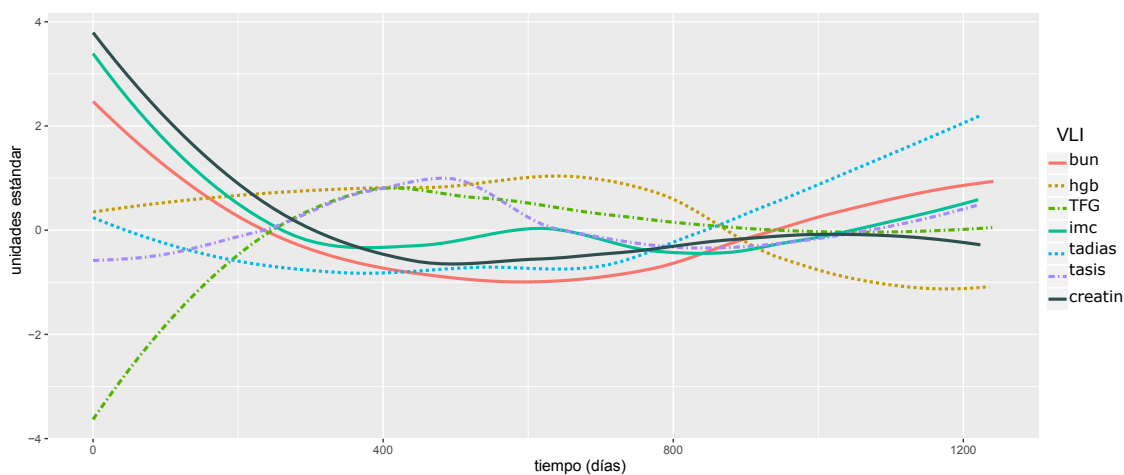


(c) Suavizadas por glm

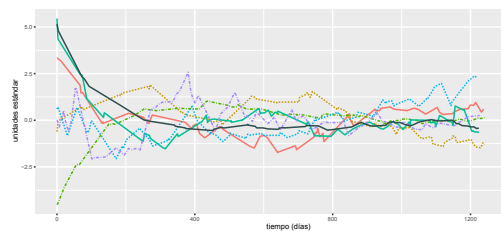
FIGURA 4.47. Trayectorias de las VLI de la clase 3

4.6.4.4. Trayectorias medias de las VLI: clase 4

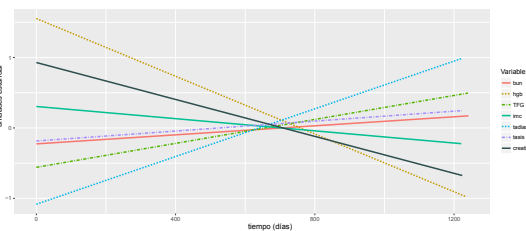
La figura 4.48(c) presenta trayectorias glm de la TFG, la tensión sistólica, el imc, la tensión diastólica y el bun con pendientes positivas (estas dos últimas con pendientes cercanas a cero). Estas trayectorias indican una progresión positiva de la ERC para esta clase. Si bien, la trayectoria suavizada por glm de la hemoglobina muestra un decrecimiento marcado en el tiempo, la realidad es que las trayectorias loess y crudas de las figuras 4.48(a) y 4.48(b) respectivamente, muestran una hemoglobina relativamente estable en el tiempo. En estos planos la TFG crece rápidamente durante los primeros 400 días, mientras que en el mismo intervalo la creatinina, el imc y el bun hacen lo contrario, para luego estabilizarse como la TFG. Por otro lado, las tensiones arteriales sistólica y diastólica, pese a la tendencia de esta última a crecer en la segunda mitad del tiempo de estudio, pueden también considerarse estables. En general, sobre todo si se observan las trayectorias crudas, se observa una tendencia estable de cada una de las variables a medida que el tiempo transcurre para esta clase.



(a) Suavizadas por loess



(b) Crudas

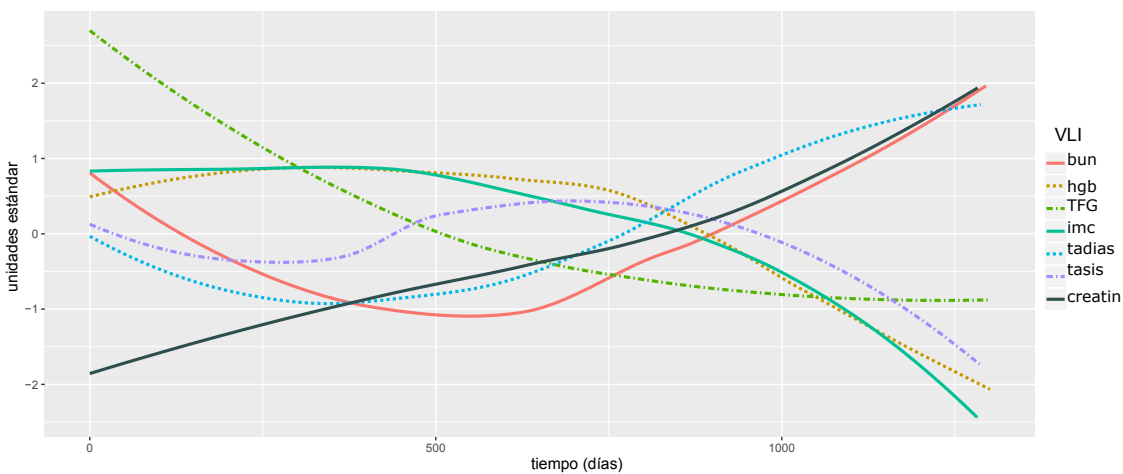


(c) Suavizadas por glm

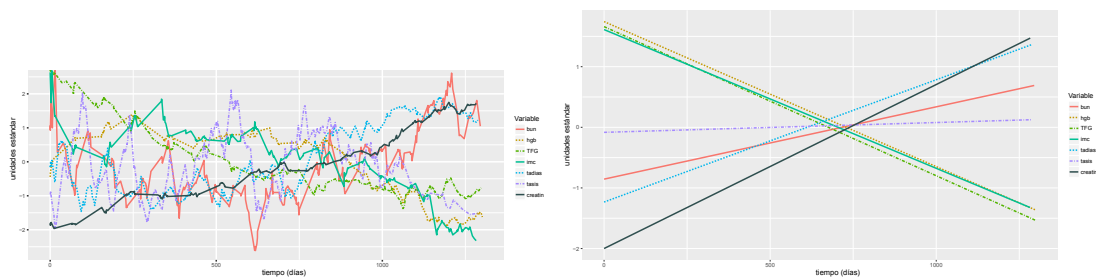
FIGURA 4.48. Trayectorias de las VLI de la clase 4

4.6.4.5. Trayectorias medias de las VLI: clase 5

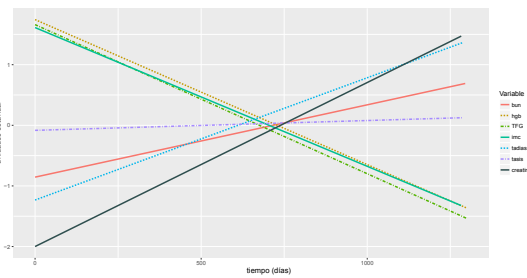
La clase 5 presenta trayectorias glm de la TFG, la hemoglobina y el imc que decaen muy rápidamente en el tiempo (prácticamente en la misma proporción), mientras que las trayectorias de la creatinina, el bun y la tensión diastólica crecen de un modo acelerado (la tensión sistólica también crece, pero muy lentamente). Estas condiciones obligan a pensar en una progresión negativa de esta clase a través de la ERC, lo cual se confirma si se observan estas mismas trayectorias, pero suavizadas por loess (figura 4.49(a)). En el plano de estas últimas trayectorias, la TFG se muestra en franco decrecimiento durante todo el periodo de estudio; Este mismo comportamiento lo presentan las trayectorias de la hemoglobina, el imc y la tensión sistólica, mientras que la tensión diastólica y el bun progresan negativamente la primera tercera parte del tiempo de estudio, para luego crecer rápidamente durante los dos restantes tercios. Un aspecto interesante que vale la pena mencionar, es la alta variabilidad que se observa en las trayectorias crudas de esta clase si se comparan con las trayectorias crudas de las clases anteriores.



(a) Suavizadas por loess



(b) Crudas

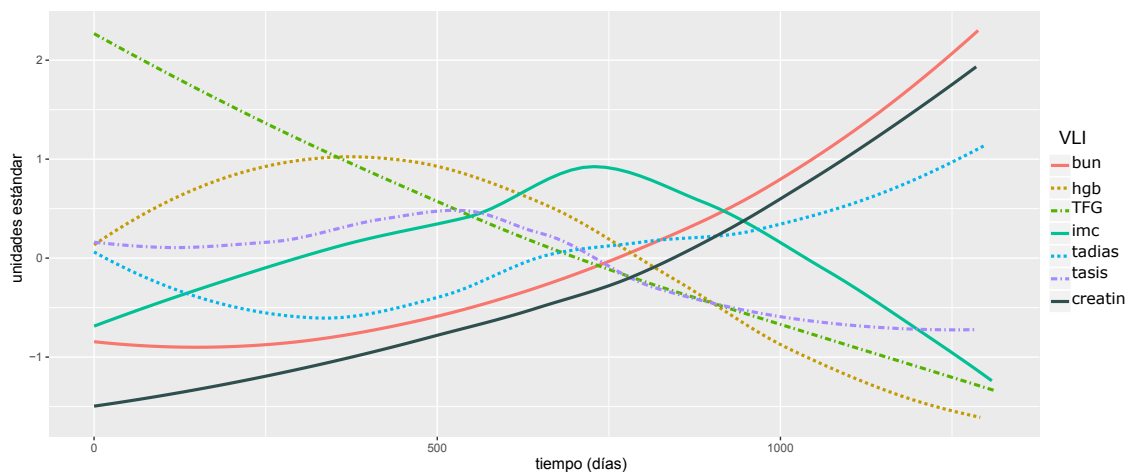


(c) Suavizadas por glm

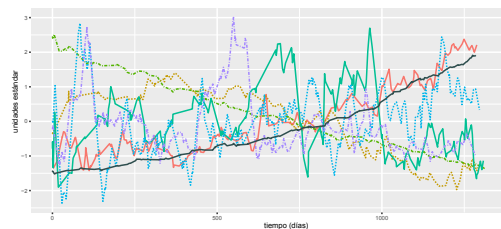
FIGURA 4.49. Trayectorias de las VLI de la clase 5

4.6.4.6. Trayectorias medias de las VLI: clase 6

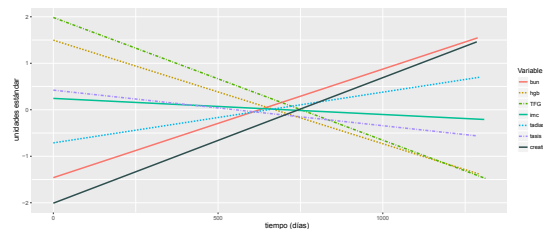
El “perfil” basado en las trayectorias glm de esta clase es muy similar al perfil de trayectorias glm de la clase 5, salvo que la trayectoria del imc decrece, pero muy lentamente 4.50(c). De este modo, se puede afirmar que la progresión de esta clase resulta decididamente negativa en su ERC. Las figuras 4.50(a) y 4.50(b) donde se muestran las trayectorias loess y crudas respectivamente, presentan una caída permanente de la TFG; la hemoglobina aumenta lentamente durante los primeros 400 días, para luego disminuir drásticamente hasta finalizar el periodo de estudio; similar comportamiento lo reporta la trayectoria del imc (aunque con mayor variabilidad) pues crece permanentemente durante los primeros 750 días, y luego decae progresivamente el resto del tiempo. Un comportamiento opuesto a estas trayectorias la presenta la tensión diastólica, pues decrece inicialmente, para luego aumentar continuamente. A su vez, la creatinina y el bun crecen de un modo acelerado durante todo el tiempo de estudio, mientras la tensión sistólica aumenta (primeros 500 días) y luego disminuye progresivamente (últimos 750 días). Nuevamente se muestran altos niveles de ruido en esta clase, que se traducen en alta variabilidad en los datos.



(a) Suavizadas por loess



(b) Crudas

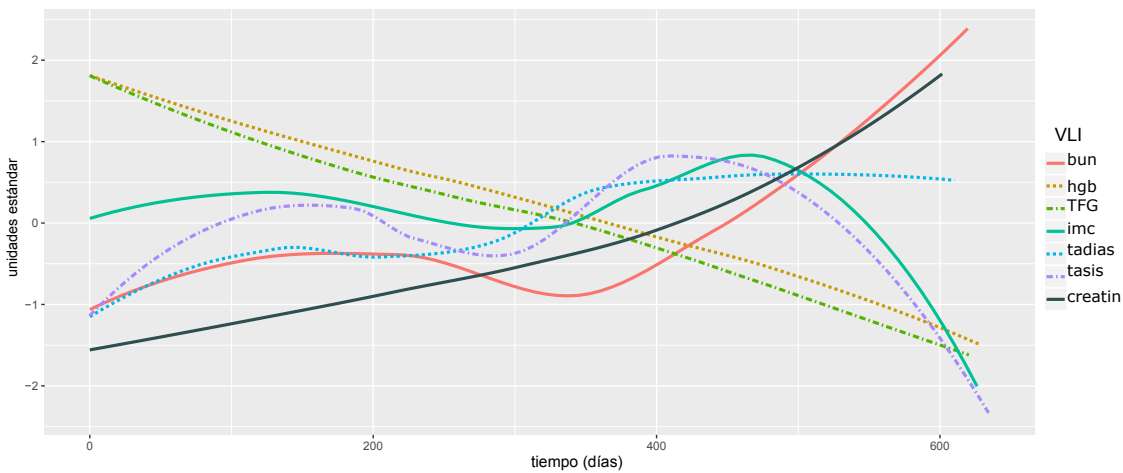


(c) Suavizadas por glm

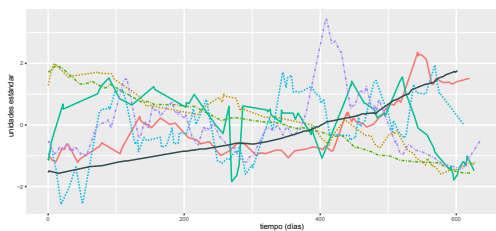
FIGURA 4.50. Trayectorias de las VLI de la clase 6

4.6.4.7. Trayectorias medias de las VLI: clase 7

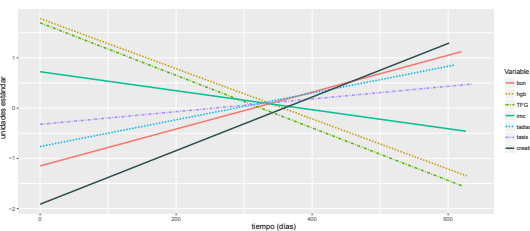
Al igual que la clase 5 y la clase 6, esta clase presenta trayectorias glm de la TFG y la hemoglobina con drásticas caídas durante de todo el periodo de estudio (figura 4.51(c)). Esta misma situación se observa en el imc (aunque con una pendiente menos pronunciada), mientras que el bun, la creatinina, la tensión sistólica y la tensión diastólica aumentan paulatinamente. En cuanto a las trayectorias de la TFG y la hemoglobina suavizadas por loess (figura 4.51(a)), se observa una tendencia decreciente casi lineal, disminuyendo casi en la misma proporción. Otro aspecto interesante lo presentan las trayectorias del imc y de la tensión sistólica que después de progresar relativamente estables durante los primeros 500 días del estudio, decaen drásticamente en los últimos días (aproximadamente 120) del mismo. Una tendencia opuesta la muestra el bun, que varía relativamente estable el primer año, punto en el cual crece vertiginosamente hasta terminar el tiempo de estudio. Finalmente, la creatinina y la tensión diastólica muestran un crecimiento sostenido en todo el periodo de observación, aunque esta última demuestra mayor variabilidad.



(a) Suavizadas por loess



(b) Crudas



(c) Suavizadas por glm

FIGURA 4.51. Trayectorias de las VLI de la clase 7

4.6.5. Comparación de las clases de la ERC mediante las VLI

En esta última sección se comparan las clases de progresión de la ERC a través de las trayectorias que describen las distintas variables longitudinales. Nuevamente las trayectorias que se presentan son suavizadas mediante loess.

4.6.5.1. Trayectorias de la TFG para las clases de progresión de la ERC

La figura 4.52 describe las trayectorias de la TFG para cada una de las siete clases. En este plano, la clase que mejor progresa a través de la ERC (mayores valores de la TFG) es la clase 4. Esta clase presenta una trayectoria que inicia con una tasa de $55 \text{ ml/min}/1.73\text{m}^2$ aproximadamente, que según la tabla 1.1 corresponde a estadio *E3a*, crece hasta alcanzar un máximo (aproximadamente 63, estadio *E2*) al finalizar el primer año de estudio, para luego decrecer muy lentamente, culminando con una tasa aproximada de $62 \text{ ml/min}/1.73\text{m}^2$ (estadio *E2*).

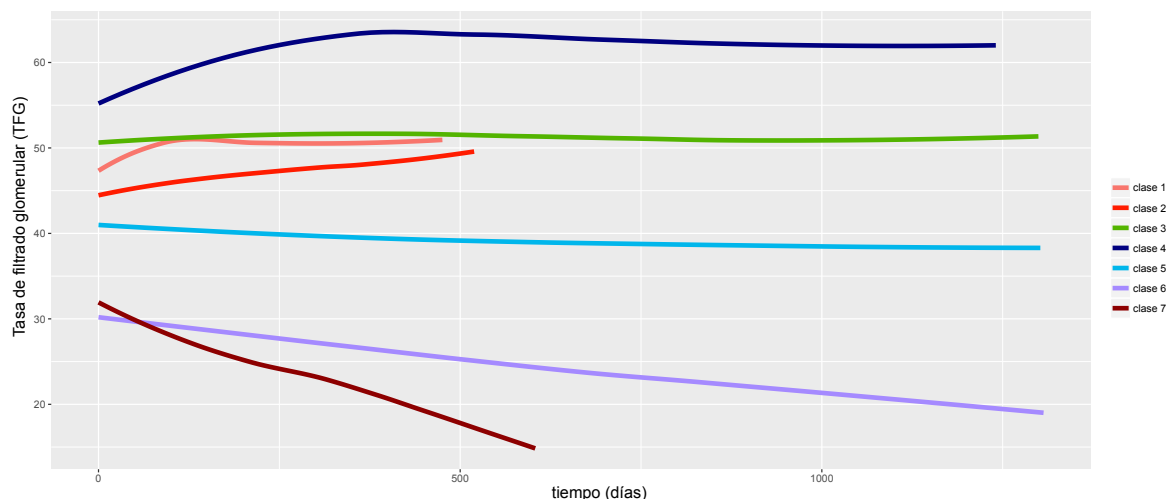


FIGURA 4.52. Trayectorias de la TFG estimada para las clases de progresión de la ERC

La clase 3, por otro lado, presenta una TFG prácticamente constante en 51 (estadio *E3a*) a través del todo el periodo de su estudio, mientras que las clases 1 y 2 presentan tasas que crecen permanentemente durante todo su tiempo de observación, con la diferencia que la primera clase inicia su trayectoria con una tasa de 47 (estadio *E3a*), mientras que la segunda clase lo hace en 44 (estadio *E3b*). La clase 5 presenta un decaimiento permanente, aunque no muy marcado en su TFG, en todo su periodo de observación, comenzando con

una tasa cercana a 41 (estadio *E3b*) y finalizando su trayectoria con una tasa aproximada a 38 (en el mismo estadio). La clase 6, por su parte, también decrece en su TFG, pero de un modo más acelerado que la clase 5, con una tasa inicial de 30 (estadio *E3b*) y una tasa al final de su estudio menor a 20 (estadio *E4*). Finalmente, la clase 7, muestra una caída dramática en su TFG, pues en menos de dos años, pasa de una tasa inicial cercana a 32 (aunque un poco mayor que la tasa inicial de la clase 6) a una tasa final de 10, lo cual implica la remisión de sus individuos inequívoca a terapia de reemplazo renal.

4.6.5.2. Trayectorias de la creatinina para las clases de progresión de la ERC

Al comparar las trayectorias de la creatinina (expresadas en g/dl) ilustradas en la figura 4.53 con las trayectorias de la TFG de la figura 4.52, automáticamente se pone de manifiesto la relación inversa presente entre ellas. Nuevamente se identifican los dos grupos de trayectorias observados en la TFG: clases 6 y 7 con altos niveles de creatinina y un acelerado crecimiento en un primer grupo, y un segundo grupo de clases, del 1 al 5, con valores de la creatinina relativamente estables.

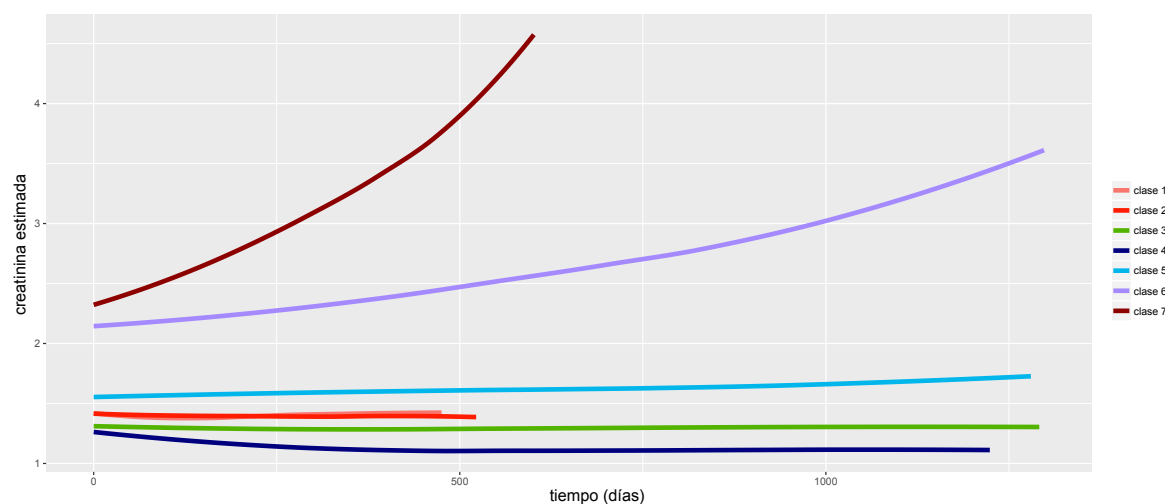


FIGURA 4.53. Trayectorias de la creatinina estimada para las clases de progresión de la ERC

4.6.5.3. Trayectorias del bun para las clases de progresión de la ERC

El nitrógeno ureico en la sangre es medido en mg/dl . El bun es producido por la descomposición de las proteínas presentes en los alimentos consumidos. Valores normales del bun varían entre 7 y 20 mg/dl ; cuando la función renal disminuye permanentemente,

es usual que el bun se eleve. El plano de la figura 4.54 muestra dos grupos de trayectorias: un primer grupo conformado por las trayectorias bun de la clase 7 y la clase 6, que se caracterizan por un crecimiento sostenido (muy dramático para la clase 7) a lo largo de sus respectivos periodos de observación (evidencian una progresión muy negativa en su progresión de la ERC); el segundo grupo, por su parte, compuesto por las clases 1 al 5 muestran niveles de bun más bajos y constantes, aunque mayores a los niveles normales, por lo que hay evidencia que confirma muchas de las afirmaciones realizadas en apartados anteriores, en cuanto al modo como cada una de las clases progresan a través de su ERC.

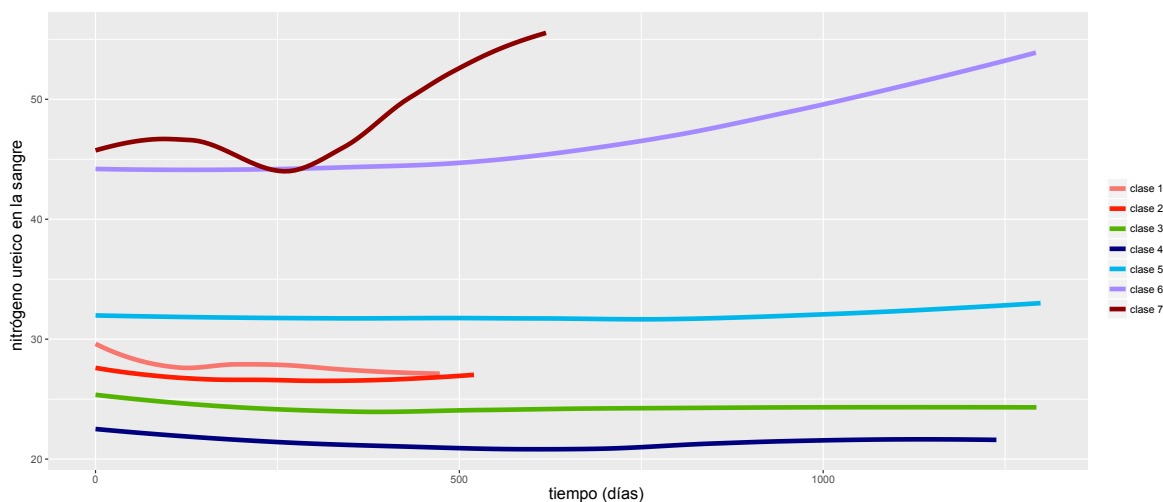


FIGURA 4.54. Trayectorias del nitrógeno ureico en la sangre para las clases de progresión de la ERC

4.6.5.4. Trayectorias de la hemoglobina para las clases de progresión de la ERC

La anemia se considera una condición asociada al mal funcionamiento de los riñones, conteos de hemoglobina menores a 13.5 g/dl son evidencia esta condición, y muy posiblemente, de daño renal (sección 1.4). De un modo general, la trayectorias de la hemoglobina de las las clases 4, 3, 2, 5 y 6, muestran una progresión estable durante la primera mitad de sus respectivos tiempos de observación, para luego decaer suavemente hasta el final de los mismos (figura 4.55). Por su parte, la clase 1 observa un crecimiento marcado en su hemoglobina durante la primera parte de su estudio, para luego, como sucedió con las anteriores clases, decaer hasta final del periodo de observación. En cuanto a la clase 7, a diferencia de las demás clases, presenta un descenso ininterrumpido y acelerado de la hemoglobina. Al comparar cada una de las trayectorias, se evidencia que la clase 4 presenta los valores más

altos de hemoglobina (lo cual indica una progresión favorable en la ERC), mientras que en su orden, las clases 3, 1, 2, 5, 6 y 7 muestran valores más bajos. Dado que los niveles de hemoglobina de las clases 6 y 7 varían por debajo de 13.5 se puede afirmar que estas dos clases muestran las progresiones de la ERC de mayor riesgo si se comparan con las otras clases. Esta configuración de las trayectorias está en concordancia con la clasificación de los tipos de progresión de la ERC realizada en apartados anteriores.

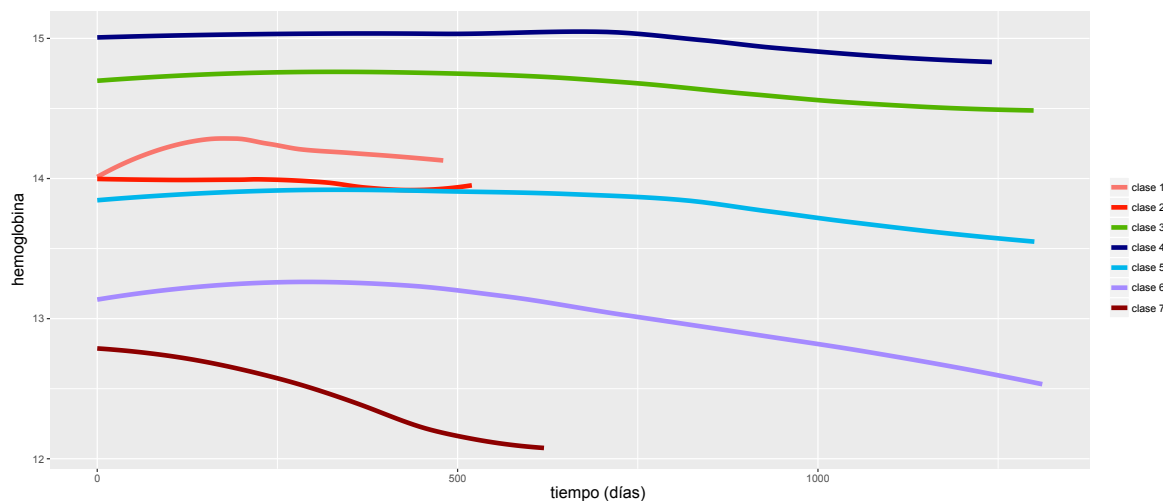


FIGURA 4.55. Trayectorias de la hemoglobina para las clases de progresión de la ERC

4.6.5.5. Trayectorias del índice de masa corporal para las clases de progresión de la ERC

El plano de la figura 4.56 muestra las trayectorias del índice de masa corporal (en kg/m^2) de las siete clases de progresión de la ERC. Dado que estos índices varían entre 26 y 27.5 para cada una de las clases, los individuos a las que ellas pertenecen presentan sobrepeso (sección 1.4). A excepción de la clase 6, todas las trayectorias del imc describen un decrecimiento en sus valores durante sus respectivos periodos de observación. La clase 4, que es la que mejor progresa en la ERC de todas las clases, presenta los menores índices (varían entre 26.2 y 26), mientras que la clase 5 (con una progresión de la ERC de riesgo moderadamente alto) muestra los índices más altos (con una caída importante de los mismos) que varían entre valores cercanos a 27.2 y a 26.8 respectivamente. Las clases que muestran caídas vertiginosas en sus imc son la clase 1, la clase 2 y la clase 7, aunque esta última llama la atención porque presenta alta variabilidad a lo largo de su tiempo

de observación, al punto que presenta un cambio brusco en sus valores casi al final del mismo. Para concluir, resulta importante señalar que el imc, tal como se expresa a través de sus trayectorias, no parece ser un buen indicador del modo como cada una de las clases progresan en la ERC. Esto puede ser debido a un efecto de enmascaramiento de otros factores que no permiten observar de un modo directo su relación con el progreso de esta enfermedad.

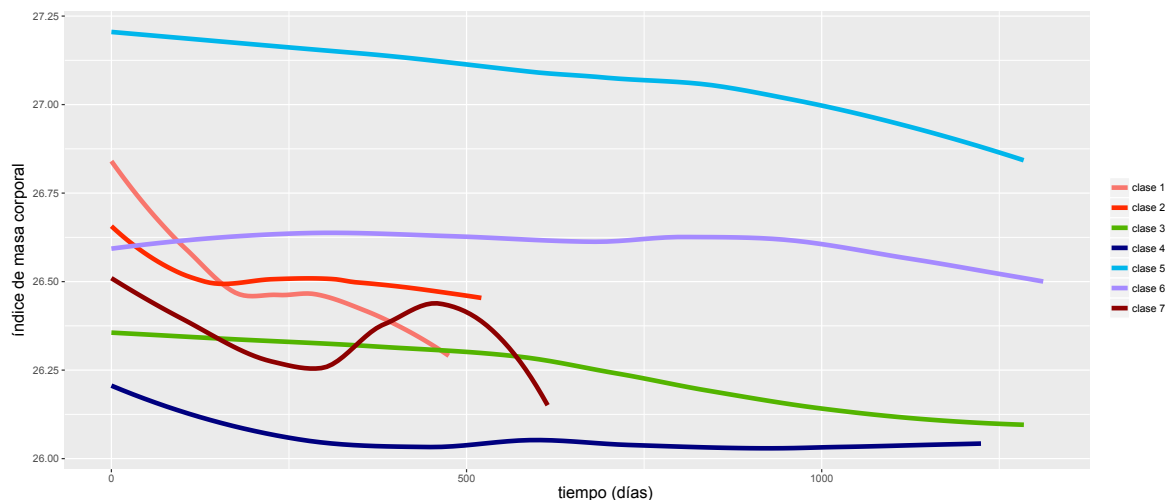


FIGURA 4.56. Trayectorias del índice de masa corporal para las clases de progresión de la ERC

4.6.5.6. Trayectorias de la tensión sistólica para las clases de progresión de la ERC

Valores mayores a 120 mmHg en la tensión sistólica son evidencia de tensión arterial alta (HTA), la cual a su vez, se considera un factor de riesgo de fallo renal (sección 1.4). En la figura 4.58 es evidente la marcada diferencia entre la trayectoria de la tensión sistólica de la clase 7 y las trayectorias de las demás clases. Esta clase muestra un crecimiento acelerado durante gran parte de su primer año de observación, punto en el cual alcanza un pico (en 160 mmHg) para luego caer drásticamente hasta finalizar de su periodo de estudio (140 mmHg). Los valores entre los que se mueven estas tensiones indican un claro riesgo de desarrollar fallo renal para esta clase. Otras clases, como la 4, la 3 y la 5 describen tensiones relativamente estables, mientras que las clases 1 y 2 muestran trayectorias que decrecen rápidamente durante la primera tercera parte de su periodo de estudio, crecen durante la segunda tercera parte, para volver a caer hasta el final. Por su parte, se observa

que la trayectoria de la clase 6 crece paulatinamente durante los dos primeros años de su progresión y luego realiza lo contrario, un poco más rápido, hasta alcanzar una tensión de 130 mmHg . Dado que todas las clases presentan tensiones mayores a 120, unas más controladas que otras, se pone en evidencia el riesgo que cada una de ellas presenta de progresar de manera no favorable a través de la ERC.

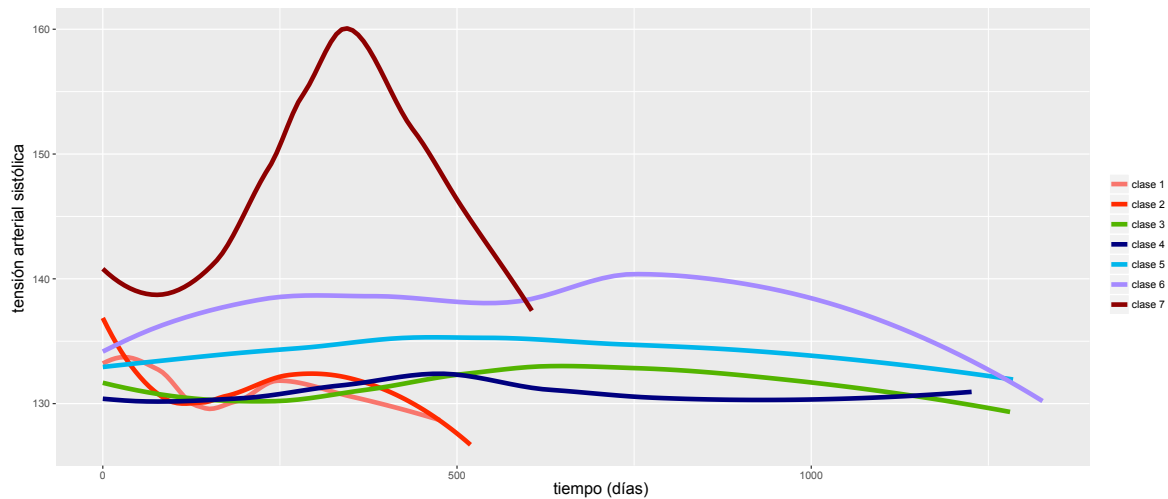


FIGURA 4.57. Trayectorias de la tensión arterial sistólica para las clases de progresión de la ERC

4.6.5.7. Trayectorias de la tensión diastólica para las clases de progresión de la ERC

Se considera que un individuo presenta HTA cuando su tensión diastólica es mayor a 90 mmHg . Si esto sucede, como en el caso de la tensión sistólica, aumenta su riesgo de desarrollar fallo renal (sección 1.4). Las trayectorias de las tensiones diastólicas (figura 4.58), a diferencia de las trayectorias sistólicas presentes en la figura 4.57, muestran una alta variabilidad a lo largo de sus respectivos tiempos de observación. Nuevamente la clase 7, igual que en el caso de la tensión sistólica, muestra un crecimiento acelerado al principio del estudio, para luego decrecer drásticamente. Al contrario, la clase 1 muestra una tensión diastólica que decae rápidamente durante los primeros días del estudio, aunque luego tiende a estabilizarse. Se observa que las clases 4, 5 y 3 presentan trayectorias de las tensiones similares, con una caída suave al principio para luego presentar un crecimiento muy suave hasta final de sus tiempos de observación respectivos. Por su parte, las clases 2 y 6 describen tensiones diastólicas mayores al grupo de clases descritas anteriormente,

aunque se diferencian en que la clase 2 muestra una trayectoria que primero aumenta y luego disminuye, mientras que la clase 6 progresa más o menos constante durante casi todo el tiempo, pero la última parte de este, aumenta de modo acelerado.

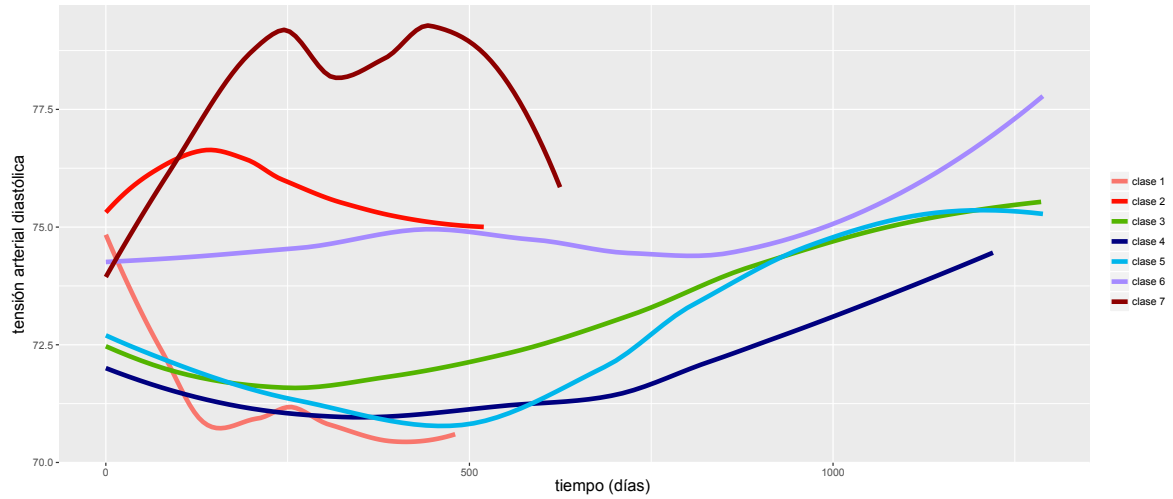


FIGURA 4.58. Trayectorias de la tensión arterial diastólica para las clases de progresión de la ERC

Conclusiones

La combinación del *análisis cualitativo armónico* (AAC) con el *análisis factorial múltiple* (AFM), permitió realizar un estudio tipológico de los perfiles de progresión de 2445 individuos a través de los distintos estadios de la *enfermedad renal crónica* (ERC), los cuales estaban adscritos a un programa de prevención secundaria y fueron observados por algo menos de 5 años (1820 días).

En lugar de trabajar directamente con las transiciones de un estadio a otro de la ERC, que constituye una *variable longitudinal cualitativa* (VLC) susceptible a ser estudiada a través un AAC, ésta se recodificó en términos de los tiempos de permanencia de cada individuo en cada estadio de la ERC (incluidas dos categorías de censura del estudio y otra de remisión a terapia de reemplazo renal) dentro de catorce periodos de corte en los que se dividió convenientemente el tiempo de estudio (sección 4.3.2). Con los datos así codificados se pudo realizar sobre ellos un *análisis de correspondencias difuso* (ACD), cuyos resultados son teóricamente equivalentes a los resultados del AAC.

Realizar el AAC como un análisis de correspondencias, permitió acceder a planos factoriales y a distintas medidas que permitieron: explorar las trayectorias de los individuos a través de su ERC y detectar patrones entre ellas; estudiar relaciones entre los periodos de corte del tiempo de estudio y, describir correspondencias entre las trayectorias y dichos periodos (sección 4.5.1). Por otro lado, a partir de las coordenadas factoriales resultantes del análisis, y tras realizarse una clasificación automática sobre ellas, se obtuvieron siete clases distintas de progresión de la ERC, que si bien no fueron estudiadas detenidamente, si se constituyeron un primer acercamiento a los perfiles de progresión de la ERC definitivos (sección 4.5.1.1).

Al incorporar el AFM al AAC se logró, entre otras cosas: obtener una síntesis general de toda la información presente en los datos considerando todos los periodos de corte, lo mismo que, síntesis parciales de la información dentro de cada uno de ellos; equilibrar la influencia de cada uno de los periodos en el análisis global tomando en cuenta sus correspondientes estructuras internas; acceder a medidas de comparación y asociación entre tales periodos y finalmente, incorporar individuos y variables transversales ilustrativas (VTI) al análisis para la caracterización de los perfiles de progresión de la ERC que fueron determinados (sección 4.5.2).

Siete clases de progresión de la ERC distintas fueron obtenidas tras aplicar una clasificación automática sobre las coordenadas factoriales de los individuos, resultantes de la combinación del AAC con el AFM (sección 4.6). La clase 3 fue la más grande, representando un poco más del 30 por ciento de la población total, mientras que las restantes seis clases se ordenaron de manera decreciente, según sus tamaños relativos, así: clase 2 (22.61 %), clase 5 (19.63 %), clase 4 (10.67 %), clase 1 (9.48 %), clase 6 (4.38 %) y clase 7 (2.08 %).

Un acercamiento inicial a estas clases se realizó a partir de la observación de las trayectorias medias de individuos artificiales ilustrativos en los distintos estadios y categorías de censura de la ERC, en cada uno de los catorce periodos de corte (sección 4.6.1). Este primer acercamiento fue complementado con la caracterización de cada clase a partir de la propia VLC codificada, las transiciones entre estadios más frecuentes dentro de los periodos de corte y el grupo de VTI. Esto permitió construir para cada clase perfiles particulares de progresión a través de la ERC e identificar posibles factores de riesgo de progresión para cada una de ellas (sección 4.6.2). Dentro de las características más relevantes de estas clases se pueden citar:

Las dos primeras clases, la 1 y la 2, se caracterizaron porque todos sus individuos fueron censurados del estudio; la primera clase porque sus integrantes salieron prematuramente del mismo por causas distintas a la remisión a terapia de reemplazo renal, mientras que para la segunda clase, la censura obedeció a que todos sus individuos sobrevivieron al tiempo de estudio. Además, se logró establecer para estas clases progresiones relativamente estables, y nivel de riesgo bajo-medio, en su ERC, con permanencias en estadios *E3a* y *E3b* y tasas de filtrado glomerular (TFG) superiores a la media global.

La clase 3, presentó una progresión a través de la ERC muy similar a la progresión mostrada por la totalidad de la población. Aunque sus TFG iniciales, promedios y finales se marcaron por encima de la media global y la mayor permanencia se registró en estadio *E3b*, se determinó que esta clase estaba en un nivel de riesgo de progresión medio, pues como causa principal de su ERC se identificó la tensión arterial alta (HTA), el cual se considera un factor de riesgo de progresión dentro de la enfermedad renal.

La clase con progresión de la ERC más favorable, fue la clase 4. Se identificó para ella que los estadios de mayor permanencia fueron los estadios *E2* y *E3a* y que sus TFG estuvieron todas por encima de los promedios globales. También fue posible determinar que la mayoría de sus individuos tenían como causa de su ERC la HTA, pero que también en su mayoría, fueron excluidos del estudio por ser remitidos a consulta externa.

Por otro lado, las clases 5, 6 y 7 mostraron las progresiones menos positivas y estables. Cada una de ellas se caracterizaron porque sus individuos eran diabéticos y porque registraron TFG iniciales, promedios y finales por debajo de las respectivas medias globales. De modo particular, la clase 5, en su mayoría mujeres, permanecieron sobre todo en estadio *E4* y tuvieron un desenlace fatal, mientras que la clase 6, catalogada dentro de un riesgo alto de progresión negativa, mostró mayor permanencia en estadio 5 y registró una proporción importante de individuos remitidos a diálisis. Finalmente, la clase 7 mostró la tendencia de progresión de la ERC menos favorable de todas clases. Esta clase se caracterizó porque el estadio de mayor permanencia fue el estadio *E5* y porque la totalidad de sus individuos fueron incluidos en terapia de reemplazo renal.

En una última etapa, cada una de las clases de progresión fueron estudiadas a través de las trayectorias en el tiempo de distintas variables longitudinales, como la propia TFG, la creatinina, el nitrógeno ureico en la sangre, la tensión arterial sistólica, entre otras (sección 4.6.4). La observación y descripción de estas trayectorias, no sólo permitió comparar cada una de las clases identificadas, sino que ratificó las distintas afirmaciones relativas a sus tipos particulares de progresión, los niveles de riesgo donde fueron incluidas y los factores que las determinaron.

En resumen, el AAC-AFM se constituyó en una metodología apropiada para la descripción de datos cualitativos longitudinales (en particular de la progresión a través de estadios de la ERC). También demostró ser un medio útil y adecuado para transformar este tipo de datos, y dar paso a procedimientos de clasificación con una estructura común. A nivel clíni-

co, los distintos tipos de progresión a través de la ERC logran ser identificados plenamente y se confirman la diabetes, el HTA y la edad como factores de progresión acelerada dentro de esta enfermedad. Por otro lado, debido a los tipos de progresión identificados (cuatro de los siete catalogados en riesgo bajo-medio), la baja proporción de individuos remitidos a TRR (93 individuos de 2445 posibles) o con desenlace fatal, entre otros aspectos, demuestran que la inclusión de individuos con ERC a programas de prevención y seguimiento, en particular el programa dentro del cual se realizó este estudio, es un medio efectivo en la intervención de la enfermedad, retrasando en gran medida el tiempo de progresión a estadios terminales, controlando los factores de progresión y mejorando, en general, la calidad de vida de los pacientes.

Desarrollos futuros y oportunidades

Incluir variables longitudinales, y no sólo las variables transversales como en el presente caso, dentro de los análisis factoriales como elementos ilustrativos puede enriquecer la descripción de los perfiles de progresión de la ERC. Al realizarse esto, debería accederse a medidas de asociación entre las distintas variables con las distintas componentes principales del análisis, permitiendo determinar de manera más precisa en cada periodo de corte, aquellos factores que pueden acelerar o retrasar la ERC a una fase terminal.

Por otro lado, al actualizar la clasificación de la ERC utilizada en este trabajo a la nueva clasificación pronóstica de la ERC basada en la TFG y la albuminuria (Eknoyan et al. 2013), se podría realizar un estudio más detallado de la progresión de los individuos a través de su enfermedad.

Finalmente, el estudio de las clases de progresión de la ERC podría trasladarse al contexto de los métodos de clasificación no supervisada a partir del análisis de correspondencias múltiples de datos funcionales categóricos (Besse 1979, Saporta 1981, Boumaza 1980), que también se basan en el AAC.

Referencias

- Cuenta de Alto Costo (2015), Situación de la enfermedad renal crónica, hipertensión arterial y diabetes mellitus en Colombia 2015, Technical report, Fondo Colombiano de Enfermedades de Alto Costo. Ministerio de la Salud y la Protección Social.
- Abascal, E., Fernández, K., Madroño, J. & Landaluce, M. (1999), ‘Técnicas factoriales de análisis de tablas múltiples: Nuevos desarrollos empíricos’, *Proyectos de Investigación PB98-0149 de la Dirección General de Enseñanza Superior del Ministerio de Educación y Cultura* y pp. 3–5.
- Abascal Fernández, E. & Landaluce Calvo, M. I. (2002), ‘Análisis factorial múltiple como técnica de estudio de la estabilidad de los resultados de un análisis de componentes principales’.
- Barbary, O. (1996), *Análisis tipológico de datos biográficos en Bogotá*, Universidad Nacional Colombia–ORSTOM, Bogotá, Colombia.
- Barbary, O. (1997), Análisis estadístico de datos biográficos: métodos, ejemplos y perspectivas en el estudio de itinerarios migratorios, *in* ‘Taller de medición de la migración internacional’, Colección COLEF, El Colegio de la Frontera Norte, pp. 11–34.
- Barbary, O. & Pinzón, L. M. S. (1998), ‘El análisis armónico cualitativo: Teoría y aplicación a la tipología de trayectorias individuales’, *Revista Colombiana de Estadística* **22**(1), 26–58.
- Bécue-Bertaut, M. & Pagès, J. (2008), ‘Multiple factor analysis and clustering of a mixture of quantitative, categorical and frequency data’, *Computational Statistics & Data Analysis* **52**(6), 3255–3268.
- Benzécri, J.-P. (1973), *L’analyse des données*, Vol. 1, Dunod Paris.

- Besse, P. C. (1979), Etude descriptive d'un processus: Approximation et interpolation, PhD thesis.
- Boumaza, R. (1980), *Contribution à l'étude descriptive d'une fonction aléatoire qualitative*, Université Paul Sabatier.
- Cabrera, S. S. (2004), 'Definición y clasificación de los estadios de la enfermedad renal crónica. prevalencia. claves para el diagnóstico precoz. factores de riesgo de enfermedad renal crónica', *Nefrología* **24**(6), 27–34.
- Carroll, J. D. (1968), Generalization of canonical correlation analysis to three or more sets of variables, *in* 'Proceedings of the 76th annual convention of the American Psychological Association', Vol. 3, pp. 227–228.
- Chevenet, F., Dolédec, S. & Chessel, D. (1994), 'A fuzzy coding approach for the analysis of long-term ecological data', *Freshwater Biology* **31**(3), 295–309.
- Corrales, M. L. (2010), Combinación del análisis factorial múltiple y el análisis armónico cualitativo en la descripción de datos longitudinales cualitativos, Master's thesis, Universidad Nacional de Colombia. Facultad de Ciencias. Departamento de Estadística, Bogotá, DC.
- Corrales, M. L. & Pardo, C. E. (2015), 'Análisis de datos longitudinales cualitativos mediante la combinación de análisis armónico cualitativo y análisis factorial múltiple', *Comunicaciones en Estadística* **8**(1), 11–32.
- Corrales, M. L. & Pardo, C. E. (2016), *qha: Qualitative Harmonic Analysis*. R package version 0.0.8.
URL: <https://CRAN.R-project.org/package=qha>
- Daville, J. & Saporta, G. (1980), 'Analyse harmonique qualitative', *Data Analysis and Informatics* pp. 375–389.
- Deville, J.-C. (1982), Analyses de données chronologiques qualitatives: comment analyser des calendriers?, *in* 'Annales de l'INSEE', pp. 45–104.
- Deville, J.-C. & Saporta, G. (1983), 'Correspondence analysis, with an extension towards nominal time series', *Journal of Econometrics* **22**(1-2), 169–189.

- Diggle, P. (2002), *Analysis of longitudinal data*, Oxford University Press.
- Dray, S., Dufour, A.-B. et al. (2007), 'The ade4 package: implementing the duality diagram for ecologists', *Journal of Statistical Software* **22**(4), 1–20.
- Eknoyan, G., Lameire, N., Eckardt, K., Kasiske, B., Wheeler, D., Levin, A., Stevens, P., Bilous, R., Lamb, E., Coresh, J. et al. (2013), 'KDIGO 2012 clinical practice guideline for the evaluation and management of chronic kidney disease', *Kidney International* **3**, 5–14.
- Escofier, B. & Pagès, J. (1984), 'L'analyse factorielle multiple: une méthode de comparaison de groupes de variables', *Data Analysis and Informatics III, North-Holland, Amsterdam* pp. 375–389.
- Escofier, B. & Pagès, J. (1992), *Análisis factoriales simples y múltiples objetivos. Métodos e interpretación*, Universidad del País Vasco, Bilbao.
- Everitt, B. S., Stahl, D., Leese, M. & Landau, S. (2011), *Cluster analysis*, John Wiley & Sons.
- Fitzmaurice, G. M., Laird, N. M. & Ware, J. H. (2012), *Applied longitudinal analysis*, Vol. 998, John Wiley & Sons.
- Gámez, N. (2012), 'Fundamentos y aplicaciones del análisis de correspondencias difuso', *Comunicaciones en Estadística* **5**(1), 7–32.
- Genolini, C., Ecochard, R., Benghezal, M., Driss, T., Andrieu, S. & Subtil, F. (2016), 'kmlshape: An efficient method to cluster longitudinal data (time-series) according to their shapes', *PloS one* **11**(6), e0150738.
- Greenacre, M. & Blasius, J. (2006), *Multiple correspondence analysis and related methods*, CRC Press.
- Haynes, R. J. & Winearls, C. G. (2010), 'Chronic kidney disease', *Surgery (Oxford)* **28**(11), 525–529.
- Himmelfarb, J. & Sayegh, M. H. (2010), *Chronic Kidney Disease, Dialysis, and Transplantation: A Companion to Brenner and Rectors The Kidney*, Elsevier Health Sciences.
- Hsu, C.-y., McCulloch, C. E., Iribarren, C., Darbinian, J. & Go, A. S. (2006), 'Body mass index and risk for end-stage renal disease', *Annals of Internal Medicine* **144**(1), 21–28.

- Hu, B., Gadegbeku, C., Lipkowitz, M. S., Rostand, S., Lewis, J., Wright, J. T., Appel, L. J., Greene, T., Gassman, J., Astor, B. C. et al. (2012), 'Kidney function can improve in patients with hypertensive CKD', *Journal of the American Society of Nephrology* **23**(4), 706–713.
- Hunsicker, L. G., Adler, S., Caggiula, A., England, B. K., Greene, T., Kusek, J. W., Rogers, N. L., Teschan, P. E., Beck, G., of Diet in Renal Disease Study Group, M. et al. (1997), 'Predictors of the progression of renal disease in the modification of diet in renal disease study', *Kidney International* **51**(6), 1908–1919.
- Husson, F., Josse, J., Lê, S. et al. (2008), 'Factominer: an R package for multivariate analysis', *Journal of Statistical Software* **25**(1), 1–18.
- Kalton, G. & Citro, C. F. (1995), 'Panel surveys: adding the fourth dimension', *Innovation: The European Journal of Social Science Research* **8**(1), 25–39.
- Lebart, L., Morineau, A. & Piron, M. (1995), 'Statistique exploratoire multidimensionnelle', *Dunod. Paris* .
- Lebart, L., Piron, M. & Morineau, A. (2006), *Statistique exploratoire multidimensionnelle. Visualisation et inférence en fouilles de données*, 4 edn, Dunod, Paris.
- Levey, A. S. & Coresh, J. (2012), 'Chronic kidney disease', *The Lancet* **379**(9811), 165–180.
- Levey, A. S., Coresh, J., Bolton, K., Culeton, B., Harvey, K. S., Ikizler, T. A., Johnson, C. A., Kausz, A., Kimmel, P. L., Kusek, J. et al. (2002), 'K/DOQI clinical practice guidelines for chronic kidney disease', *American Journal of Kidney Diseases* **39**(2 SUPPL. 1).
- Levey, A. S., Perrone, R. D. & Madias, N. E. (1988), 'Serum creatinine and renal function', *Annual Review of Medicine* **39**(1), 465–490.
- Levey, A. S., Stevens, L. A. & Coresh, J. (2009), 'Conceptual model of CKD: applications and implications', *American Journal of Kidney Diseases* **53**(3), S4–S16.
- Li, L., Astor, B. C., Lewis, J., Hu, B., Appel, L. J., Lipkowitz, M. S., Toto, R. D., Wang, X., Wright, J. T. & Greene, T. H. (2012), 'Longitudinal progression trajectory of GFR among patients with CKD', *American Journal of Kidney Diseases* **59**(4), 504–512.

- Manzini, J. L. (2000), 'Declaración de Helsinki: Principios éticos para la investigación médica sobre sujetos humanos', *Acta Bioethica* **6**(2), 321–334.
- McCullagh, P. (1984), 'Generalized linear models', *European Journal of Operational Research* **16**(3), 285–292.
- Mezzano, S. & Aros, C. (2005), 'Enfermedad renal crónica: clasificación, mecanismos de progresión y estrategias de renoprotección', *Revista Médica de Chile* **133**(3), 338–348.
- Mills, K. T., Xu, Y., Zhang, W., Bundy, J. D., Chen, C.-S., Kelly, T. N., Chen, J. & He, J. (2015), 'A systematic analysis of worldwide population-based data on the global burden of chronic kidney disease in 2010', *Kidney International* .
- Mitch, W., Walser, M., Buffington, G. & Lemann, J. (1976), 'A simple method of estimating progression of chronic renal failure', *The Lancet* **308**(7999), 1326–1328.
- Norris, K. C., Greene, T., Kopple, J., Lea, J., Lewis, J., Lipkowitz, M., Miller, P., Richardson, A., Rostand, S., Wang, X. et al. (2006), 'Baseline predictors of renal disease progression in the african american study of hypertension and kidney disease', *Journal of the American Society of Nephrology* **17**(10), 2928–2936.
- Pagès, J. et al. (2004), 'Multiple factor analysis: main features and application to sensory data', *Revista Colombiana de Estadística* **27**(1), 1–26.
- Pardo, C. E. & Del Campo, P. C. (2007), 'Combinación de métodos factoriales y de análisis de conglomerados en R: el paquete FactoClass', *Revista Colombiana de Estadística* **30**, 231–245.
- Pinzón, L. M. (1998), Manejo del tiempo en el análisis armónico cualitativo. Movilidad residencial en la áreas urbanas de Bogotá, Master's thesis, Universidad Nacional de Colombia. Facultad de Ciencias. Departamento de Estadística, Bogotá, DC.
- R Core Team (2016), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.
URL: <https://www.R-project.org/>
- Robette, N. & Thibault, N. (2008), 'Comparing qualitative harmonic analysis and optimal matching. an exploratory study of occupational trajectories', *Population* **63**(4), 533–556.

- Saporta, G. (1981), Méthodes exploratoires d'analyse de données temporelles, PhD thesis, Université Pierre et Marie Curie-Paris VI.
- Saporta, G. (1996), L'analyse harmonique qualitative, une synthèse de la théorie, *in* 'Seminario de capacitación e investigación. Recolección y análisis de datos longitudinales', pp. 111–120.
- Subdirección General de Calidad y Cohesión, Consejería de Sanidad de las CCAA & Ministerio de Sanidad Servicios Sociales e Igualdad (2015), Documento marco sobre enfermedad renal crónica (ERC) dentro de la estrategia de abordaje a la cronicidad en el SNS, Technical report, Ministerio de Sanidad Servicios Sociales e Igualdad.
- Tenenhuis, M. & Young, F. W. (1985), 'An analysis and synthesis of multiple correspondence analysis, optimal scaling, dual scaling, homogeneity analysis and other methods for quantifying categorical multivariate data', *Psychometrika* **50**(1), 91–119.
- Wickham, H. (2016), *ggplot2: elegant graphics for data analysis*, Springer.
- Wouters, O. J., O'Donoghue, D. J., Ritchie, J., Kanavos, P. G. & Narva, A. S. (2015), 'Early chronic kidney disease: diagnosis, management and models of care', *Nature Reviews Nephrology* **11**(8), 491–502.
- Zhong, Y., Muñoz, A., Schwartz, G. J., Warady, B. A., Furth, S. L. & Abraham, A. G. (2014), 'Nonlinear trajectory of GFR in children before RRT', *Journal of the American Society of Nephrology* **25**(5), 913–917.

APÉNDICE A

Aspectos éticos y de confidencialidad del Estudio

Dado que el análisis estuvo inmerso en un estudio de tipo observacional, no hubo en ningún momento intervención directa sobre los individuos, y por tanto, se consideró este trabajo como sin riesgo para ellos de acuerdo a la Resolución 8430 del *Ministerio de Salud de Colombia*. La identidad de los sujetos fue protegida en todo momento dado que la base de datos se codificó para garantizar confidencialidad de los mismos.

Además, esta propuesta fue presentada a un comité de ética en investigación institucional para que evaluara su viabilidad y pertinencia y el investigador condujo el estudio de acuerdo con las buenas prácticas clínicas en investigación y la declaración de Helsinki (Manzini 2000). Finalmente, se realizó un convenio específico entre la Universidad Nacional de Colombia y la entidad que financió este proyecto y suministró la información, de tal modo, que se aseguraron los compromisos éticos y de confidencialidad.

APÉNDICE B

Variables transversales y longitudinales ilustrativas de la ERC

indv	iniTFG	mediaTFG	finTFG	edad	diasaport	genero	escol	dxca	dxcv	causa	dxhta	dxdm	cegre
1	32.06	31.81	23.57	85.52	547	h	edubasic	no	si	hta	si	no	muerte
2	38.89	33.00	24.37	79.29	334	h	edusecu	no	no	dm	si	si	muerte
3	39.00	40.66	37.67	83.17	303	h	edusecu	no	no	hta	si	no	muerte
4	25.06	44.79	55.29	85.05	127	h	edusecu	no	no	hta	si	no	finest
5	44.48	49.68	51.35	79.89	1344	h	edusup	no	no	hta	si	no	finest
6	55.12	58.52	56.85	79.41	471	h	edusecu	no	no	hta	si	no	consext
7	32.64	52.37	56.23	84.86	1722	h	edubasic	no	no	hta	si	no	finest
8	49.66	51.63	44.79	81.55	1398	h	edubasic	no	no	dm	si	si	finest
9	47.66	47.82	46.56	92.06	1187	h	edubasic	no	no	otra	si	no	finest
10	57.51	56.99	52.48	87.79	312	h	edusecu	no	no	hta	si	no	abansusp

TABLA B.1. Tabla de variables transversales ilustrativas del AAC-AFM (primeros 10 individuos)

indv	tiempo	TFG	bun	hgb	tasis	tadias	creatin	imc
1	0	32.06	42.8	15.1	136	68	2.1	26.14
1	7	29.75	38.2	15.5	121	73	2.2	25.61
1	34	35.53	39.9	13.7	156	74	1.9	25.17
1	65	35.31	38.7	14.3	148	63	1.9	25.44
1	104	34.89	29.6	14.6	115	64	2	24.82
1	215	31.32	48	14.9	153	83	2.14	25.17
1	246	31.31	38.6	14.7	157	80	2.1	25
1	279	29.70	41.3	15.3	154	83	2.24	25
1	310	32.16	37.2	15.7	147	78	2.1	25.35
1	344	31.98	41.8	15.2	138	75	2.1	25
1	381	28.78	49.2	14.3	122	58	2.3	25.26
1	422	31.44	40	14.4	143	69	2.13	25.17
1	454	29.97	42	14.3	146	75	2.2	25.35
2	0	38.89	39	12.1	150	86	1.8	28.25
2	31	33.45	49	13.9	146	79	2.05	29.01
2	76	41.23	50.8	11.8	177	97	1.7	29.94
2	114	35.42	31	12.4	113	62	2	29.52
2	173	32.68	30.18	12.3	120	80	2.1	29.68
2	198	24.96	54.5	14.9	110	70	2.6	30.78
2	506	24.37	55	13.9	190	110	2.7	26.56

TABLA B.2. Tabla de variables longitudinales ilustrativas del AAC-AFM (primeros 2 individuos)

Variable	Definición operacional	Categorías	Tipo	Escala
dxcv	Antecedentes de enfermedad cardiovascular.	Si: Con antecedentes	VTI	Nominal
		No: Sin antecedentes		
edad	Edad del paciente al inicio del estudio.		VTI	Discreta
genero	Género definido al momento del ingreso al estudio.	m: mujer	VTI	Nominal
		h: hombre		
iniTFG	TFG reportada en la primera consulta en $ml/min/1.73m^2$		VTI	Intervalo
mediaTFG	Media de las TFG repetidas reportadas para el i -ésimo individuo en $ml/min/1.73m^2$		VTI	Intervalo
finTFG	TFG reportada en la última consulta en $ml/min/1.73m^2$		VTI	Intervalo
diasport	Días de permanencia del individuo dentro del estudio		VTI	Discreta
escol	Nivel educativo registrado al ingreso de estudio.	ninguna: ningún tipo de formación	VTI	Ordinal
		mprima: primaria no culminada		
		edubasic: básica primaria culminada		
		edusecu: básica secundaria culminada		
		edusup: con algún tipo de educación superior		
causa	Diagnóstico causal de la ERC.	desc: causa desconocida no diagnosticada	VTI	Nominal
		dm: diabetes mellitus		
		glo: glomerulonefritis		
		hta: hipertensión arterial		
		otra: otras causas diagnosticadas		
dxdm	Diagnóstico de diabetes mellitus al ingreso al estudio.	si: diagnóstico positivo	VTI	Nominal
		no: diagnóstico negativo		
dxhta	Diagnóstico de hipertensión al ingreso al estudio.	si: diagnóstico positivo	VTI	Nominal
		no: diagnóstico negativo		
dxca	Diagnóstico de cáncer no avanzado al ingreso al estudio	si: diagnóstico positivo	VTI	Nominal
		no: diagnóstico negativo		
cegre	Causa de egreso del estudio.	abansusp: abandono o suspensión del tratamiento	VTI	Nominal
		cambaseg: cambio de aseguradora o prestadora		
		consex: cambio a consulta externa		
		cuidpali: cuidados paliativos		
		finest: final del estudio		
		muer: muerte		
		perdseg: pérdida de seguimiento		
		trasplante: trasplante		
		dialisis: ingreso a diálisis		
TFG	Medidas repetidas de la TFG en $ml/min/1.73m^2$		VLI	Intervalo
bun	Medidas repetidas del nitrógeno ureico en la sangre en mg/dl		VLI	Intervalo
hgb	Medidas repetidas de la hemoglobina en g/dl		VLI	Intervalo
tasis	Medidas repetidas de la tensión arterial sistólica en $mmHg$		VLI	Intervalo
tadias	Medidas repetidas de la tensión arterial diastólica en $mmHg$		VLI	Intervalo
creatin	Medidas repetidas de la creatinina estimada en g/dl		VLI	Intervalo
imc	Medidas repetidas del índice de masa corporal en kg/m^2		VLI	Intervalo

TABLA B.3. Definición operativa de las variables ilustrativas del AAC-AFM

B.1. Análisis descriptivo de las VTI de la población con ERC

En esta sección se desarrolla un análisis descriptivo global de las variables transversales ilustrativas (VTI) utilizadas para caracterizar las clases de progresión de la ERC. Una primera parte se centra en el análisis de la tabla B.4, que muestra el modo como las distintas categorías de las VTI cualitativas se distribuyen a lo largo de la población general con ERC. En una segunda parte, la descripción se centra en la distribución de las VTI cuantitativas (tabla B.5 y figura B.1), mientras que en una tercera y última parte, la descripción se enfoca en el manera como dichas variables se comportan a través de las distintas categorías de las VTI categóricas.

B.2. Variables categóricas transversales ilustrativas

Hombres y mujeres son reportados aproximadamente con la misma frecuencia dentro de los individuos estudiados (47.4% y 52.6% respectivamente), mientras que por el nivel de escolaridad, la mayor proporción la presentan aquellos individuos que finalizaron la educación secundaria (60.5%), seguida de aquellos que finalizaron la educación básica (23%), tuvieron algún tipo de educación superior (9.7%), no terminaron la primaria (6.3%), y finalmente, no realizaron ningún tipo de estudios (0.5%).

Las principales causas de la ERC son la tensión arterial alta (HTA) y la diabetes, condiciones que en conjunto representan un poco menos del 80 por ciento de la población total. Este resultado concuerda con los distintos hallazgos reportados en la literatura especializada (ver capítulo 1, sección 1.2), que presentan estas afecciones como los factores de susceptibilidad y de inicio de la ERC más frecuentes. Otras causas de la ERC que aparecen en menor proporción son el conjunto de enfermedades distintas a la HTA, la diabetes y la glomerulonefritis (17.6%), la propia glomerulonefritis (3.3%) y por último, cuando esta causa es desconocida (1.6%).

Por otro lado, la mayoría de los individuos observados fueron diagnosticados negativamente con algún tipo de cáncer, no presentaban antecedentes de enfermedades cardiovasculares ni padecían de diabetes mellitus (situación favorable pues estas condiciones se consideran factores de riesgo de la ERC), aunque si hipertensos (85.3%).

Variable	Categoría	Frecuencia	Porcentaje
genero	masculino	1160	47.4
	femenino	1285	52.6
escol	edubasic	563	23.0
	edusecu	1479	60.5
	edusup	238	9.7
	mprima	153	6.3
	ninguna	12	0.5
dxca	no	2416	98.8
	si	29	1.2
dxcv	no	2344	95.9
	si	101	4.1
causa	dm	439	18
	glo	81	3.3
	hta	1455	59.5
	otra	431	17.6
	desconocida	39	1.6
dxhta	no	360	14.7
	si	2085	85.3
dxdm	no	1827	74.7
	si	618	25.3
cegre	abansusp	68	2.8
	cambaseg	84	3.4
	consext	276	11.3
	cuidpali	13	0.5
	dialisis	91	3.7
	finest	1724	70.5
	muerte	73	3
	perdseg	114	4.7
trasplante	2	0.1	

TABLA B.4. Síntesis de la distribución de las categorías de las VTI categóricas

En cuanto a las razones por las cuales egresaron del estudio, la mayor frecuencia la presentan aquellos individuos que fueron censurados por concluir el mismo, es decir, sobrevivieron al estudio (70.5%). Otra proporción importante la constituyen los individuos que fueron remitidos a consulta externa (el 11.3% de la población), lo cual es un hecho positivo, pues este tipo de remisiones generalmente se realizan cuando el individuo presenta una progresión estable y positiva a través de la ERC. El 3% de la población falleció en el transcurso del periodo de estudio, mientras que 3.7% y el 0.1% fueron incluidos en diálisis y trasplante de riñón respectivamente, información que habla positivamente del programa de prevención y seguimiento de la ERC, por cuanto el principal objetivo del mismo es retrasar en mayor medida posible la llegada de los individuos a este tipo de procedimientos.

B.3. Variables cuantitativas transversales ilustrativas

La descripción de las VTI cuantitativas se hace principalmente en términos de la mediana y el rango intercuartílico, y no a través de la media y la desviación estándar, si bien, éstas últimas también son reportadas en la tabla B.5. Las dos razones para realizar esto son: primero, que en el contexto clínico, en el que está inmerso este trabajo, la descripción de este tipo de variables usualmente se hace a través de estos dos estadísticos y segundo, que las distribuciones de las VTI clínicas no parecen distribuirse de manera normal.

Como se ya se ha indicado, la tabla B.5 presenta las principales medidas descriptivas de las VTI cuantitativas, mientras que los histogramas (acompañados de sus correspondientes diagramas de caja en la parte superior) de la figura B.1 dan cuenta de su distribución aproximada en la población.

Variables	Mínimo	1er. cuartil	Mediana	Media	3er. cuartil	Máximo	Desv. estándar
iniTFG	15.07	40.01	47.56	45.95	53.93	60.00	9.74
mediaTFG	12.90	40.66	48.57	46.84	54.43	83.32	10.94
finTFG	4.71	40.39	49.51	47.48	56.74	99.45	14.01
edad	21.71	63.75	71.75	70.23	78.27	96.92	11.57

TABLA B.5. Medidas descriptivas de las VTI cuantitativas

De manera general, los individuos entraron al estudio con una mediana de la TFG inicial (TFG estimada en la primera consulta) de 47.56 (en $ml/min/1.73m^2$), que según la tabla 1.1 indica estadio *E3a*, es decir, TFG ligera a moderadamente disminuida. La distribución de estas TFG¹ (cuyo valores mínimos y máximos son 15.07 y 60 respectivamente) presentan un sesgo negativo marcado, es decir, las mayores frecuencias las presentan TFG iniciales altas (figura B.1), tanto es así, que el 50% de la población reportó tasas entre 40.6 y 54.93. Aunque distintos valores atípicos menores a $20ml/min/1.73m^2$ son observados, la distribución de las TFG iniciales indican que los individuos ingresaron al estudio en unas condiciones tolerables en su ERC.

La TFG promedio (promedio de las distintas medidas repetidas por consulta de la TFG para cada individuo) presenta una mediana de 48.57 (estadio *E3a*). El rango de esta tasa varía entre los 12.9 y los 83 (mayor que la TFG inicial) y su distribución es algo sesgada negativamente. El 50% de la población tiene TFG promedio entre 40.66 y 54.43 (también mayor que en TFG inicial). Valores atípicos de tasas muy pequeñas o muy grandes son

¹La forma del histograma obedece a la inclusión de individuos en estadios inferiores a *E2*, es decir, con TFG estimada menor a 60 al iniciar el estudio.

observados, aunque esto es posible debido a la existencia de individuos con progresiones muy negativas en su ERC e individuos, que al contrario, mostraron mejoría.

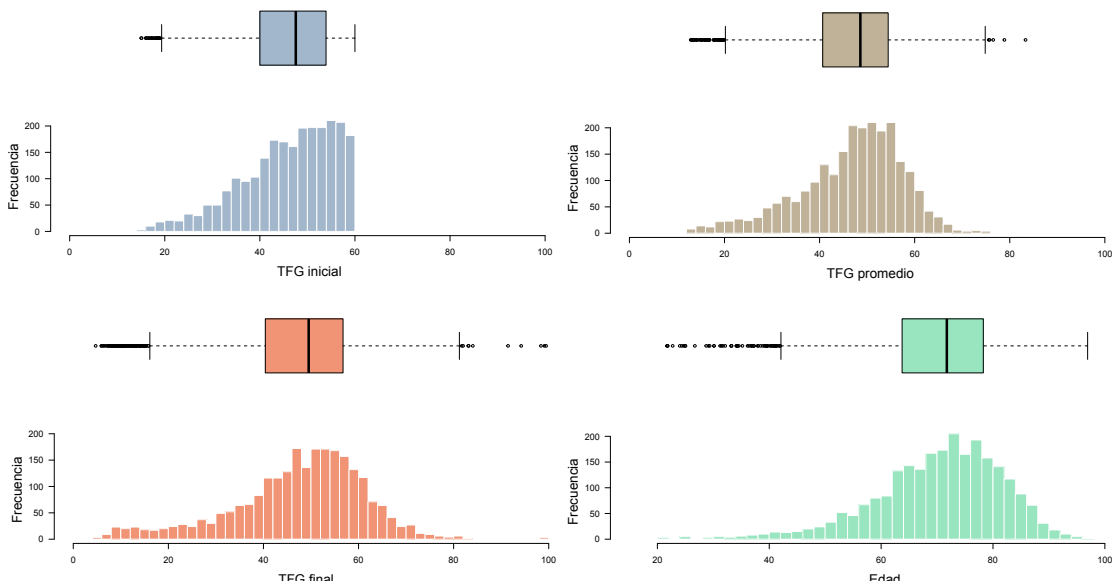


FIGURA B.1. Histogramas de frecuencias y diagramas de caja de las VTI cuantitativas

La TFG final (TFG estimada reportada en la última consulta) muestra, al igual que la TFG promedio, una distribución ligeramente asimétrica negativa. Se reporta una mediana de esta tasa de 49.51 (mayor un poco a las TFG inicial y promedio) y un rango intercuartílico que varía entre 40.39 y 99.45 (muy cercano al de las otras tasas). Se observan también valores extremos, con tasas muy pequeñas (por ejemplo 4.71) o tasas muy grandes (99.45).

Al comparar las distribuciones de las tres tasas, incluidas sus medianas (que están muy cercanas) y sus rangos intercuartílicos (que varían muy poco de uno a otro), se puede afirmar que la intervención del programa de prevención y seguimiento en el que fueron incluidos los individuos con ERC, cumplió en gran medida el objetivo de mantener las TFG en rangos aceptables, controlando de este modo la progresión de la enfermedad.

Finalmente, la mediana de la edad fue de 71.75 años, es decir individuos muy longevos, aunque también se reportaron individuos menores a 40 años (considerados atípicos), tanto así, que se identificó la presencia de un individuo de 21 años.

B.4. VTI cuantitativas por categorías de las VTI categóricas

La distribución de las VTI cuantitativas a través de la distintas categorías de las VTI categóricas se estudian por medio de diagramas de cajas. Por ejemplo, en la figura B.2 las distribuciones de las edades de los individuos se contrastan para las categorías de la HTA, los antecedentes de enfermedad cardiovascular, el cáncer, el género y la diabetes.

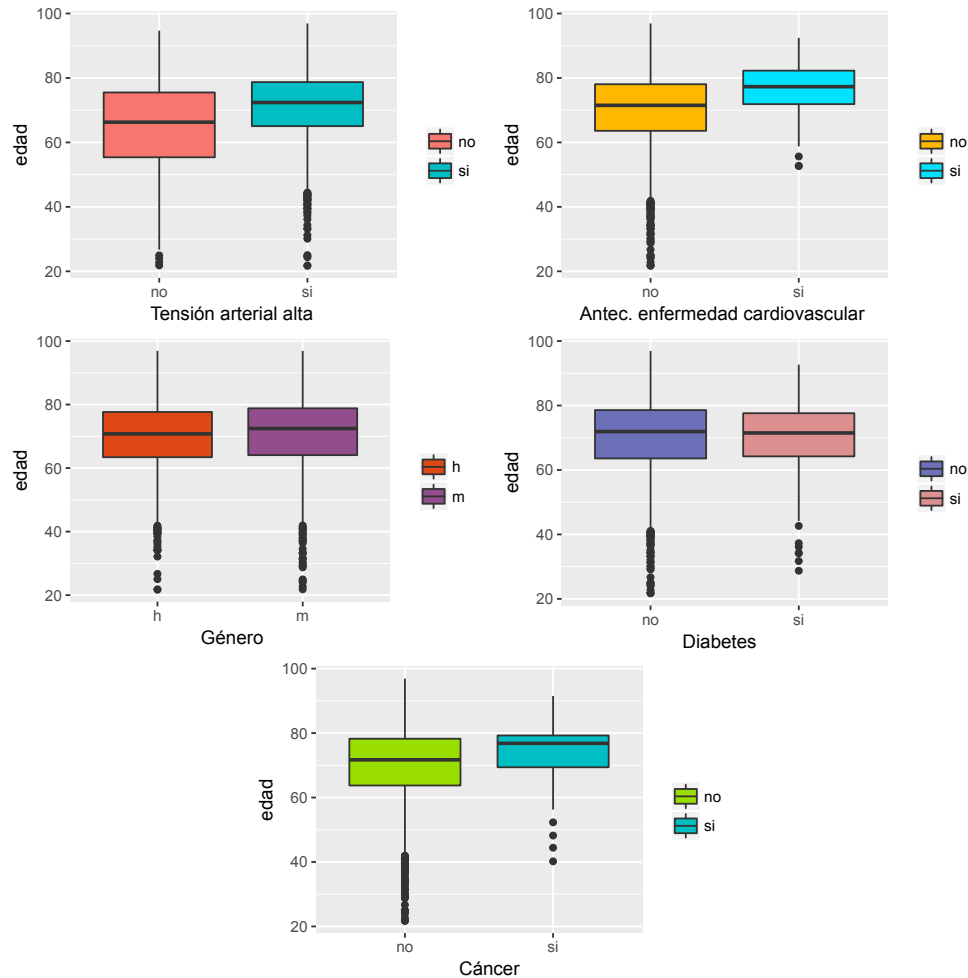


FIGURA B.2. Variable edad por categorías de las VTI categóricas (I)

El gráfico B.2 muestra que las medianas de las edades para todos los casos oscilan entre los 70 y 75 años (individuos hipertensos, con cáncer y con antecedentes de enfermedad cardiovascular son ligeramente mayores que sus complementos). Los hipertensos diagnosticados con cáncer y con antecedentes cardiovasculares son mayores en promedio que aquellos que no presentan estas condiciones. El género y el diagnóstico de diabetes, no muestran mucha diferencia en relación con la manera como se distribuyen las edades.

Resulta interesante, por otro lado, que muchos de los individuos menores de 40 años en distintas categorías se consideren atípicos.

En la figura B.3 se comparan las edades a través de las distintas causas de egreso del estudio, las causas de la ERC y el nivel de escolaridad.

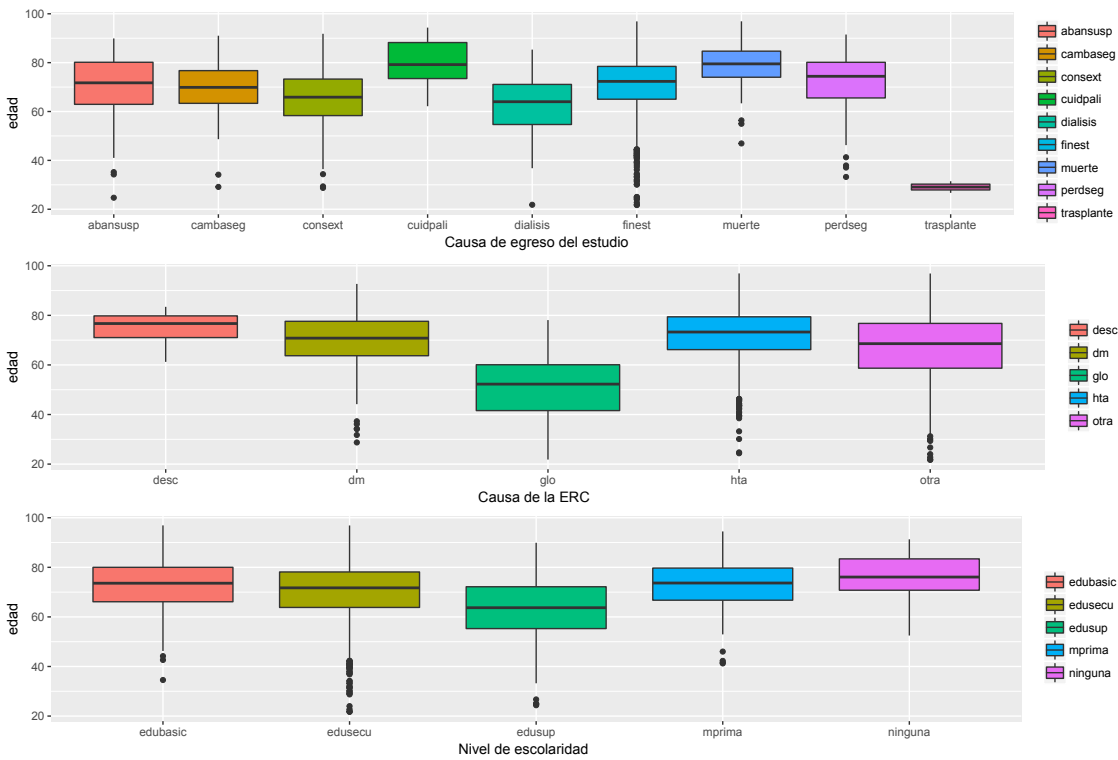


FIGURA B.3. Variable edad por categorías de las VTI categóricas (II)

Individuos remitidos a cuidados paliativos y fallecidos presentan mayor edad (alrededor de los 80 años) que los individuos que egresaron del estudio por otras razones. Quienes fueron remitidos a diálisis presentan una mediana de edad aproximada a los 60 años (menor a las otras causas de egreso, salvo trasplante). Los 2 individuos que fueron remitidos a trasplante son muy jóvenes si se compara con la media poblacional. Se destacan también, aquellos individuos que fallecieron durante el periodo de estudio, pues la mediana de su edad es relativamente alta, lo cual puede indicar que este desenlace obedeció a factores distintos a la ERC propiamente dicha. Por último, se observa que los individuos cuya causa de censura del estudio fue haber sobrevivido al mismo, muestran un número importante de individuos menores de 45 años considerados atípicos.

En cuanto a las causas de la ERC, los hipertensos, diabéticos y cuya causa de la ERC es una enfermedad desconocida son los de mayor edad (mediana aproximada a los 70 años), mientras que los más jóvenes son aquellos cuya causa es la glomerulonefritis (mediana cercana a los 50 años).

Por el nivel de escolaridad, la menor edad la presentan los individuos con estudios superiores (mediana de la edad cercana a los 60 años), mientras que las otras categorías (muy cercanas entre sí) tienen medianas de 70 años. Cabe anotar también, el reducido número de individuos menores a 40 años entre aquellos que terminaron la educación básica, no culminaron la primaria o no tuvieron ningún tipo de formación (incluso en esta categoría no hay individuos menores de 50 años).

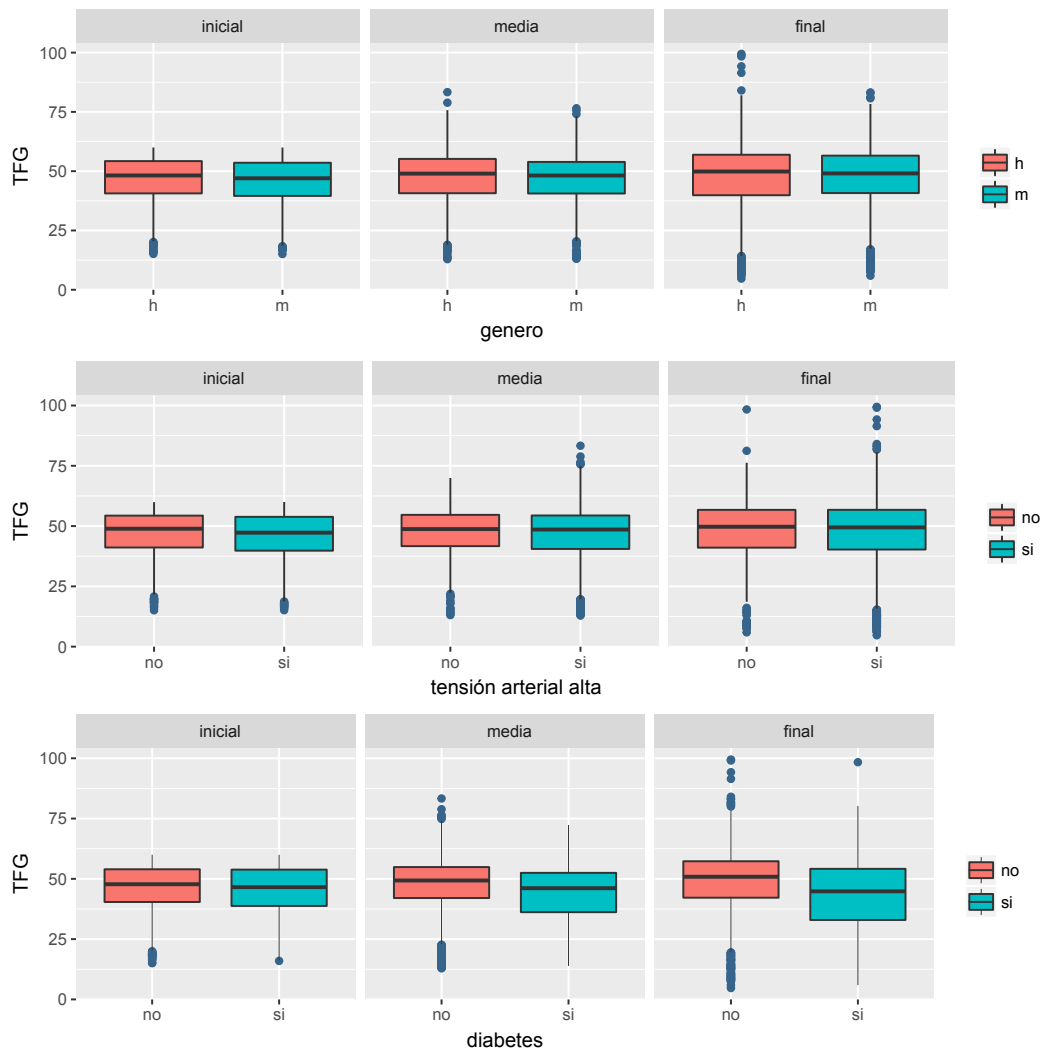


FIGURA B.4. TFG inicial, media y final por género, HTA y diabetes

Por otro lado, la comparación de las TFG inicial, promedio y final a través de las distintas categorías del género y los diagnósticos de HTA y diabetes se describen en los correspondientes diagramas de caja de la figura B.4.

Si se comparan las tres tasas (la inicial, promedio y final) entre hombres y mujeres, éstas muestran distribuciones muy similares con medianas cercanas a 50, lo cual puede significar que el género no es un factor determinante en la progresión de la TFG. Una situación similar se presenta en la variable HTA, donde las medianas de cada una de las tasas (todas ellas cercanas a 50) no varían mucho una con respecto a las otras, si bien, se alcanza a observar que los hipertensos presentan tasas levemente inferiores a los no hipertensos. En cuanto al caso de diabéticos y no diabéticos, es evidente el descenso que se observa desde la tasa de ingreso hasta la tasa final en los diagnosticados con diabetes; los no diabéticos, con medianas de las tasas cercanas a 50 en cada uno de los casos, presentan en general, tasas mayores que los diabéticos. Este último hecho puede indicar que la diabetes es un factor de riesgo de progresión de la ERC dentro de esta población.

La figura B.5 muestra como se distribuyen las tasas en términos de las distintas causas de egreso del estudio. De una manera general, se observa que la configuración de los diagramas de caja para las tres tasas (inicial, promedio y final) a través de estas causas es prácticamente la misma: las menores tasas las registran quienes son remitidos a diálisis y a trasplante, seguidos de quienes egresaron por desenlace fatal, mientras que las mayores tasas las observan quienes fueron censurados por causas no clínicas, es decir, salieron a consulta externa, se perdió su seguimiento o cambiaron de aseguradora. Esta relativa constancia en las tres tasas puede indicar un adecuado control de la enfermedad, por parte del programa de prevención, donde fueron incluidos los individuos. Pese a esta intervención, los individuos que egresaron por ser incluidos en diálisis muestran un descenso evidente de la tasa inicial a la tasa final: tasa inicial cercana a 38 y sesgo negativo (alta frecuencia de tasas iniciales menores a 35); tasa promedio cercana 22, pero con un sesgo positivo y, una tasa final próxima a 15 con la presencia de un número importante de individuos atípicos. Quienes finalizaron el estudio presentan tasas relativamente estables, aunque resulta interesante observar que las tasas iniciales presentan mayor variabilidad si se comparan con la tasa promedio y la tasa final. Esta disminución de la variabilidad también es reflejo del control por parte del programa de intervención.

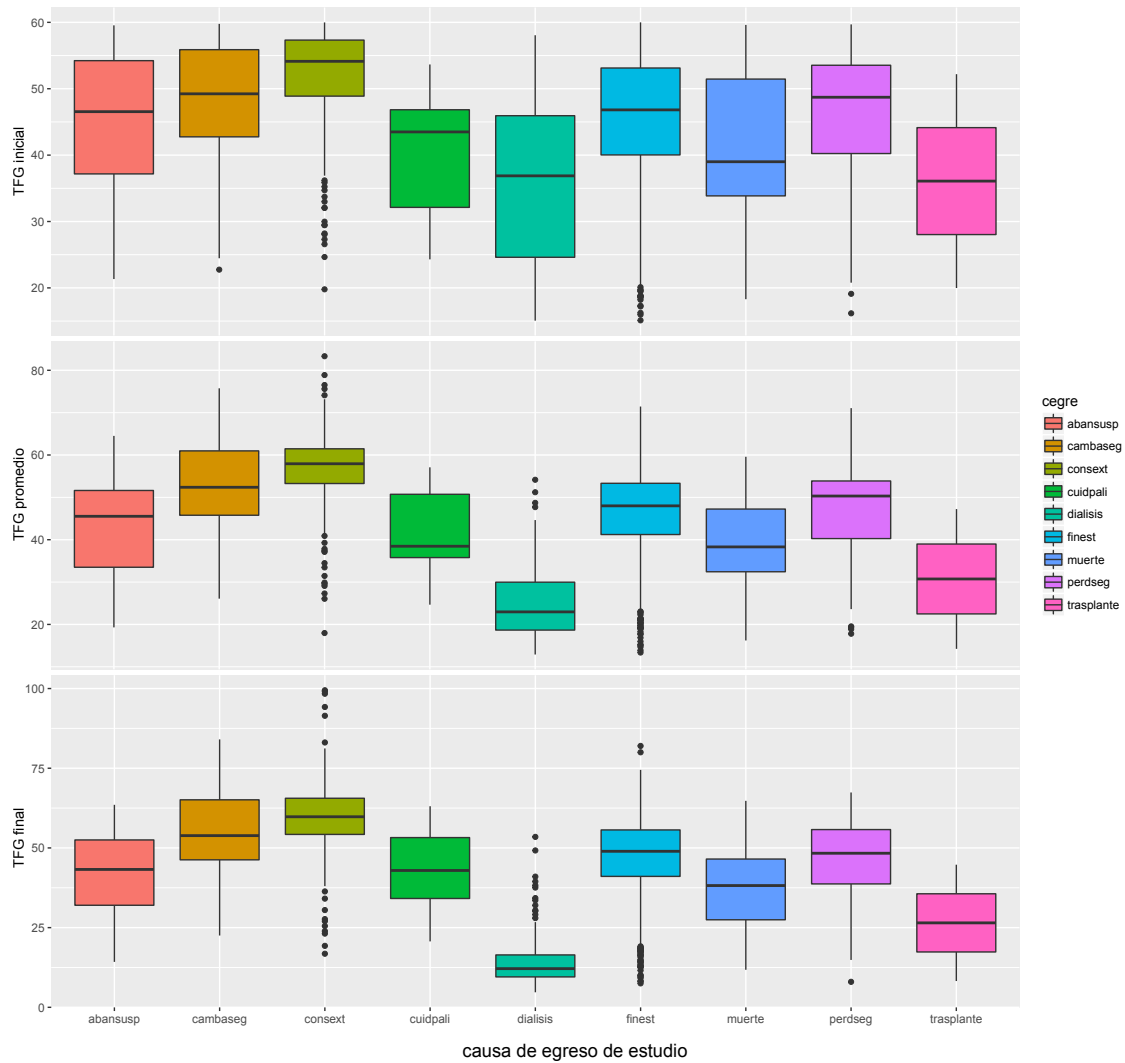


FIGURA B.5. TFG inicial, media y final por causa de egreso del estudio

Para concluir esta sección, los boxplots de la figura B.6 muestran las TFG inicial, promedio y final en relación a las distintas causas de la ERC. Las tasas promedio y final muestran distribuciones muy parecidas para cada una de las causas de la ERC. También se puede observar, que para dichas tasas, sus medianas son más o menos constantes (entre 45 y 50), lo que da indicios de un buen control del programa de prevención sobre la progresión de la ERC. Las TFG iniciales muestran una distribución distinta, con respecto a las otras dos, debido muy seguramente al criterio de sólo incluir individuos con tasas iguales o menores a 60 al principio del estudio, aunque también presenta medianas cercanas a los otros dos tipos de tasas.

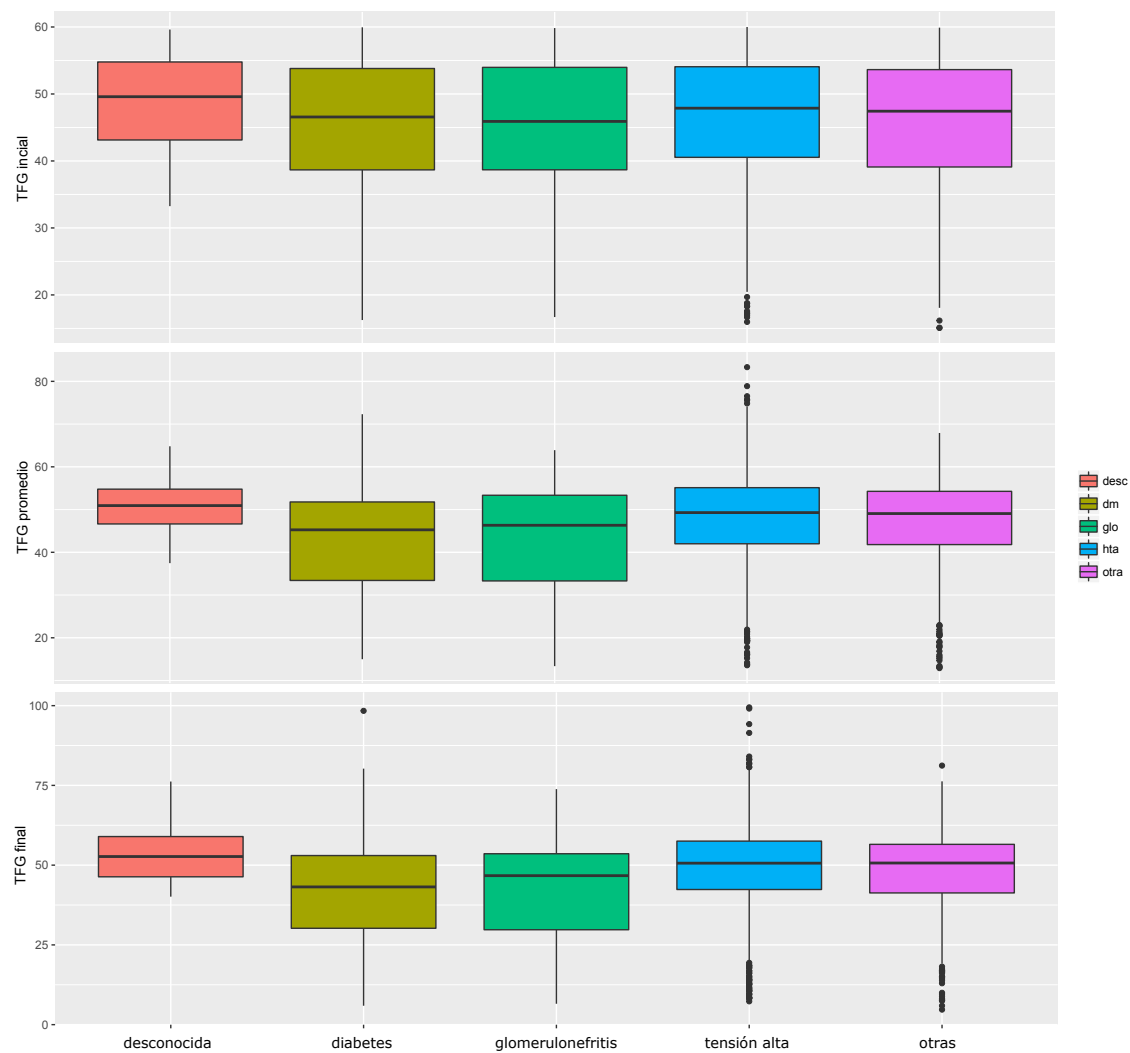


FIGURA B.6. TFG inicial, media y final por causas de la ERC

Cuando la causa de la ERC es la diabetes o es la glomerulonefritis, las tasas son siempre menores (en cada uno de los tres casos) que las tasas registradas para las otras causas, lo que indica que estas dos condiciones son factores de riesgo de aceleración de la ERC.

APÉNDICE C

Permanencias y transiciones entre estadios de la ERC y censuras

Estadios y censuras	Frecuencia	Porcentaje
CA	1724	17.15
CB	628	5.15
E2	1314	10.29
E3a	4160	7.53
E3b	3142	32.11
E4	965	22.43
E5	170	4.11
TRR	93	1.23
Total	12196	100

TABLA C.1. Frecuencias de permanencia en estadios de la ERC y censuras

Transición	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	Total
E2-CA	0	0	0	5	18	17	11	22	5	17	12	22	22	46	197
E2-CB	0	0	0	2	14	12	18	14	29	20	18	21	26	17	191
E2-E3a	4	56	83	80	81	76	84	76	84	65	81	46	58	16	890
E2-E3b	1	5	5	2	4	2	0	2	2	3	2	5	1	0	34
E3a-CA	0	0	0	31	71	62	78	63	50	84	55	100	123	208	925
E3a-CB	0	0	0	8	16	18	17	19	19	31	29	36	30	27	250
E3a-E2	198	123	115	123	77	91	75	81	78	84	73	64	50	26	1258
E3a-E3b	145	151	164	141	132	126	120	123	155	103	126	101	82	30	1699
E3a-E4	1	5	3	1	5	1	5	1	0	1	0	1	2	0	26
E3b-CA	0	0	0	11	27	26	21	28	22	45	43	44	65	134	466
E3b-CB	0	0	0	2	12	10	14	10	6	5	14	14	21	20	128
E3b-TRR	0	0	0	0	1	1	1	0	2	2	1	1	0	1	10
E3b-E2	11	8	6	3	3	5	2	1	0	1	3	4	1	0	48
E3b-E3a	193	207	174	149	118	133	117	132	132	120	106	109	94	32	1816
E3b-E4	63	57	62	57	45	40	53	47	48	40	52	44	38	27	673
E4-CA	0	0	0	5	7	4	3	8	7	7	7	16	17	37	118
E4-CB	0	0	0	1	4	4	4	5	3	8	3	4	5	12	53
E4-TRR	0	0	0	0	1	2	3	2	4	0	0	1	0	3	16
E4-E2	5	0	1	0	0	2	0	0	0	0	0	0	0	0	8
E4-E3a	7	4	3	1	0	1	0	0	0	0	0	0	1	0	17
E4-E3b	66	65	56	48	39	30	48	40	30	40	36	36	33	18	585
E4-E5	2	9	14	18	9	14	10	12	13	13	20	14	7	13	168
E5-CA	0	0	0	0	2	0	1	0	1	2	2	1	2	7	18
E5-CB	0	0	0	0	1	0	1	1	0	0	1	1	1	0	6
E5-TRR	0	0	0	0	5	7	7	9	5	1	10	7	6	8	65
E5-E4	2	4	7	7	2	12	3	6	4	7	9	6	5	6	80
Total	698	694	693	695	694	696	696	702	699	699	703	698	690	688	9745

TABLA C.2. Transiciones entre estadios y censuras por los catorce periodos de corte del tiempo de estudio

APÉNDICE D

Primeros planos factoriales del AAC y del AAC-AFM

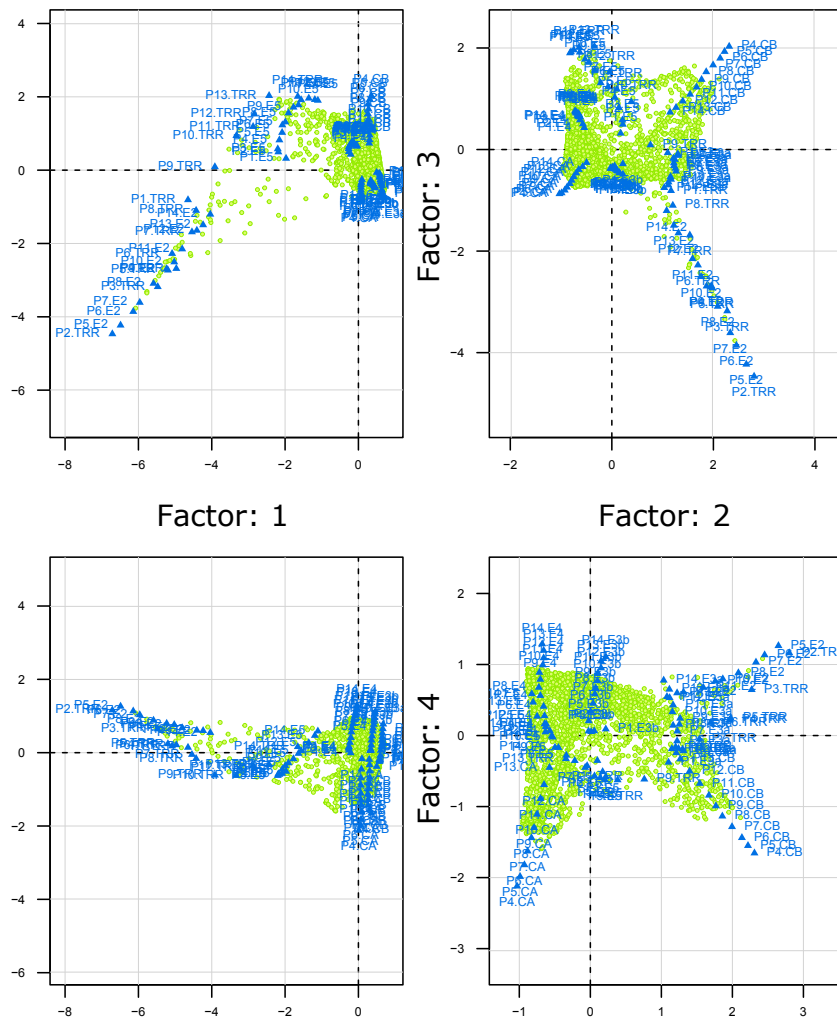


FIGURA D.1. Primeros planos factoriales del AAC

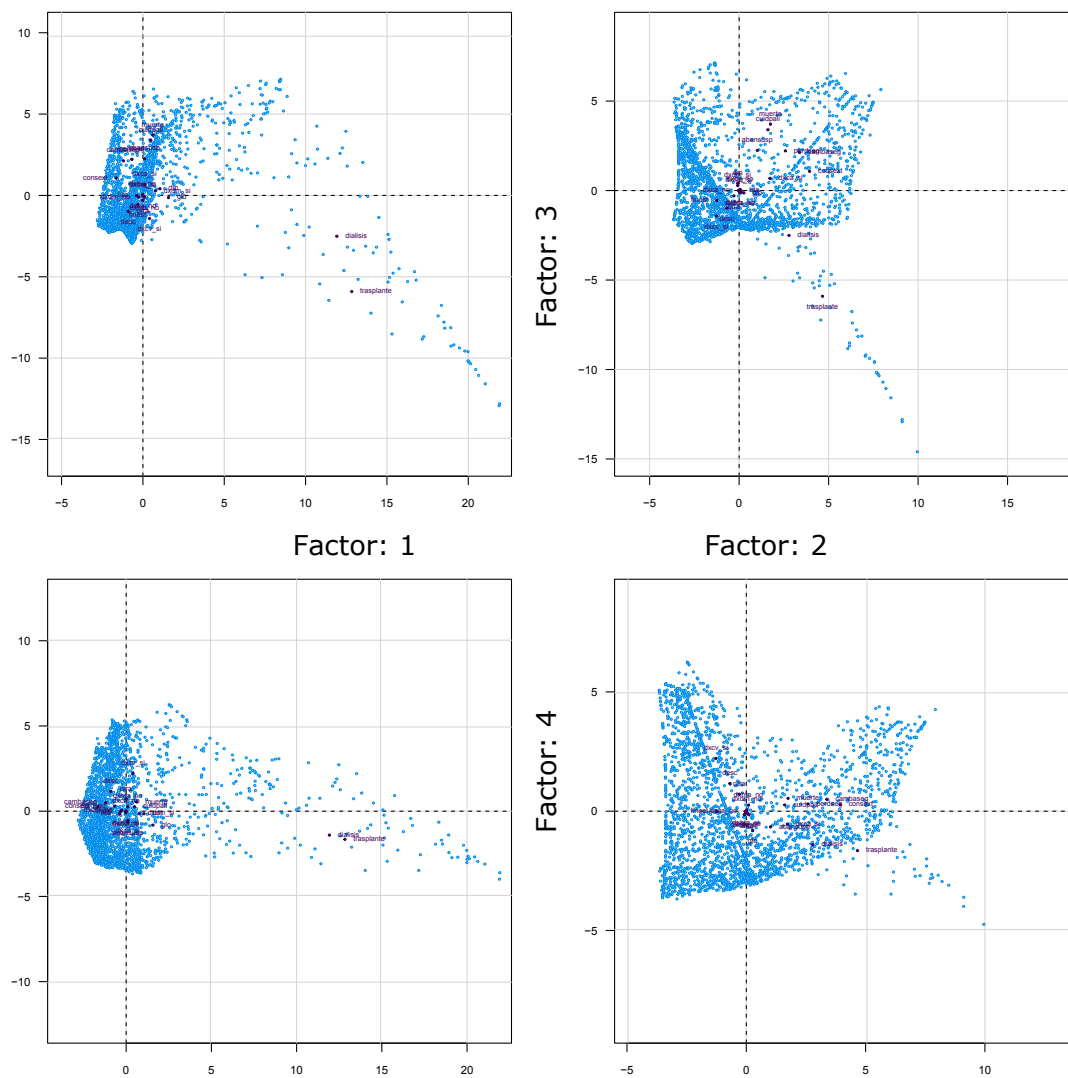


FIGURA D.2. Primeros planos factoriales del AAC-AFM con individuos y centroides de las categorías de las VTI categóricas proyectados

APÉNDICE E

Caracterización de las clases de progresión a partir de la VLC y las VTI

Periodo/estadio	V.test	Mean.Class	Global
P1.E4	-3.4	0.232	0.324
P2.E3a	6.2	0.203	0.098
P2.E4	-3.6	0.210	0.305
P3.E3a	6.5	0.238	0.118
P3.E4	-3.8	0.191	0.291
P4.E3a	5.9	0.253	0.136
P4.E3b	-2.8	0.427	0.507
P4.E4	-2.2	0.224	0.282
P5.CB	20.9	0.150	0.014
P5.E3a	4.6	0.230	0.137
P5.E3b	-4.1	0.345	0.464
P5.E4	-2.9	0.192	0.266
P6.CB	30.0	0.353	0.033
P6.E3b	-5.4	0.271	0.424
P6.E4	-3.6	0.159	0.251
P7.CB	37.9	0.569	0.054
P7.E3b	-6.6	0.201	0.388
P8.CB	44.6	0.792	0.075
P8.E3b	-9.1	0.105	0.354
P9.CB	47.9	0.958	0.096
P10.CB	44.4	0.998	0.120
P11.CB	39.4	1.000	0.146
P12.CB	35.3	1.000	0.176
P13.CB	31.6	1.000	0.209
P14.CB	28.8	1.000	0.240

TABLA E.1. Caracterización de la clase 1 según valores test de la VLC

Categoría	Test.Value	Class.Cat	Cat.Class	Global
cegre.consext	11.5	32.2	38.4	11.3
cegre.perdseg	10.4	45.6	22.4	4.7
cegre.cambaseg	8.3	42.9	15.5	3.4
cegre.muerte	7.9	43.8	13.8	3.0
cegre.cuidpali	2.8	38.5	2.2	0.5
cegre.abansusp	4.2	26.5	7.8	2.8
escol.edubasic	2.5	12.1	29.3	23.0
dxca.si	2.4	24.1	3.0	1.2
dxcv.no	2.2	9.7	98.3	95.9

TABLA E.2. Caracterización de la clase 1 según valores test de las VTI categóricas

Variable	V.test	Mean.Class	Global
IniTFG	2.325	47.365	45.95
MediaTFG	4.26	49.748	46.837
FinTFG	3.916	50.905	47.477
diasaort	-17.523	352.647	863.022

TABLA E.3. Caracterización de la clase 1 según valores test de las VTI cuantitativas

Periodo/estadio	V.test	Mean.Class	Global
P1.E3b	-3.4	0.510	0.568
P1.E4	3.8	0.386	0.324
P2.E4	2.8	0.349	0.305
P3.E3b	3.7	0.586	0.522
P4.E3b	4.6	0.585	0.507
P5.CA	22.4	0.204	0.046
P5.E4	-3.2	0.216	0.266
P6.CA	30.8	0.417	0.094
P6.E3b	-4.8	0.341	0.424
P6.E4	-6.8	0.148	0.251
P7.CA	38.3	0.625	0.141
P7.E3b	-9.6	0.226	0.388
P8.CA	45.3	0.840	0.191
P9.CA	47.4	0.965	0.231
P10.CA	44.2	0.999	0.277
P11.CA	38.1	1.000	0.338
P12.CA	33.6	1.000	0.399
P13.CA	28.6	1.000	0.483
P14.CA	23.0	1.000	0.604

TABLA E.4. Caracterización de la clase 2 según valores test de la VLC

Periodo/estadio	V.test	Mean.Class	Global
P1.E3b	18.0	0.816	0.568
P1.E4	-12.5	0.160	0.324
P2.E3b	21.2	0.823	0.530
P2.E4	-14.3	0.123	0.305
P3.E3b	23.1	0.842	0.522
P3.E4	-14.9	0.103	0.291
P4.E3b	25.0	0.853	0.507
P5.E3b	28.8	0.864	0.464
P6.E3b	31.4	0.857	0.424
P7.E3b	33.9	0.849	0.388
P8.E3b	37.2	0.847	0.354
P9.E3b	37.8	0.812	0.322
P9.E4	-6.1	0.129	0.195
P10.E3b	35.5	0.722	0.281
P10.E4	-4.5	0.128	0.175
P11.CA	-11.9	0.173	0.338
P11.E3b	30.4	0.588	0.234
P11.E4	-2.4	0.129	0.153
P12.CA	-8.5	0.277	0.399
P12.CB	-5.9	0.110	0.176
P12.E3b	25.5	0.451	0.183
P13.CA	-4.0	0.426	0.483
P13.CB	-4.8	0.151	0.209
P13.E3b	19.9	0.297	0.124
P14.CB	-3.8	0.192	0.240
P14.E3b	13.4	0.125	0.053

TABLA E.7. Caracterización de la clase 3 según valores test de la VLC

Categoría	Test.Value	Class.Cat	Cat.Class	Global
cegre.finest	21.1	32.1	99.1	70.5
escol.edusecu	17.1	33.5	89.7	60.5
causaERC.otra	6.6	35.0	27.3	17.6
causaERC.desc	3.0	43.6	3.1	1.6
dxca.no	3.4	22.9	100.0	98.8
dxcv.si	10.2	69.3	12.7	4.1
genero.m	2.6	24.6	57.1	52.6

TABLA E.5. Caracterización de la clase 2 según valores test de las VTI categóricas

Variable	V.test	Mean.Class	Global
IniTFG	-4.314	44.378	45.95
FinTFG	4.325	49.744	47.477
edad	2.16	71.168	70.233
diasaport	-30.363	333.383	863.022

TABLA E.6. Caracterización de la clase 2 según valores test de las VTI cuantitativas

Categoría	Test.Value	Class.Cat	Cat.Class	Global
cegre.finest	5.6	34.4	77.9	70.5
dxcv.no	4.6	32.0	98.4	95.9
causaERC.hta	3.7	33.9	64.8	59.5
escol.edusup	3.4	40.8	12.7	9.7
dxdm.no	3.2	32.8	78.7	74.7
escol.edubasic	3.0	36.1	26.7	23.0

TABLA E.8. Caracterización de la clase 3 según valores test de las VTI categóricas

Variable	V.test	Mean.Class	Global
diasaport	19.684	1139.154	863.022
IniTFG	14.376	50.164	45.95
MediaTFG	12.261	50.873	46.837
FinTFG	9.326	51.409	47.477

TABLA E.9. Caracterización de la clase 3 según valores test de las VTI cuantitativas

Periodo/estadio	V.test	Mean.Class	Global
P1.E2	13.1	0.145	0.146
P1.E3a	9.5	0.781	0.568
P2.E2	41.1	0.479	0.118
P2.E3a	4.8	0.459	0.330
P3.E2	22.7	0.615	0.105
P3.E3b	-4.3	0.357	0.522
P4.E2	26.5	0.704	0.124
P4.E3a	15.1	0.289	0.507
P5.E2	33.5	0.770	0.117
P5.E3a	8.3	0.221	0.164
P6.E2	25.2	0.764	0.138
P6.E3a	16.6	0.220	0.124
P7.E2	24.2	0.740	0.236
P7.E3a	5.8	0.257	0.258
P8.E2	37.0	0.704	0.144
P8.E3a	6.4	0.292	0.164
P9.E2	36.6	0.636	0.256
P10.E3a	33.4	0.549	0.717
P11.CB	3.0	0.208	0.146
P11.E3a	30.7	0.447	0.212
P12.CA	-10.4	0.108	0.399
P12.CB	5.0	0.286	0.176
P12.E2	26.2	0.342	0.061
P12.E3a	3.4	0.252	0.191
P13.CA	-10.2	0.195	0.483
P13.CB	7.8	0.392	0.209
P13.E2	22.4	0.230	0.022
P13.E3a	3.1	0.176	0.106
P14.CA	-9.6	0.343	0.604
P14.CB	9.4	0.473	0.240

TABLA E.10. Caracterización de la clase 4 según valores test de la VLC

Categoría	Test.Value	Class.Cat	Cat.Class	Global
cegre.consext	12.8	37.3	39.5	11.3
cegre.cambaseg	2.7	20.2	6.5	3.4
causaERC.hta	4.0	12.6	70.5	59.5
escol.mprima	2.4	16.3	9.6	6.3
escol.edubasic	3.1	14.2	30.7	23.0
escol.edusup	3.0	16.4	14.9	9.7
genero.h	2.8	12.4	55.2	47.4

TABLA E.11. Caracterización de la clase 4 según valores test de las VTI categóricas

Variable	V.test	Mean.Class	Global
IniTFG	12.864	53.283	45.95
MediaTFG	21.923	60.871	46.837
FinTFG	18.484	62.631	47.477
diasaport	9.457	1121.008	863.022

TABLA E.12. Caracterización de la clase 4 según valores test de las VTI cuantitativas

Periodo/estadio	V.test	Mean.Class	Global
P1.E3b	-16.9	0.251	0.568
P1.E4	20.8	0.695	0.324
P2.E3b	-15.5	0.238	0.530
P2.E4	23.6	0.716	0.305
P3.E3b	-17.2	0.199	0.522
P3.E4	27.4	0.762	0.291
P4.E3b	-17.5	0.177	0.507
P4.E4	29.3	0.784	0.282
P5.E3b	-15.7	0.167	0.464
P5.E4	31.9	0.803	0.266
P6.E3b	-14.4	0.153	0.424
P6.E4	34.2	0.818	0.251
P7.E3b	-14.0	0.130	0.388
P7.E4	37.5	0.824	0.228
P8.E3b	-12.1	0.137	0.354
P8.E4	39.5	0.810	0.207
P9.E3b	-9.6	0.152	0.322
P9.E4	37.0	0.749	0.195
P10.CA	-9.0	0.116	0.277
P10.E3b	-8.8	0.131	0.281
P10.E4	34.2	0.664	0.175
P11.CA	-5.8	0.228	0.338
P11.E3b	-7.3	0.118	0.234
P11.E4	29.4	0.545	0.153
P12.CA	-3.8	0.325	0.399
P12.E4	25.7	0.445	0.129
P13.CB	-6.2	0.107	0.209
P13.E4	20.0	0.294	0.089
P14.CA	2.1	0.642	0.604
P14.CB	-5.5	0.146	0.240
P14.E4	14.4	0.128	0.038

TABLA E.13. Caracterización de la clase 5 según valores test de la VLC

Categoría	Test.Value	Class.Cat	Cat.Class	Global
cegre.finest	5.8	22.5	80.8	70.5
cegre.muerte	2.8	32.9	5.0	3.0
dxcv.no	4.7	20.3	99.2	95.9
dxdm.si	2.9	23.5	30.2	25.3
escol.mprima	3.1	29.4	9.4	6.3
escol.edubasic	2.4	22.9	26.9	23.0
causaERC.dm	2.8	24.1	22.1	18.0
genero.m	2.1	21.0	56.2	52.6

TABLA E.14. Caracterización de la clase 5 según valores test de las VTI categóricas

Variable	V.test	Mean.Class	Global
diasaport	14.138	1132.786	863.022
edad	2.79	71.554	70.233
IniTFG	-11.901	41.205	45.95
MediaTFG	-18.054	38.753	46.837
FinTFG	-15.745	38.448	47.477

TABLA E.15. Caracterización de la clase 5 según valores test de las VTI cuantitativas

Periodo/estadio	V.test	Mean.Class	Global
P1.E4	2.1	0.411	0.324
P1.E5	19.6	0.515	0.071
P2.E5	24.1	0.588	0.065
P3.E5	25.2	0.613	0.065
P4.E5	28.3	0.655	0.062
P5.E5	31.9	0.747	0.062
P6.E5	35.1	0.766	0.055
P7.E5	37.2	0.817	0.057
P8.E4	-2.4	0.124	0.207
P8.E5	40.4	0.825	0.050
P9.E4	-2.6	0.104	0.195
P9.E5	39.1	0.771	0.046
P10.E5	34.7	0.651	0.043
P10.TRR	19.0	0.140	0.007
P11.CA	-4.5	0.140	0.338
P11.E5	31.2	0.525	0.035
P11.TRR	19.6	0.139	0.007
P12.E5	24.4	0.377	0.030
P12.TRR	16.9	0.108	0.005
P13.CA	-3.3	0.334	0.483
P13.CB	-2.7	0.106	0.209
P13.E5	18.4	0.268	0.025
P14.CA	-3.4	0.457	0.604
P14.CB	-2.3	0.150	0.240
P14.E5	12.8	0.128	0.014

TABLA E.16. Caracterización de la clase 6 según valores test de la VLC

Periodo/estadio	V.test	Mean.Class	Global
P1.E3b	-5.6	0.213	0.568
P1.E5	14.4	0.548	0.071
P2.E3b	-6.2	0.133	0.530
P2.E5	13.9	0.509	0.065
P2.TRR	16.5	0.060	0.001
P3.E3b	-6.4	0.114	0.522
P3.E5	14.1	0.513	0.065
P3.TRR	20.6	0.141	0.004
P4.E3b	-6.8	0.070	0.507
P4.E5	12.3	0.441	0.062
P4.TRR	25.6	0.301	0.009
P5.E5	10.4	0.388	0.062
P5.TRR	26.5	0.331	0.009
P6.E4	-2.3	0.121	0.251
P6.E5	7.9	0.290	0.055
P6.TRR	26.6	0.322	0.009
P7.E5	6.6	0.254	0.057
P7.TRR	22.7	0.244	0.007
P8.TRR	21.1	0.205	0.006
P9.TRR	8.8	0.081	0.005
P9.E5	5.6	0.346	0.046
P10.TRR	47.2	1.000	0.007
P11.TRR	44.6	1.000	0.007
P12.TRR	39.9	1.000	0.005
P13.TRR	38.6	1.000	0.006
P14.TRR	36.7	1.000	0.018

TABLA E.19. Caracterización de la clase 7 según valores test de la VLC

Categoría	Test.Value	Class.Cat	Cat.Class	Global
cegre.dialisis	8.0	28.6	24.3	3.7
causaERC.dm	4.5	8.7	35.5	18.0
dxdm.si	4.0	7.3	42.1	25.3
escol.edubasic	2.3	6.0	31.8	23.0

TABLA E.17. Caracterización de la clase 6 según valores test de las VTI categóricas

Variable	V.test	Mean.Class	Global
edad	-4.51	65.299	70.233
IniTFG	-16.045	31.171	45.95
MediaTFG	-22.108	23.967	46.837
FinTFG	-21.337	19.21	47.477

TABLA E.18. Caracterización de la clase 6 según valores test de las VTI cuantitativas

Categoría	Test.Value	Class.Cat	Cat.Class	Global
causaERC.dm	4.0	4.8	41.2	18.0
causaERC.glo	3.2	8.6	13.7	3.3
dxdm.si	3.7	4.0	49.0	25.3
cegre.dialisis	17.5	51.6	92.2	3.7
cegre.trasplante	3.5	100.0	3.9	0.1

TABLA E.20. Caracterización de la clase 7 según valores test de las VTI categóricas

Variable	V.test	Mean.Class	Global
diasaport	-6.952	413.858	863.022
edad	-7.431	58.318	70.233
IniTFG	-10.464	31.823	45.95
MediaTFG	-15.483	23.36	46.837
FinTFG	-16.538	15.365	47.477

TABLA E.21. Caracterización de la clase 7 según valores test de las VTI cuantitativas

Variable	Categoría	clase 1	clase 2	clase 3	clase 4	clase 5	clase 6	clase 7	Global
Género	masculino	50	42.9	48	55.2	43.8	55.1	56.9	47.4
	femenino	50	57.1	52	44.8	56.2	44.9	43.1	52.6
Nivel de escolaridad	ninguna	0.4	0.2	0.7	0	0.8	0.9	0	0.5
	mprima	8.2	1.3	5.9	9.6	9.4	8.4	5.9	6.3
	edubasic	29.3	6.5	26.7	30.7	26.9	31.8	25.5	23
	edusecu	53.4	89.7	54	44.8	51.9	49.5	56.9	60.5
	edusup	8.6	2.4	12.7	14.9	11	9.3	11.8	9.7
Diagnóstico de cáncer	no	97	100	99.2	98.1	98.1	100	96.1	98.8
	si	3	0	0.8	1.9	1.9	0	3.9	1.2
Antecedente cardiovascular	no	98.3	87.3	98.4	97.7	99.2	99.1	92.2	95.9
	si	1.7	12.7	1.6	2.3	0.8	0.9	7.8	4.1
Causa de la ERC	desc	1.7	3.1	1.8	0.4	0.6	0	0	1.6
	dm	18.1	15	14.6	14.6	22.1	35.5	41.2	18
	glo	2.2	2	3.7	1.9	4.2	4.7	13.7	3.3
	hta	60.3	52.6	64.8	70.5	60.2	39.3	31.4	59.5
	otra	17.7	27.3	15.1	12.6	12.9	20.6	13.7	17.6
Diagnóstico de HTA	no	16.8	16.3	13.9	11.9	13.3	20.6	15.7	14.7
	si	83.2	83.7	86.1	88.1	86.7	79.4	84.3	85.3
Diagnóstico de diabetes	no	75	77.4	78.7	77.8	69.8	57.9	51	74.7
	si	25	22.6	21.3	22.2	30.2	42.1	49	25.3
Causa de egreso de estudio	abansusp	7.8	0	3.2	1.1	3.8	4.7	0	2.8
	cambaseg	15.5	0	2.5	6.5	2.3	0.9	0	3.4
	consext	38.4	0	9.5	39.5	1.7	3.7	0	11.3
	cuidpali	2.2	0	0.3	0	1	0.9	0	0.5
	dialisis	0	0	0.8	0	2.5	24.3	92.2	3.7
	finest	0	100	77.9	48.3	80.8	57.9	3.9	70.5
	muerte	13.8	0	1.4	0.4	5	4.7	0	3
	perdseg	22.4	0	4.5	4.2	2.9	2.8	0	4.7
	trasplante	0	0	0	0	0	0	3.9	0.1

TABLA E.22. Síntesis de los perfiles de las siete clases de progresión de la ERC a través de las VTI categóricas

Clase	Variable	Mínimo	1er. Cuartil	Mediana	Media	3er. Cuartil	Máximo	Desv. estandar
1	IniTFG	18.30	41.89	49.94	47.37	55.94	59.78	10.16
	FinTFG	11.77	41.40	52.20	50.90	61.04	99.45	15.51
	MediaTFG	16.21	42.56	52.38	49.75	58.14	83.32	12.38
	edad	24.72	63.14	71.35	70.55	80.75	94.48	12.41
	diasaport	47.00	232.20	348.50	352.60	462.80	695.00	141.36
2	IniTFG	15.97	37.33	45.02	44.38	52.00	60.00	9.37
	FinTFG	7.51	44.28	51.57	49.74	56.78	81.99	10.76
	MediaTFG	13.89	42.23	48.77	47.04	53.65	62.64	8.60
	edad	21.71	64.81	72.55	71.17	79.10	96.89	11.41
	diasaport	122.00	216.00	323.00	333.40	439.00	706.00	140.50
3	IniTFG	19.11	46.40	50.94	50.16	55.07	59.93	6.36
	FinTFG	9.52	47.01	51.80	51.41	56.23	76.15	7.32
	MediaTFG	30.88	47.86	51.04	50.87	53.91	60.58	4.00
	edad	28.73	64.16	71.95	70.42	77.95	94.68	10.84
	diasaport	572.00	878.00	1117.00	1139.00	1371.00	1821.00	311.66
4	IniTFG	17.32	51.85	55.39	53.28	57.75	59.99	7.25
	FinTFG	39.70	57.81	61.89	62.63	67.34	99.09	8.07
	MediaTFG	49.31	58.30	60.27	60.87	62.84	75.75	4.26
	edad	39.51	63.58	69.72	69.35	75.76	91.55	9.19
	diasaport	493.00	833.00	1126.00	1121.00	1362.00	1813.00	341.73
5	IniTFG	20.56	36.82	41.60	41.21	45.34	59.55	6.99
	FinTFG	7.31	33.33	39.11	38.45	44.25	73.83	8.94
	MediaTFG	20.85	35.30	39.49	38.75	42.25	54.78	4.96
	edad	24.95	65.12	73.32	71.55	80.01	94.34	11.79
	diasaport	488.00	835.50	1130.00	1133.00	1380.00	1798.00	339.06
6	IniTFG	15.11	25.08	29.78	31.17	35.35	58.06	9.56
	FinTFG	4.71	12.89	19.38	19.21	24.70	38.53	7.78
	MediaTFG	12.90	20.69	24.56	23.97	27.22	33.79	4.42
	edad	24.01	56.60	65.64	65.30	74.39	96.92	13.93
	diasaport	441.00	861.70	1067.00	1118.00	1358.00	1805.00	335.49
7	IniTFG	15.07	19.85	28.53	31.82	41.23	58.06	13.01
	FinTFG	6.28	9.71	12.68	15.36	16.19	44.74	8.67
	MediaTFG	13.09	15.93	20.26	23.36	29.51	48.71	9.45
	edad	21.83	51.34	58.21	58.32	69.13	85.34	14.18
	diasaport	136.00	292.40	410.00	413.90	540.00	1098.00	180.43

TABLA E.23. Síntesis de los perfiles de las siete clases de progresión de la ERC a través de las VTI cuantitativas

APÉNDICE F

Permanencias y transiciones de las clases en estadios y censuras de la ERC

Clase	Estudio	E2	E3a	E3b	E4	E5	TRR	CA	CB
1	352.6	244.9	254.7	194.6	151.4	80.0	0.0	0.0	1199.2
2	333.4	135.2	251.5	178.7	135.5	188.6	0.0	1222.6	0.0
3	1139.2	137.6	477.0	143.0	98.7	74.5	465.8	415.5	514.7
4	1121.0	353.1	183.3	117.8	158.9	0.0	0.0	331.1	575.6
5	1132.8	126.4	161.2	409.7	107.2	86.3	498.6	444.5	463.8
6	1117.7	0.0	162.2	164.8	369.3	123.2	463.8	455.1	461.4
7	413.9	141.0	183.5	172.3	208.5	166.6	1156.7	636.5	0.0

TABLA F.1. Permanencias medias (en días) de las clases en estadios y censuras de la ERC y en el tiempo de estudio

Transición de	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	Total
E2-CB	0	0	0	2	14	12	18	13	6	0	0	0	0	0	65
E2-E3a	3	7	11	10	13	5	6	2	0	0	0	0	0	0	57
E3a-CB	0	0	0	8	16	18	17	19	13	1	0	0	0	0	92
E3a-E2	38	19	18	9	10	10	5	4	1	0	0	0	0	0	114
E3a-E3b	5	13	13	13	4	4	4	2	0	0	0	0	0	0	58
E3b-CB	0	0	0	2	12	10	14	10	4	1	0	0	0	0	53
E3b-E2	3	1	1	0	2	0	0	0	0	0	0	0	0	0	7
E3b-E3a	12	15	8	2	4	5	5	0	1	0	0	0	0	0	52
E3b-E4	8	7	6	9	8	2	4	0	0	0	0	0	0	0	44
E4-CB	0	0	0	1	4	4	4	5	1	0	0	0	0	0	19
E4-E3a	2	1	0	1	0	0	0	0	0	0	0	0	0	0	4
E4-E3b	7	5	6	6	9	2	3	0	0	0	0	0	0	0	38
E4-E5	0	0	1	3	1	0	2	0	0	0	0	0	0	0	7
E5-CB	0	0	0	0	1	0	1	1	0	0	0	0	0	0	3
E5-E4	0	0	0	3	0	1	0	0	0	0	0	0	0	0	4
Total	80	68	66	69	98	74	83	56	26	2	0	0	0	0	622

TABLA F.2. Frecuencias de transiciones entre estadios y censuras por periodo: clase 1

Transición de	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	Total
E2-E3a	0	1	1	1	0	0	1	0	0	0	0	0	1	0	5
E2-E3b	0	1	1	0	1	0	0	1	0	0	0	0	0	0	4
E3a-CA	0	0	0	0	0	0	0	0	7	13	10	17	16	20	83
E3a-CB	0	0	0	0	0	0	0	0	0	1	4	1	5	0	11
E3a-E2	3	1	1	0	0	0	1	0	1	0	1	0	1	0	9
E3a-E3b	49	50	61	41	41	50	46	34	53	38	34	37	23	7	564
E3a-E4	0	3	1	0	1	0	4	1	0	0	0	0	0	0	10
E3b-CA	0	0	0	0	0	0	0	4	17	33	32	31	48	90	255
E3b-CB	0	0	0	0	0	0	0	0	2	3	8	9	11	13	46
E3b-TRR	0	0	0	0	0	0	0	0	0	1	1	1	0	1	4
E3b-E2	1	0	0	1	0	1	0	0	0	0	1	1	0	0	5
E3b-E3a	42	45	36	34	40	39	34	60	49	42	45	36	35	4	541
E3b-E4	17	19	21	17	9	21	25	32	36	27	42	36	31	23	356
E4-CA	0	0	0	0	0	0	0	0	2	4	5	5	9	21	46
E4-CB	0	0	0	0	0	0	0	0	1	4	2	3	4	7	21
E4-TRR	0	0	0	0	0	0	0	0	1	0	0	1	0	1	3
E4-E3a	0	1	2	0	0	0	0	0	0	0	0	0	0	0	3
E4-E3b	20	26	22	17	18	9	26	27	24	26	28	29	25	15	312
E4-E5	0	0	0	0	0	0	0	0	0	0	3	1	1	4	9
E5-TRR	0	0	0	0	0	0	0	0	0	0	0	3	1	1	5
Total	132	147	147	112	110	120	137	159	193	193	216	212	213	211	2302

TABLA F.6. Frecuencias de transiciones entre estadios y censuras por periodo: clase 5

Transición de	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	Total
E3a-E3b	4	0	4	0	1	1	0	0	0	0	0	0	0	0	10
E3a-E4	1	0	0	0	1	0	1	0	0	0	0	0	0	0	3
E3b-CA	0	0	0	0	0	0	0	1	0	0	2	2	1	2	8
E3b-CB	0	0	0	0	0	0	0	0	0	0	0	0	3	0	3
E3b-E3a	0	1	1	1	0	1	0	0	0	0	0	0	0	0	4
E3b-E4	12	14	17	18	19	8	15	12	12	12	9	7	4	3	162
E4-CA	0	0	0	0	0	0	0	3	4	3	2	11	7	14	44
E4-CB	0	0	0	0	0	0	0	0	1	4	1	1	1	5	13
E4-E3b	10	8	12	17	6	11	8	12	6	13	8	7	5	3	126
E4-E5	1	0	1	2	4	6	3	7	12	13	17	11	5	8	90
E5-CA	0	0	0	0	0	0	0	0	0	1	2	1	1	5	10
E5-CB	0	0	0	0	0	0	0	0	0	0	1	1	1	0	3
E5-TRR	0	0	0	0	0	0	0	0	0	1	10	3	5	6	25
E5-E4	1	0	0	1	2	8	3	4	4	7	9	6	3	4	52
Total	29	23	35	39	33	35	30	39	39	54	61	50	36	51	554

TABLA F.7. Frecuencias de transiciones entre estadios y censuras por periodo: clase 6

Transición de	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	Total
E3a-E3b	4	3	1	3	1	1	0	1	0	0	0	0	0	0	14
E3b-TRR	0	0	0	0	1	1	1	0	2	0	0	0	0	0	5
E3b-E3a	2	2	1	0	1	0	0	0	0	0	0	0	0	0	6
E3b-E4	4	2	8	3	2	2	4	0	0	0	0	0	0	0	25
E4-TRR	0	0	0	0	1	2	3	2	3	0	0	0	0	0	11
E4-E3b	0	5	2	0	1	0	1	0	0	0	0	0	0	0	9
E4-E5	1	9	11	11	3	7	5	5	1	0	0	0	0	0	53
E5-TRR	0	0	0	0	5	7	7	9	5	0	0	0	0	0	33
E5-E4	1	4	7	3	0	2	0	2	0	0	0	0	0	0	19
Total	13	27	30	22	15	22	21	19	11	1	0	0	1	0	182

TABLA F.8. Frecuencias de transiciones entre estadios y censuras por periodo: clase 7

Síntesis descriptiva de las clases de progresión

Clase 1
Estatus: clase caracterizada por egreso prematuro del periodo de estudio
Progresión a través de la ERC: tendencia positiva y estable
Nivel de riesgo: bajo-medio
Días de permanencia en estudio: 352.65 días (por debajo media global)
Estadios de mayor permanencia: <i>E3a</i> (por encima de la permanencia media global)
Tipo de censura predominante: <i>CB</i> (de <i>P4</i> a <i>P10</i>)
Transiciones predominantes: <i>E3a – E2</i> , <i>E2 – CB</i> , <i>E3a – CB</i> y <i>E3b – CB</i>
TFG inicial: 47.36 (por encima de la media global)
TFG promedio: 49.75 (por encima de la media global)
TFG final: 50.9 (por encima de la media global)
Género: No determinante
Edad: No determinante
Escolaridad: Básica primaria
Diabetes: No determinante
Hipertensión: No determinante
Cáncer: Si
Causa de la ERC: No determinante
Ante. Enf. cardiovascular: No
Causa de egreso estudio: muerte, consulta externa, cuidados paliativos, abandono, pérdida seguimiento

TABLA G.1. Síntesis de la caracterización de la clase 1

Clase 2
Estatus: clase caracterizada por finalizar el periodo de estudio
Progresión a través de la ERC: tendencia positiva y estable
Nivel de riesgo: bajo-medio
Días de permanencia en estudio: 333.38 días (por debajo media global)
Estadios de mayor permanencia: <i>E3b</i> (por encima de la permanencia media global)
Tipo de censura predominante: <i>CA</i> (de <i>P5</i> a <i>P10</i>)
Transiciones predominantes: <i>E3b – E3a</i> , <i>E3a – CA</i>
TFG inicial: 44.38 (por encima de la media global)
TFG promedio: No determinante (cerca a la media global)
TFG final: 49.74 (por encima de la media global)
Género: femenino
Edad: 71.16 años (por encima de la media global)
Escolaridad: educación secundaria
Diabetes: no determinante
Hipertensión: no determinante
Cáncer: no
Causa de la ERC: otras distintas a HTA, diabetes y glomerulonefritis y enfermedad desconocida
Ante. Enf. cardiovascular: si
Causa de egreso estudio: finalización de periodo de estudio

TABLA G.2. Síntesis de la caracterización de la clase 2

Clase 3
Estatus: clase caracterizada por progreso similar a la población general
Progresión a través de la ERC: tendencia estable
Nivel de riesgo: medio
Días de permanencia en estudio: 1139.15 días (por encima de la media global)
Estadios de mayor permanencia: <i>E3b</i> (por encima de la permanencia media global)
Tipo de censura predominante: <i>CA</i> y <i>CB</i> (de <i>P11</i> a <i>P12</i>)
Transiciones predominantes: <i>E3b – E3a</i> , <i>E3a – E3b</i> , <i>E3a – CA</i>
TFG inicial: 50.16 (por encima de la media global)
TFG promedio: 50.87 (por encima de la media global)
TFG final: 51.4 (por encima de la media global)
Género: No determinante
Edad: No determinante
Escolaridad: educación básica y educación superior
Diabetes: no determinante
Hipertensión: no determinante
Cáncer: no determinante
Causa de la ERC: HTA
Ante. Enf. cardiovascular: no
Causa de egreso estudio: finalización de periodo de estudio

TABLA G.3. Síntesis de la caracterización de la clase 3

Clase 4

<p>Estatus: clase caracterizada por mostrar el mejor progreso a través de la ERC</p> <p>Progresión a través de la ERC: tendencia positiva</p> <p>Nivel de riesgo: bajo</p> <p>Días de permanencia en estudio: 1121.15 días (por encima de la media global)</p> <p>Estadios de mayor permanencia: <i>E2</i> y <i>E3a</i> (por encima de la permanencia media global)</p> <p>Tipo de censura predominante: No determinante</p> <p>Transiciones predominantes: <i>E3a – E2</i>, <i>E2 – E3a</i>, <i>E3a – CA</i>, <i>E2 – CA</i></p> <p>TFG inicial: 53.28 (por encima de la media global)</p> <p>TFG promedio: 60.87 (por encima de la media global)</p> <p>TFG final: 62.63 (por encima de la media global)</p> <p>Género: masculino</p> <p>Edad: No determinante</p> <p>Escolaridad: menor a primaria, educación básica y educación superior</p> <p>Diabetes: no determinante</p> <p>Hipertensión: no determinante</p> <p>Cáncer: no determinante</p> <p>Causa de la ERC: HTA</p> <p>Ante. Enf. cardiovascular: no</p> <p>Causa de egreso estudio: consulta externa y cambio de aseguradora</p>
--

TABLA G.4. Síntesis de la caracterización de la clase 4

Clase 5

<p>Estatus: clase caracterizada por progresar principalmente en estadio 4</p> <p>Progresión a través de la ERC: tendencia estable</p> <p>Nivel de riesgo: medio-alto</p> <p>Días de permanencia en estudio: 1132 días (por encima de la media global)</p> <p>Estadios de mayor permanencia: <i>E4</i> (por encima de la permanencia media global)</p> <p>Tipo de censura predominante: No determinante</p> <p>Transiciones predominantes: <i>E3a – E3b</i>, <i>E3b – E3a</i>, <i>E3a – CA</i></p> <p>TFG inicial: 41.2 (por debajo de la media global)</p> <p>TFG promedio: 38.44 (por debajo de la media global)</p> <p>TFG final: 38.75 (por debajo de la media global)</p> <p>Género: femenino</p> <p>Edad: No determinante</p> <p>Escolaridad: menor a primaria y educación básica</p> <p>Diabetes: si</p> <p>Hipertensión: no determinante</p> <p>Cáncer: no determinante</p> <p>Causa de la ERC: diabetes</p> <p>Ante. Enf. cardiovascular: no</p> <p>Causa de egreso estudio: finalización del periodo de estudio y fallecimiento</p>

TABLA G.5. Síntesis de la caracterización de la clase 5

Clase 6

<p>Estatus: clase caracterizada por progresar principalmente en estadio 5</p> <p>Progresión a través de la ERC: tendencia negativa</p> <p>Nivel de riesgo: alto</p> <p>Días de permanencia en estudio: no determinante (similar a la media global)</p> <p>Estadios de mayor permanencia: <i>E5</i> (por encima de la permanencia media global)</p> <p>Tipo de censura predominante: No determinante</p> <p>Transiciones predominantes: <i>E4 – E3b</i>, <i>E3b – E4</i> y <i>E3a – CA</i></p> <p>TFG inicial: 31.17 (por debajo de la media global)</p> <p>TFG promedio: 23.97 (por debajo de la media global)</p> <p>TFG final: 19.21 (por debajo de la media global)</p> <p>Género: no determinante</p> <p>Edad: 65.3 años (por debajo de la media global)</p> <p>Escolaridad: no determinante</p> <p>Diabetes: si</p> <p>Hipertensión: no determinante</p> <p>Cáncer: no determinante</p> <p>Causa de la ERC: diabetes</p> <p>Ante. Enf. cardiovascular: no determinante</p> <p>Causa de egreso estudio: ingreso a diálisis</p>

TABLA G.6. Síntesis de la caracterización de la clase 6

Clase 7

<p>Estatus: clase caracterizada por progresar principalmente en estadio 5 e ingresar a <i>TRR</i> de manera temprana</p> <p>Progresión a través de la ERC: tendencia negativa</p> <p>Nivel de riesgo: muy alto</p> <p>Días de permanencia en estudio: 413.86 días (por debajo de la media global)</p> <p>Estadios de mayor permanencia: <i>E5</i> (por encima de la permanencia media global)</p> <p>Tipo de censura predominante: <i>TRR</i></p> <p>Transiciones predominantes: <i>E4 – E5</i> y <i>E5 – TRR</i></p> <p>TFG inicial: 31.82 (por debajo de la media global)</p> <p>TFG promedio: 23.36 (por debajo de la media global)</p> <p>TFG final: 15.36 (por debajo de la media global)</p> <p>Género: no determinante</p> <p>Edad: 58.3 años (por debajo de la media global)</p> <p>Escolaridad: no determinante</p> <p>Diabetes: si</p> <p>Hipertensión: no determinante</p> <p>Cáncer: no determinante</p> <p>Causa de la ERC: diabetes y glomerulonefritis</p> <p>Ante. Enf. cardiovascular: no determinante</p> <p>Causa de egreso estudio: ingreso a diálisis y trasplante</p>
--

TABLA G.7. Síntesis de la caracterización de la clase 7