

# VDR

E2: Análisis de la VRD actual

22/01/2020

## Historial de revisiones

Versión	Objeto	Fecha	Autor(es)
0.1	Creación	16/01/2020	Luis Manuel Morán
0.2	Revisión	22/01/2020	Luis Manuel Morán
0.3	Añadir estudios	28/01/2020	Luis Manuel Morán
0.4	Añadir estudios adicionales	20/02/2020	Luis Manuel Morán
0.5	Añadir estudio variantes	28/02/2020	Luis Manuel Morán
0.6	Ajustes finales	12/03/2020	Luis Manuel Morán

## Índice

<b>Introducción</b>	<b>4</b>
<b>Análisis de los datos suministrados</b>	<b>4</b>
2.1 Datos utilizados	4
2.2 Criterios de exclusión de casos	5
2.3 Tiempo hasta colonoscopia	8
2.4 Tiempo hasta extracción	9
2.5 Origen VDR	10
2.6 Origen colonoscopias	11
2.7 Visitas previas a oncología	13
2.8 CCR de intervalo	15
2.9 Cirugía Urgente	16
2.10 Mortalidad	16
2.11 Visitas a médico de atención primaria	17
2.12 Tiempo y distribución cirugía	22
2.13 Análisis de trazas de pacientes previas a extracción	24
2.14 Términos clínicos seleccionados, descartados y agrupaciones	30
2.15 Frecuencia términos agrupados	31
2.16 Presencia de términos agrupados	34
2.17 Temporalidad de términos agrupados	36

## 1. Introducción

Este informe corresponde con el entregable “**E2: Análisis de la VDR actual**”, definido en el Anexo I del contrato firmado entre la Fundación Biomédica Galicia Sur y Gradient en 26/05/2019.

El referido informe contiene las respuestas a las preguntas realizadas por el equipo del Dr. Modesto Martínez, PI del proyecto AES PI17/02176 “Desarrollo de la minería de procesos para la evaluación de vías rápidas de diagnóstico”, las cuales se han contestado a través del análisis de los datos incluidos en los siguientes ficheros:

- VDR\_DatosEstructurados\_Seudonimizados\_v.04.00.ods, suministrados el 20/11/2019.
- VDR\_DatosNoEstructurados\_Seudonimizados.ods, suministrados el 15/11/2019.

También se ha empleado la información contenida en el documento VDR-DT-v.01.03-Metodología.odt, suministrado en 20/11/2019.

## 2. Análisis de los datos suministrados

A partir de estos datos, siguiendo las instrucciones del Dr. Ismael Said Criado se han extraído distintos resultados, a mayores de los proporcionados anteriormente en el documento “20191017\_Informe preliminar preprocesado de datos de VDR” entregado con anterioridad.

### 2.1 Datos utilizados

En análisis previos se comprobaron las **limitaciones** que la **escasez de casos** producía en el estudio de los procesos de los pacientes y sus estadísticos. Esta escasez venía provocada, no sólo por los datos disponibles, sino también por los filtros que debían pasar para asegurar, en la medida de lo posible, que la extracción con la que se contaba para cada paciente, fuera la que había llevado al diagnóstico de Cáncer Colorrectal (en posterioridad denominado CCR).

En esta segunda fase, se decidió de forma consensuada, intentar incluir la mayor cantidad de casos posibles, dentro de un umbral de duda razonable, de que las extracciones contempladas fueran las diagnósticas.

Partiendo del evento central, denominado “**extracción**”, consistente en el análisis por parte de Anatomía patológica de una muestra como positiva para CCR.

En la siguiente tabla puede comprobarse desde cuándo se cuentan con datos para las distintas tablas.

	2012	2013	2014	2015	2016	2017	2018
APA							
CIRUGÍA							
PRUEBAS							
EPISODIOS Y COND.							
FARMACIA							
GASTO FARMACÉUTICO							
HEMOGRAMA							
HOSPITALIZACIÓN							
INFORMES							
Lista Espera Quirúrgica							
MÉDICO FAMILIA							
MORTALIDAD							
PAC							
URGENCIAS							

Fig1: Fechas de las distintas tablas

## 2.2 Criterios de exclusión de casos

De los **473 pacientes** presentes en los datos proporcionados, la totalidad de ellos contaban con el evento extracción, en al menos una ocasión.

De estos pacientes **12 no contaban** con una solicitud de una prueba de **colonoscopia**, evento denominado “colonoscopia” o la solicitud de una prueba de colonoscopia por vía rápida, evento denominado “colonoscopia vía rápida”.

Estos son los NASI pseudoanonimizados de estos pacientes:

- 086f0910a1799eb7c8755d40d1594930
- 0dec1307795ab6035da0c8f3c968f170
- 1e617d7c94eab2ee5cb7edce858b8a96
- 4cb65c53ca7243df60a9d0590f04da37
- 51093482f14eef0042e59802fe639419
- 5695def7d4584bcb1d5b2bee7c659c80

- 89b5dc0ab3f08ff8325b005fe28d9d5c
- a9e3f5f2684ac240c9a55dcb8d9357e
- b022c80c6bacf263ebc350e674cea570
- d67b2ad6a06bb6dfab3975b8ef75ba8e
- d7ba4a89d6f0891656f41f54230cbfd3
- dd7ded21085a3a495c72bb8d7756d3a3

Se denominarán en futuras apariciones como **pacientes\_sin\_colonosopia**.

De estos, la gran mayoría (10) tienen, en fechas cercanas a la extracción, eventos de urgencias, ingresos hospitalarios o cirugías. Algunos cuentan con pruebas como 'POLIPECTOMIA ENDOSCOPICA COLON' o 'GASTROSCOPIA'. Solo un paciente, nasi 4cb65c53ca7243df60a9d0590f04da37, presenta únicamente visitas al médico de atención primaria, usando términos como "anemia" y "estreñimiento", y otro, el d67b2ad6a06bb6dfab3975b8ef75ba8e, que no presenta ningún otro evento anterior.

Adicionalmente, **9 pacientes** no presentan fechas de **colonoscopia anteriores** a la extracción. Previamente y cerca de la extracción, como anteriormente, varios presentan cirugías, ingresos en urgencias o en el hospital. Esta es la lista de los NASI pseudoanonimizados de estos pacientes:

- ea6d285ac76a4725787b2a80f68e5398
- 8ba8cf36f20ff526d92803236a275071
- 359ba49e53924271c2c1586e1ca9dff
- ecea4caa1dd89f95d21251d036bb618b
- 70db4ab43cc432b538ffde4bf32dfa27
- 66b8058f214bf90b477cd4d8bc6e99fd
- eb40b8751742f4a54fb61a2fece84819
- b445f00487989f45129b0027d23595c5
- be5dc86689aa2419376855841b2871af

Se denominarán como **pacientes\_sin\_colonosopia\_anterior**.

Adicionalmente en un estudio preliminar de tiempos hasta colonoscopia, se descubren 7 pacientes con tiempos excesivos entre el evento "colonoscopia" o "colonoscopia\_via\_rapida" y el evento "extraccion", lo que lleva a deducir algún tipo de anomalía.

Estos son los nasi de estos pacientes junto al tiempo (en días) descubierto. Entre la colonoscopia y la extracción:

- 7f4858a9a8dcbebf6f50654f1bd43b30 - 134 días
- 3fd22c6d650bbf48e28eff20b21a1083 - 84 días

- 6d6ea85fa8cd4fcbca28218003dcf9a0 - 14 días
- b97364306675b9e3048c27c01ad06e7c - 34 días
- 2d1d5cc13abe02db484b366080d8380e - 63 días

Entre colonoscopia de vía rápida y extracción:

ffe754091300e2134616c0bef3918913 - 98 días  
2750423e48dee069c9546c62d76c3e41 - 38 días

Se denominarán como **pacientes\_de\_tiempos\_anomalos**.

Dado el interés tratado en reuniones previas sobre la posibilidad de la existencia en el estudio de pacientes provenientes de programa de cribado, se han filtrado también los pacientes con colonoscopia de código de prestación 3540, que ha resultado solo ser uno, el de NASI pseudoanonimizado:

- a6c173b0afb4813c16a856c35ddaa6ae

Se denominarán como **pacientes\_de\_cribado**.

Ignorando estos cinco grupos de pacientes, de los 473 pacientes iniciales quedan **444 pacientes**, con una distribución:

- 132 pacientes que van por VDR
  - 36 mujeres
  - 96 hombres
- 312 pacientes que no van por VDR
  - 124 mujeres
  - 188 hombres

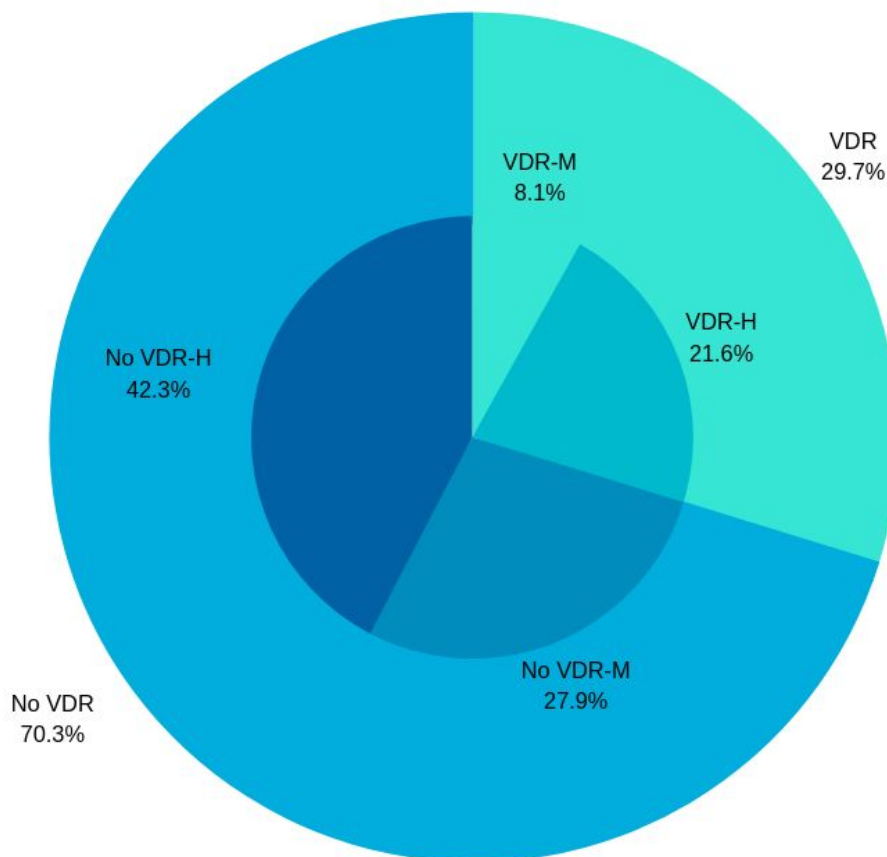


Fig2: Distribución de los pacientes en VDR y No VDR

## 2.3 Tiempo hasta colonoscopia

Se plantea realizar un estudio de tiempo para comprobar posibles diferencias en estos para los procesos de los pacientes que de VDR y los que no lo son, para el tiempo transcurrido entre la **petición de la cita** de colonoscopia hasta la cita de la **colonoscopia**.

Los resultados son los siguientes:

	VDR	No VDR
--	-----	--------



Media (días)	7.082	33.897
Desviación estándar (días)	6.795	91.547
Máximo (días)	67	1094
Mínimo (días)	0	0
Mediana (días)	6	7

Fig3: Tabla de tiempos desde petición de colonoscopia hasta cita

Lo que representa una reducción de un **~79.1%** en media y un **~92.3%** en desviación estándar.

A continuación puede verse una representación gráfica:

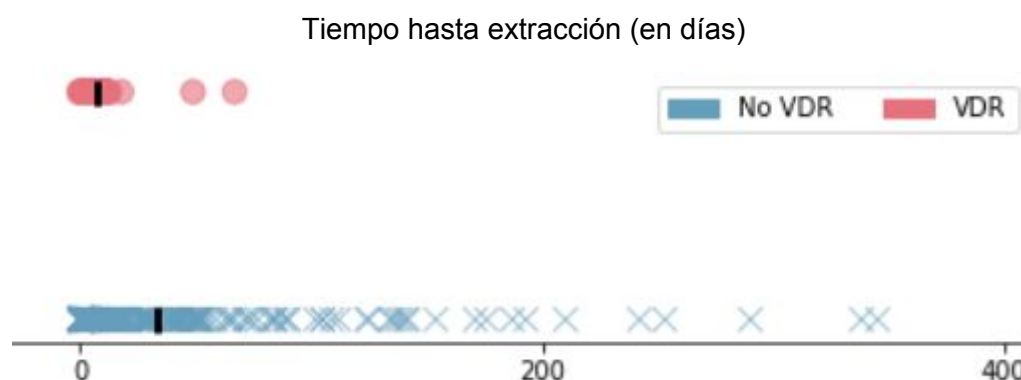


Fig4: Representación de tiempos desde petición de colonoscopia hasta cita

## 2.4 Tiempo hasta extracción

Se realiza un estudio de tiempos, en el que se planteó comprobar posibles diferencias de tiempos en los procesos de los pacientes que de VDR y los que no lo son, para el tiempo transcurrido entre la cita de la **colonoscopia** y el evento **extracción**.

Los resultados son los siguientes:

	VDR	No VDR
Media (días)	1.587	1.968

Desviación estándar (días)	0.615	1.012
Máximo (días)	5.5	17.5
Mínimo (días)	0.5	0.5
Mediana (días)	1.5	1.54

Fig5: Tabla de tiempos desde evento colonoscopia hasta extracción

Lo que representa una reducción de un **~19.3%** en media y un **~39.2%** en desviación estándar.

A continuación puede verse una representación gráfica:

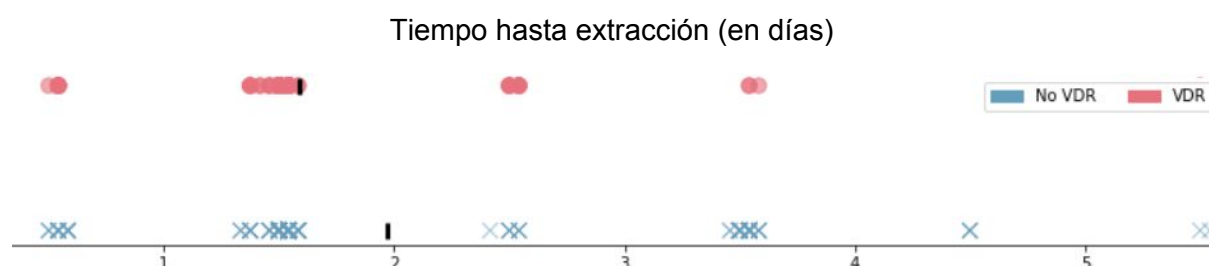


Fig6: Representación de tiempos desde evento colonoscopia a extracción

## 2.5 Origen VDR

Una de las preguntas que se plantearon como de interés en los estadios iniciales del proyecto es desde **qué servicio** era **introducido el paciente en la VDR**. Para ello se ha realizado un estudio del proceso de los pacientes, de forma que se plantea la hipótesis de que el evento anterior más cercano al de “Colonoscopia vía rápida” podría aportar información al respecto.

Para obtener estos datos se encontraron varias complejidades a tener en cuenta, lo que llevó a adoptar ciertas aproximaciones:

- Un proceso de paciente puede contener eventos de colonoscopia y colonoscopia VDR anteriores a extracción: Se consideró el más cercano a la extracción.
- Para determinada colonoscopia hay que tener en cuenta no la fecha de la cita, si no la de la prescripción.
- En caso de coincidir varios eventos, se considera el más específico: Especialistas sobre atención primaria, urgencia sobre ellos.

El resultado fue el siguiente:

- Médico de familia: 51
- Digestivo: 42
- Urgencia: 13
- Otros: 9
- Oncología: 6
- Cirugía: 6
- Hospital: 4
- Medicina interna: 1

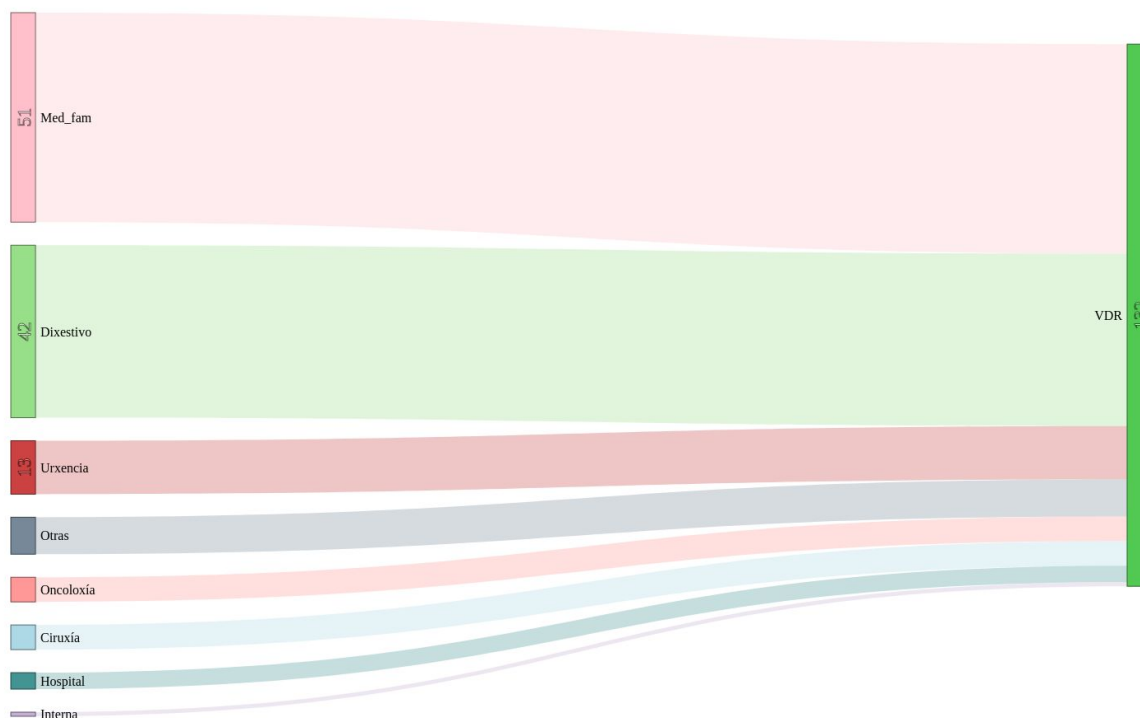


Fig7: Diagrama Sankey mostrando una aproximación de qué servicio añade al paciente a VDR

## 2.6 Origen colonoscopias

Se propone desde el equipo médico, contar con una visualización sobre el **origen** de las distintas **colonoscopias**, que se representa en el siguiente gráfico:

Colonoscopia	Número
Colonoscopia	503
Vía Rápida CCR	139
Colonoscopia H. Salnés	133
Colonoscopia V.R.	17
Colonoscopia tras teleconsulta	8
Colonoscopia con Biopsia	2
Colonoscopia Cribado CCR	1

Fig8: Tabla origen colonoscopias

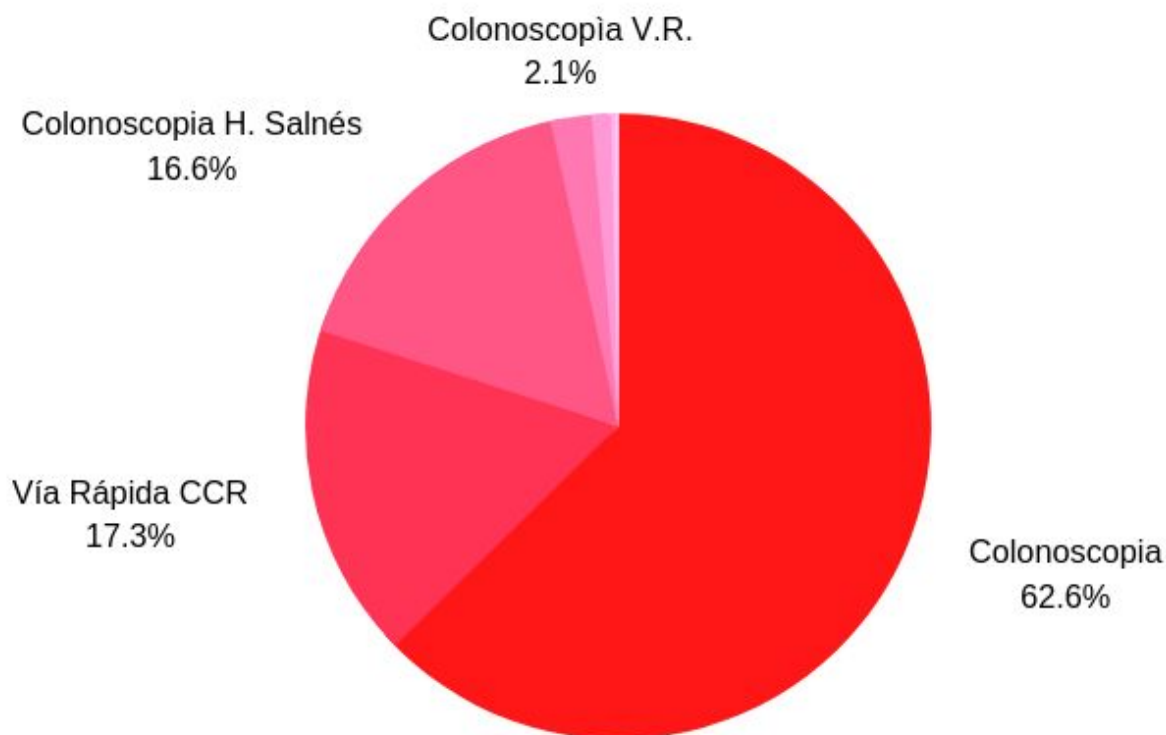


Fig9: Gráfico origen colonoscopias

## 2.7 Visitas previas a oncología

Se realiza un estudio para valorar cuántos pacientes de los disponibles presentan visitas a Oncología previas a la extracción de la muestra. Esto podría deberse a que el paciente tuviera un caso de cáncer (CCR o de otro tipo) activo cuando se recogió esta muestra.

Se detectan 43 pacientes con visitas a Oncología previas al evento extracción. Se distribuyen de la siguiente forma:

Total	VDR	NoVDR	Sin colonoscopias previas
-------	-----	-------	---------------------------

43	11	30	2
----	----	----	---

Fig10: Tabla pacientes con visitas oncológicas previas a extracción



Fig11: Distribución de pacientes

Se realiza a continuación un análisis de los tiempos de esta última visita a oncología para las dos tipologías de pacientes estudiados (con VDR y no VDR):

	VDR	No VDR
Media (días)	13.45	148.77
Desviación estándar (días)	17.60	171.19
Máximo (días)	50	603
Mínimo (días)	4	0
Mediana (días)	6	98.5

Fig12: Tabla de tiempos de visitas oncológicas previas a extracción

Tiempo última visita oncología hasta extracción (en días)

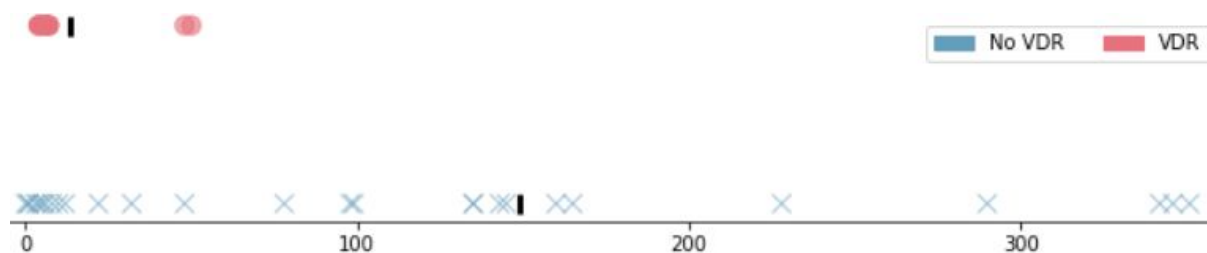


Fig13: Representación de tiempos de visitas oncológicas previas a extracción

## 2.8 CCR de intervalo

Se planteó por parte del equipo médico la posibilidad de hacer un estudio de los **pacientes** a los que se le practicó **en los años anteriores** a la **colonoscopia** que llevó a la extracción de la muestra positiva, y por tanto diagnóstica, **de resultado negativo**.

Hace un estudio así no es posible con los datos actuales por las siguientes limitaciones:

- Falta de certeza en cuál es la colonoscopia diagnóstica del paciente.
- Falta de datos de extracciones previas al año 2016.
- Datos únicamente de extracciones de resultado positivo.

Con todas estas limitaciones, se hace un estudio detectando la presencia de colonoscopias previas a la que llevó a la primera extracción positiva con la que contamos, y se representa este número para pacientes de VDR y que no son de VDR. Para minimizar el efecto de la presencia de colonoscopias que no se llevaron a cabo (cancelaciones, o cambios a VDR) se establece un plazo mínimo de 60 días.

Los resultados son los siguientes:

	VDR	NoVDR
Ausencia colonoscopias previas	128	284
Presencia colonoscopias previas	4	28
% de pacientes con col. previas	3,03%	8,97%

Fig14: Tabla colonoscopias previas a diagnóstica

## 2.9 Cirugía Urgente

En este estudio se buscan pacientes (de entre todos los disponibles, no sólo los seleccionados en el punto 2.1) que cuentan con una **cirugía urgente** de fecha **cercana a la extracción** de muestra positiva en la biopsia.

Estableciendo como umbral de fecha cercana a 30 días, se encuentran **12 pacientes**, cuyos nasi se listan a continuación:

- fa38a6eb40fd17825ca528c80636d4e1
- 70db4ab43cc432b538ffde4bf32dfa27
- 53d87818194d36ccb64171c0a65b2fc0
- ffe754091300e2134616c0bef3918913
- 83ab64e9b474cc4f767fcabf5bcad194
- 584d8b198d2875c655f021b003264f33
- ada216cd9c8e7280fa07022626d73b8b
- 2750423e48dee069c9546c62d76c3e41
- f82846e5031be242c41b4d461e504fd2
- 075020dd0f8e8356751c48bd6871b176
- aaf72611bc984d6cf3b86621a96dac9f
- 71b9ee837d6ceb2b843cb69e416184c6

De ellos 1 cuenta con colonoscopia de vía rápida, 8 con colonoscopias (sin ser VDR) y **3 no cuentan con colonoscopias previas**:

- 70db4ab43cc432b538ffde4bf32dfa27
- ffe754091300e2134616c0bef3918913
- 2750423e48dee069c9546c62d76c3e41

## 2.10 Mortalidad

Se realiza un estudio de la mortalidad de los pacientes seleccionados, dependiendo de si forman parte de la VDR o no.

	VDR	NoVDR
Supervivientes	103	221
Fallecimientos	29	91



% de fallecimientos	21,97%	29,17%
---------------------	--------	--------

Fig15: Tabla mortalidad

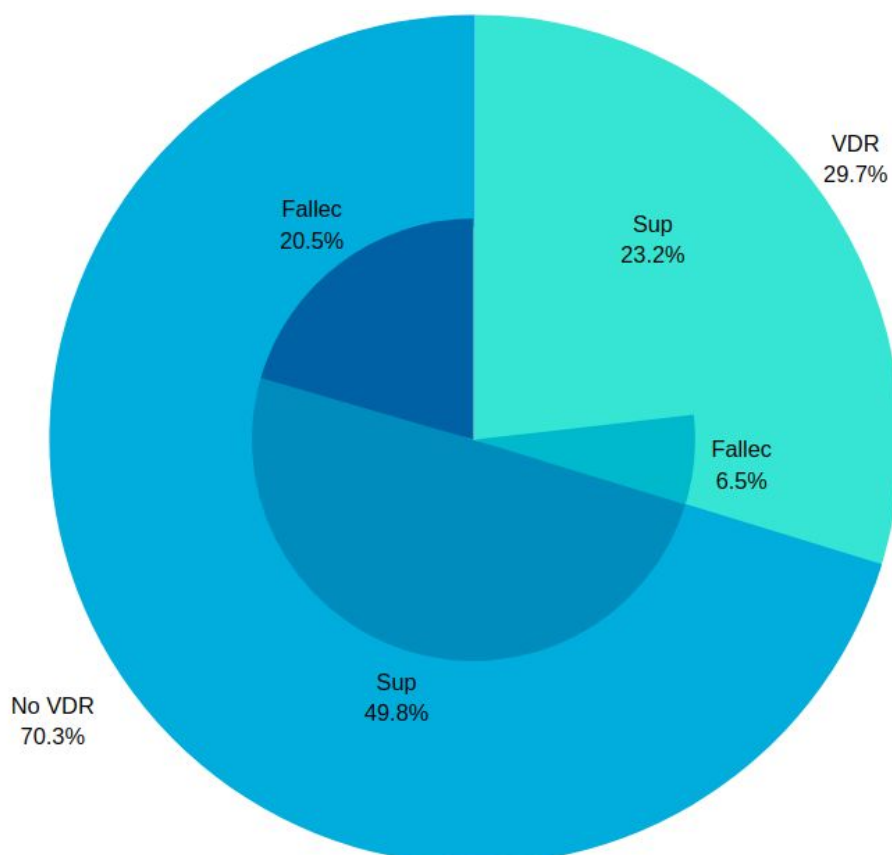


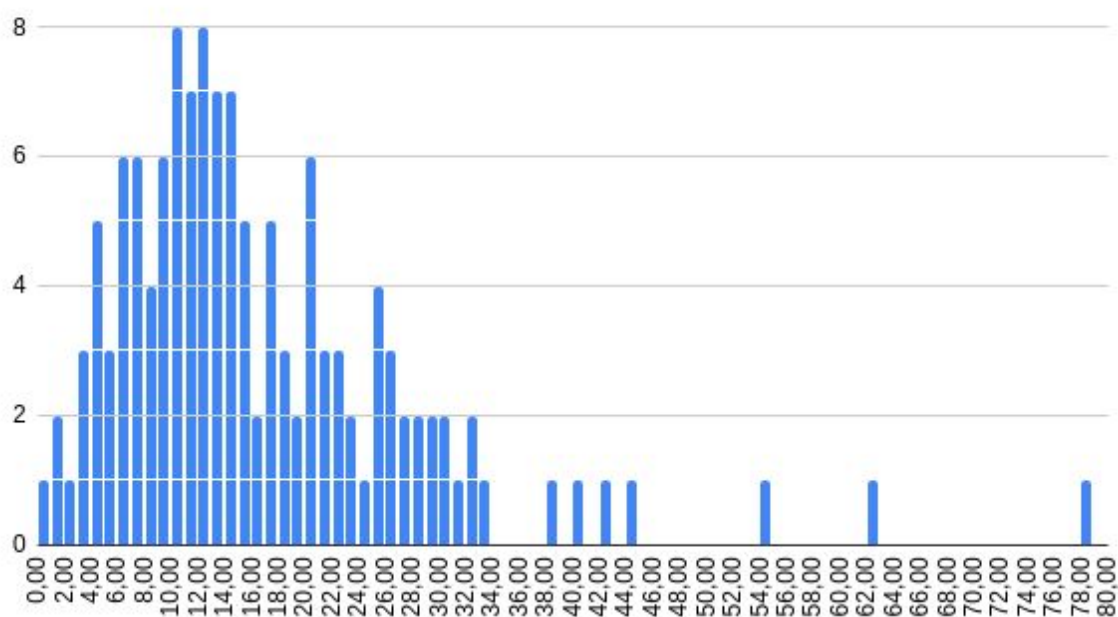
Fig16: Gráfico mortalidad

## 2.11 Visitas a médico de atención primaria

Se comparan en esta ocasión el **número de visitas al médico de atención primaria** de los pacientes, en un periodo de un año anterior a la fecha de extracción. Se pretende constatar diferencias en los patrones descubiertos respecto a los pacientes de VDR y los que no lo son.

Se obtienen los siguientes **histogramas** de número de visitas al médico de atención primaria, en el **año anterior** a la extracción. El **eje horizontal** representa el **número de visitas al médico** de atención primaria, y el **eje vertical**, **número de pacientes**.

VDR



No VDR

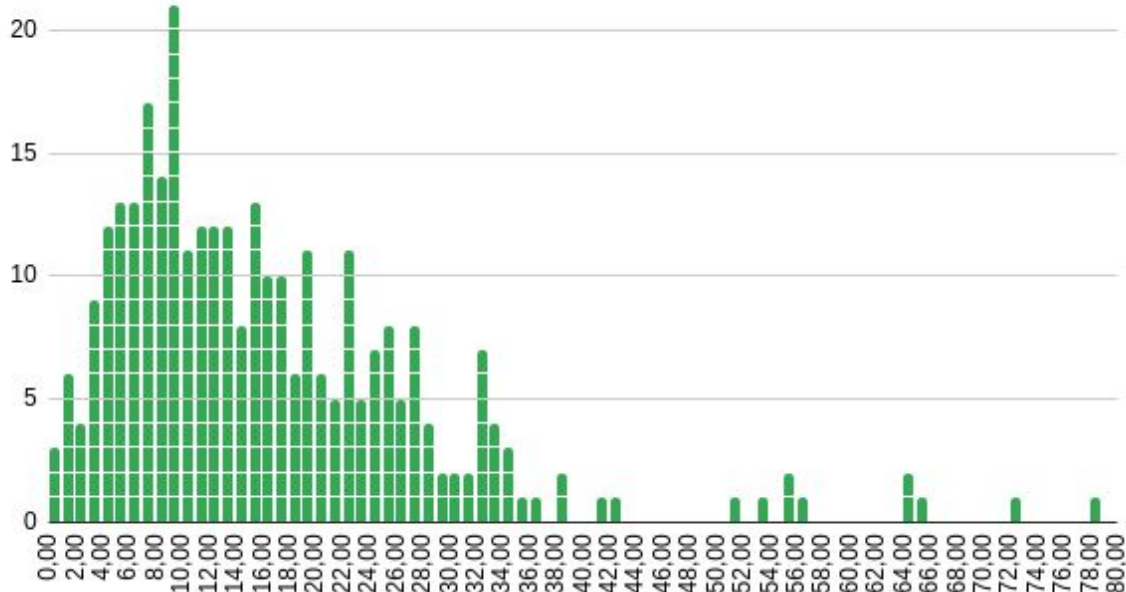


Fig17: Histograma de número de visitas a MAP en 1 año previo a extracción

Se aprecian diferencias notables en los valores absolutos, pero hay que tener en cuenta la diferencia en número de pacientes. También se detectan en el patrón, aunque esto no es muy apreciable en gráficos separados. Por ello se confecciona el siguiente gráfico, en el que se comparan los **histogramas** de forma que se **ajustan las frecuencias según el número de pacientes** (es decir, se ajustan la cantidad de pacientes con cierto número de visitas con el número de pacientes, haciendo ambos histogramas comparables) para así hacer más comparables visualmente los patrones:

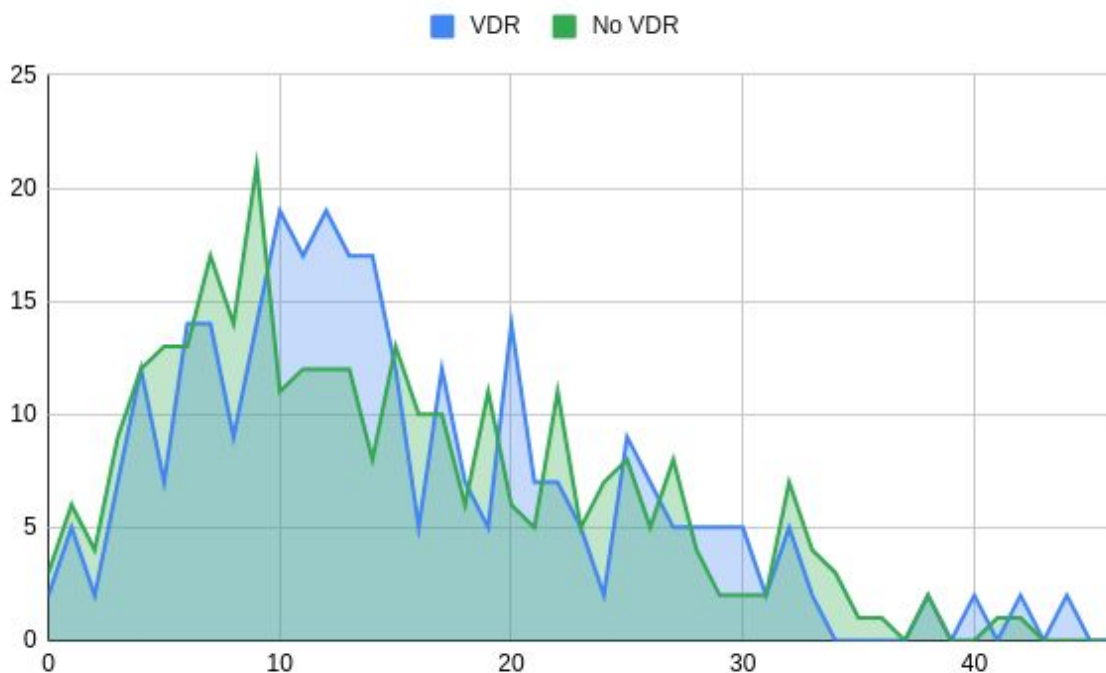
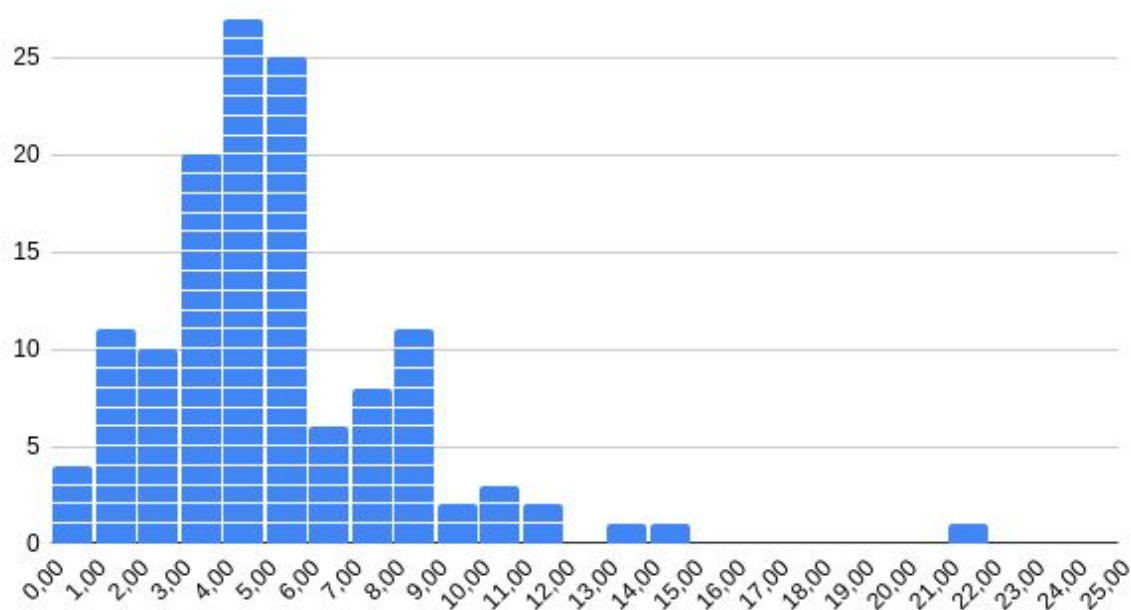


Fig18: Gráfico ajustado comparando número de visitas a MAP en 1 año previo a la extracción

También se han estudiado estos mismos gráficos en el plazo de **dos meses** anteriores a la extracción, donde los patrones pueden verse más claramente:

## VDR



## No VDR

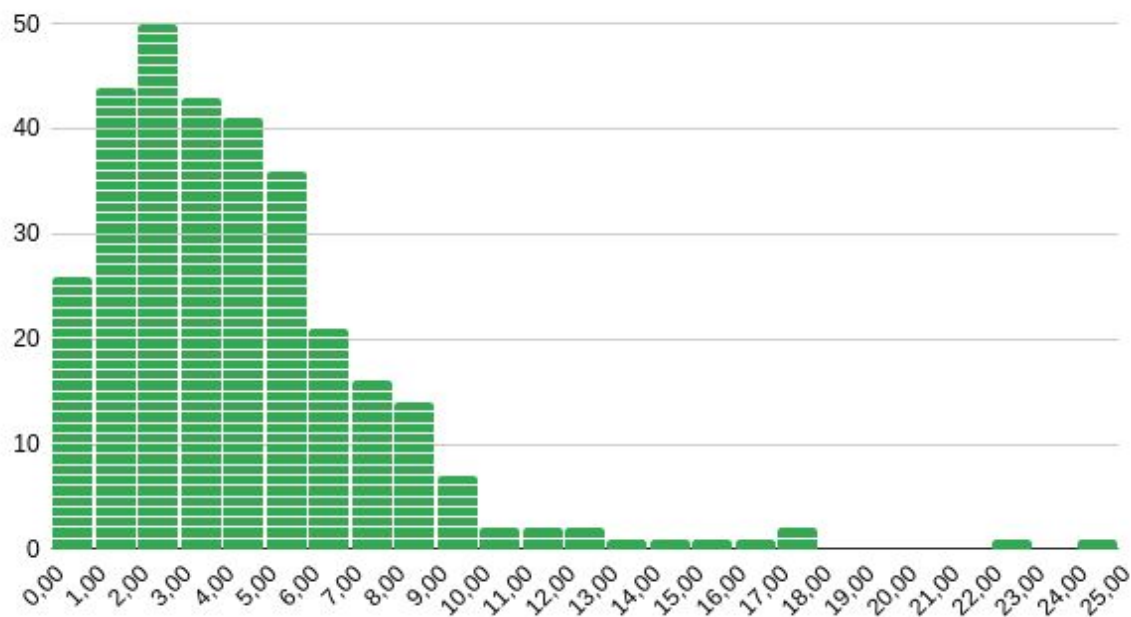


Fig19: Histograma de número de visitas a MAP en 2 meses previos a extracción

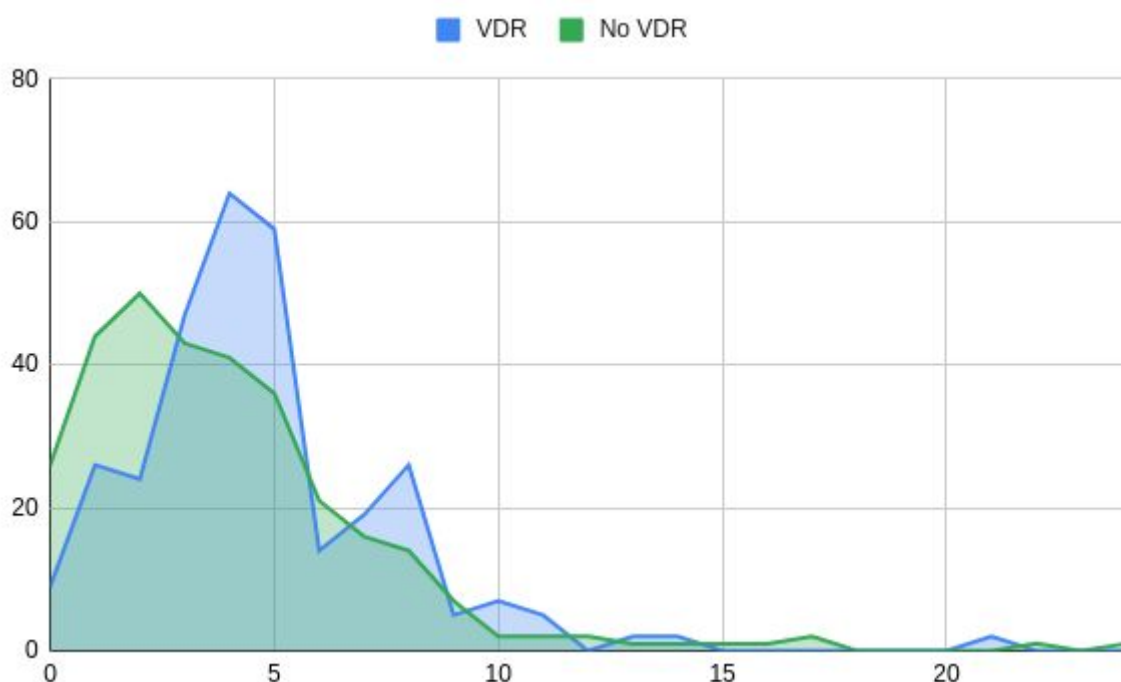


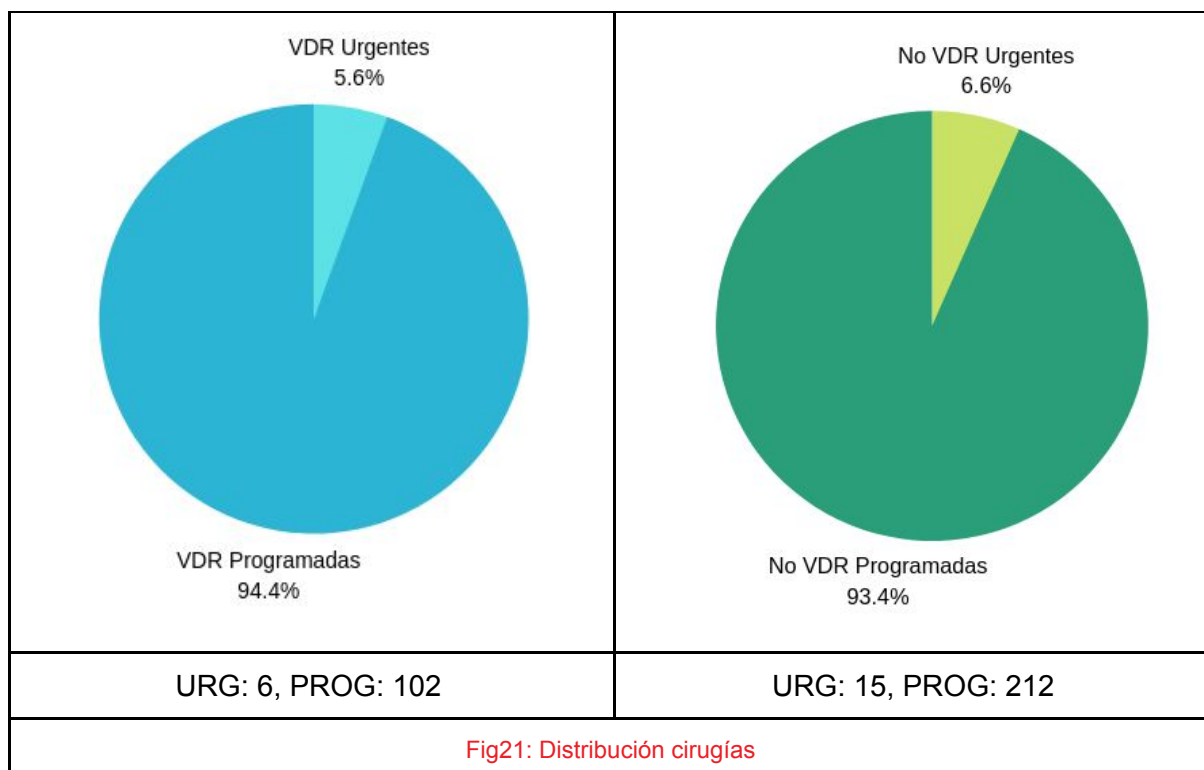
Fig20: Gráfico ajustado comparando número de visitas a MAP en 2 meses previos a la extracción

## 2.12 Tiempo y distribución cirugía

Se realiza un estudio sobre la **distribución del tipo de cirugías, urgentes o programadas**, que se realizan a los pacientes dependiendo de su pertenencia a VDR o no. Para ello se localizan las cirugías generales o digestivas más próximas y posteriores a la extracción para los distintos pacientes.

En este proceso se descubre que existen **120 pacientes** que no tienen cirugías que cumplan con estas características. De los 444 de nuestro estudio, existen los **335 pacientes** que sí, y su distribución es la siguiente:

VDR	No VDR
-----	--------



Con respecto a un estudio de tiempos, se encuentran los siguientes resultados:

	VDR URG	VDR PROG	No VDR URG	No VDR PROG
Media (días)	41.13	60.17	12.44	52.18
Desviación estándar (días)	60.17	44.66	22.96	62.84
Máximo (días)	162.5	216.45	92.42	533.42
Mínimo (días)	6.46	8.42	0.42	2.58
Mediana (días)	19.96	34.46	3.46	33.44

Fig22: Tabla de tiempos desde extracción hasta cirugía

En estos resultados puede verse una gran dispersión, que puede verse representada en la siguiente gráfica:

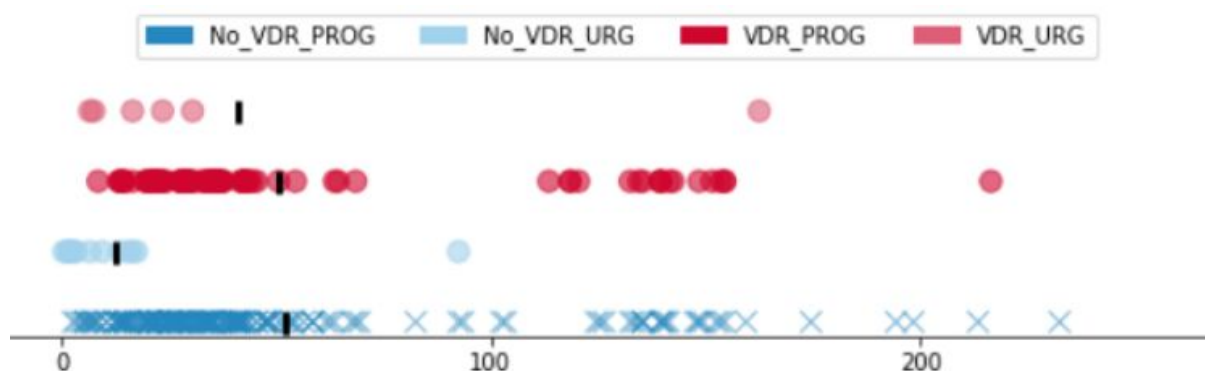


Fig23: Representación de tiempos desde evento colonoscopia a extracción

## 2.13 Análisis de trazas de pacientes previas a extracción

Aplicar al proceso obtenido técnicas de tradiciones de minería de procesos es complejo, debido al **gran tamaño de las trazas y su enorme variabilidad**, encontrando, incluso en la versión filtrada, 21 actividades diferentes y trazas de hasta 131 eventos, haciendo que cada una de estas trazas fuera única en el proceso.

Para comenzar con un estudio exploratorio que permita encontrar ciertos patrones con los que poder clasificar los pacientes de una forma amplia dependiendo de ciertos eventos clave, **se limitaron las actividades** a estudiar a las siguientes:

- HOSP\_ING
- CIR\_URG
- CIR\_PROG
- URGENCIA, PAC - Combinados en URGENCIA
- COLONOSCOPIA
- COLONOSCOPIA VIA RAPIDA
- EXTRACCION
- MEDICINA INTERNA, DIXESTIVO, ONCOLOGIA MEDICA - Combinados en ESPECIALISTA



Gracias a esta aproximación simplificada, **se pudieron encontrar ciertos patrones**, que además presentaban diferencias para los procesos de pacientes que participaban de VDR y los que no.

En la siguiente tabla se pueden encontrar estos patrones para los eventos anteriores a la extracción con un periodo temporal de 90 días:

VDR	
34 trazas 25,76%	colonoscopia_via_rapida extraccion
28 trazas 21,21%	especialista colonoscopia_via_rapida extraccion
16 trazas 12,12%	urgencia colonoscopia_via_rapida extraccion
7 trazas 5,30%	especialista especialista colonoscopia_via_rapida extraccion
6 trazas 4,55%	urgencia especialista colonoscopia_via_rapida extraccion
5 trazas 3,79%	urgencia urgencia colonoscopia_via_rapida extraccion
No VDR	

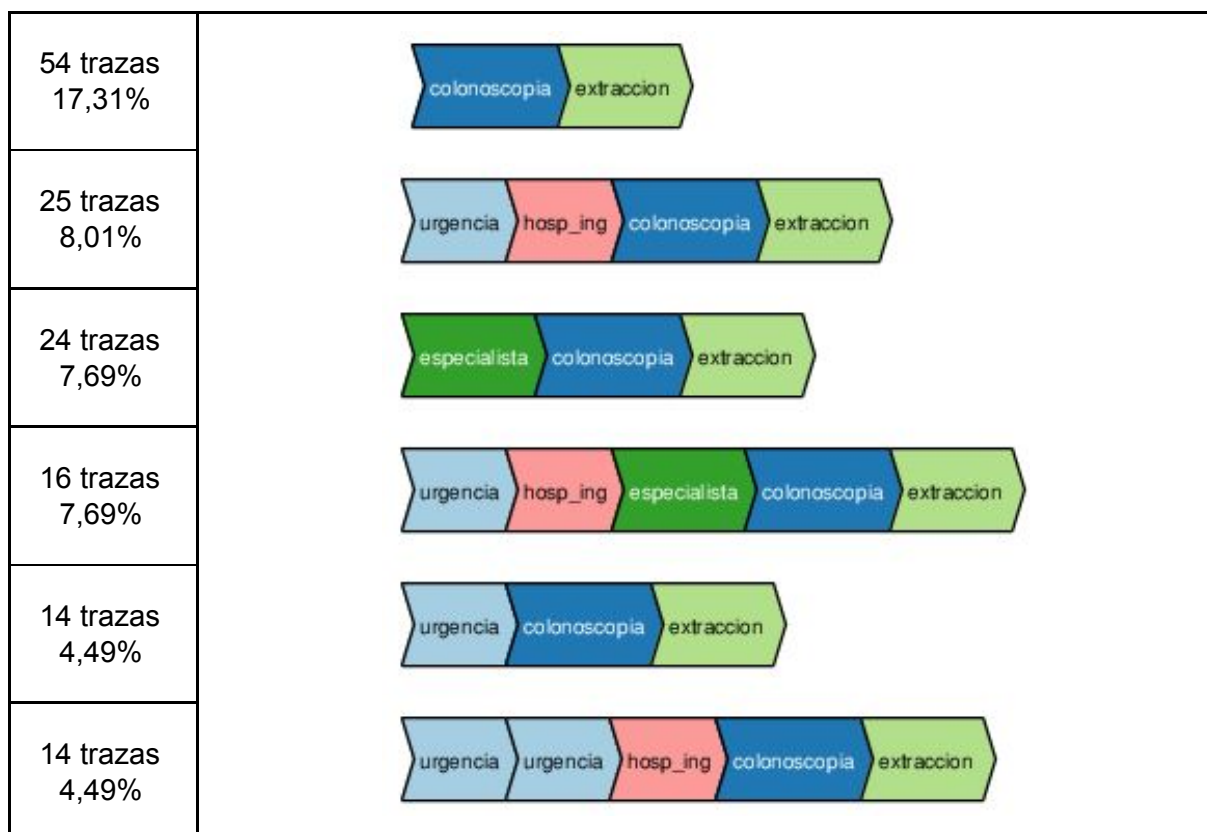
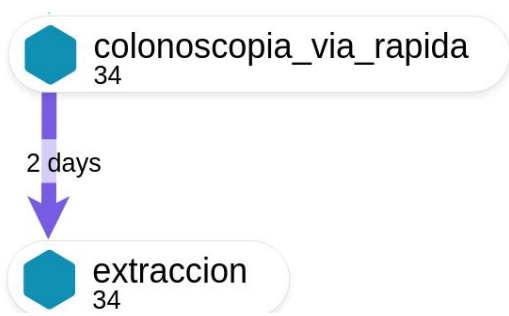


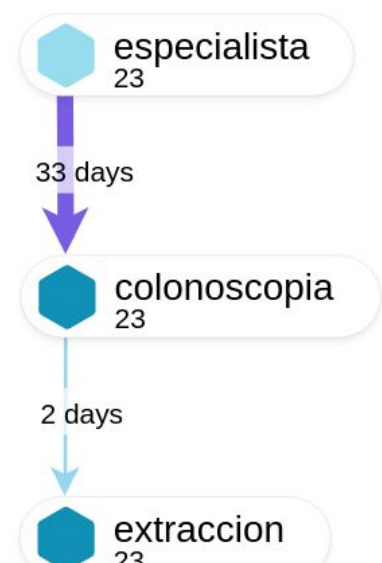
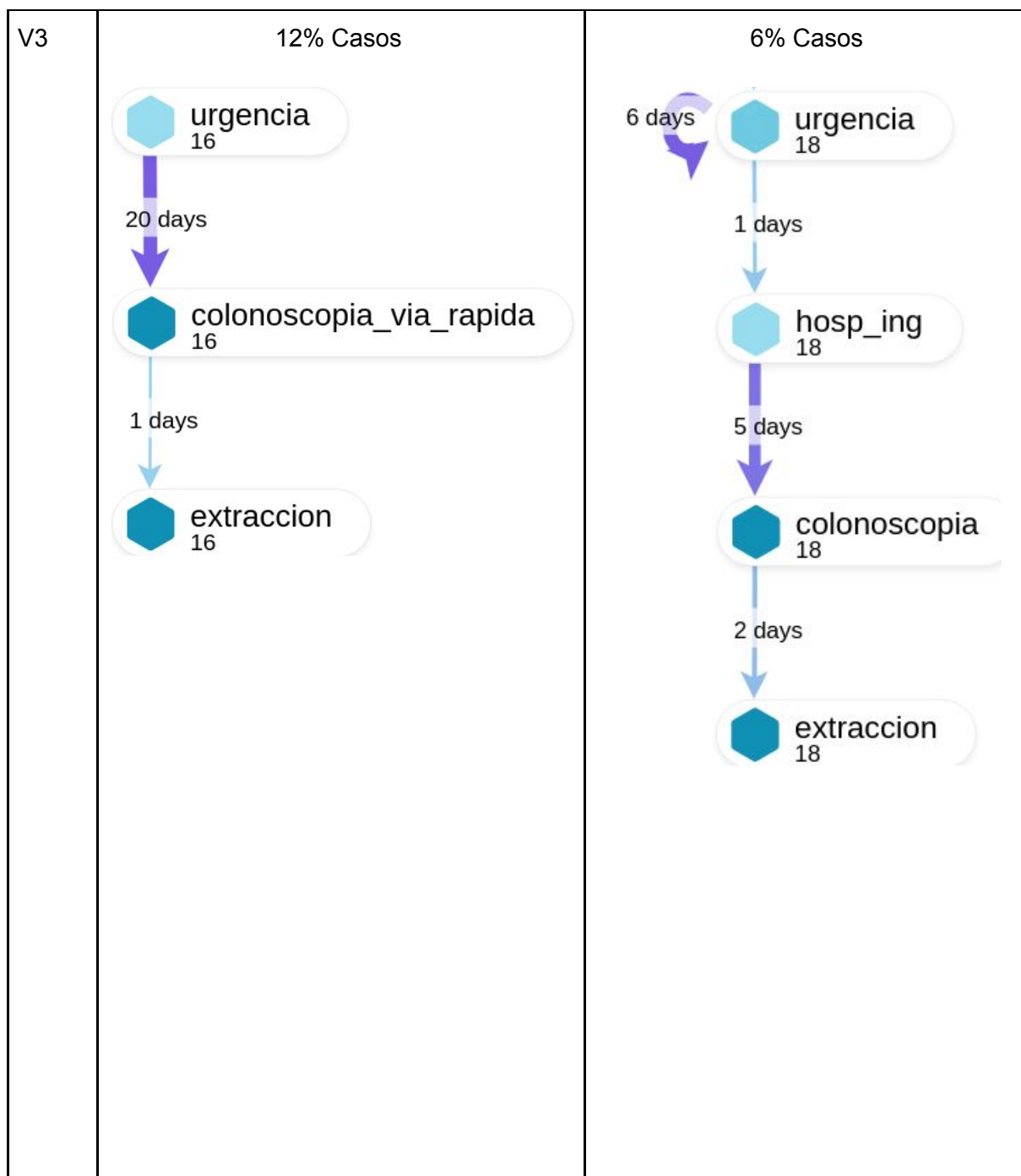
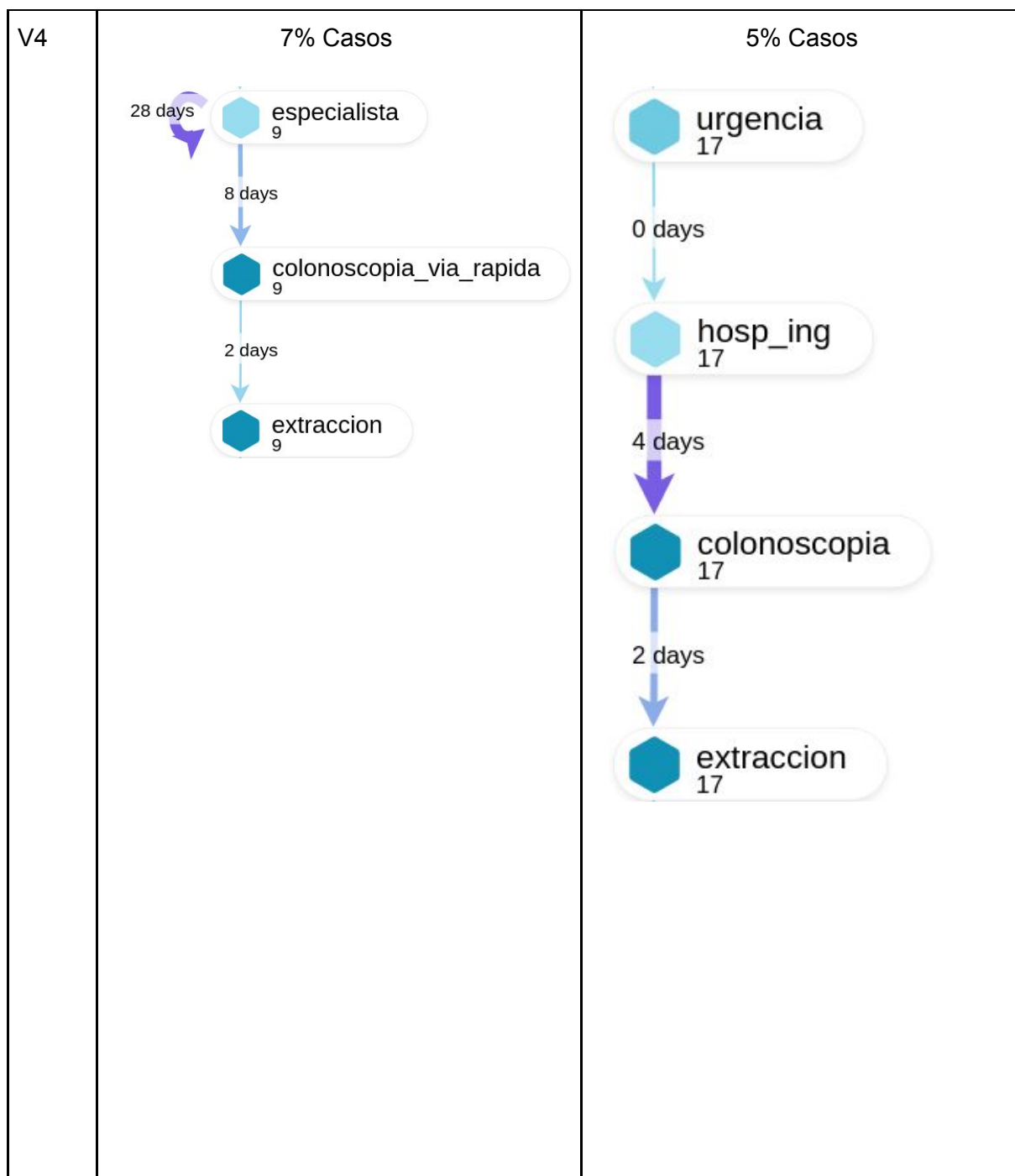


Fig24: Trazas simplificadas de los procesos de pacientes previos 90 días a la extracción

A continuación se extenderá la información previa con mayor detalle para cada una de las variantes, utilizando distinta algoritmia, que tiene también en cuenta los casos similares con respecto a tiempos, que como puede verse hace variar las variantes de casos de no VDR:

	VDR	No VDR
V1	<p>26% Casos</p>  <pre> graph TD     A[colonoscopia_via_rapida 34] -- 2 days --&gt; B[extraccion 34] </pre>	<p>17% Casos</p>  <pre> graph TD     A[colonoscopia 54] -- 2 days --&gt; B[extraccion 54] </pre>
V2	<p>21% Casos</p>  <pre> graph TD     A[especialista 28] -- 10 days --&gt; B[colonoscopia_via_rapida 28]     B -- 2 days --&gt; C[extraccion 28] </pre>	<p>7% Casos</p>  <pre> graph TD     A[especialista 23] -- 33 days --&gt; B[colonoscopia 23]     B -- 2 days --&gt; C[extraccion 23] </pre>





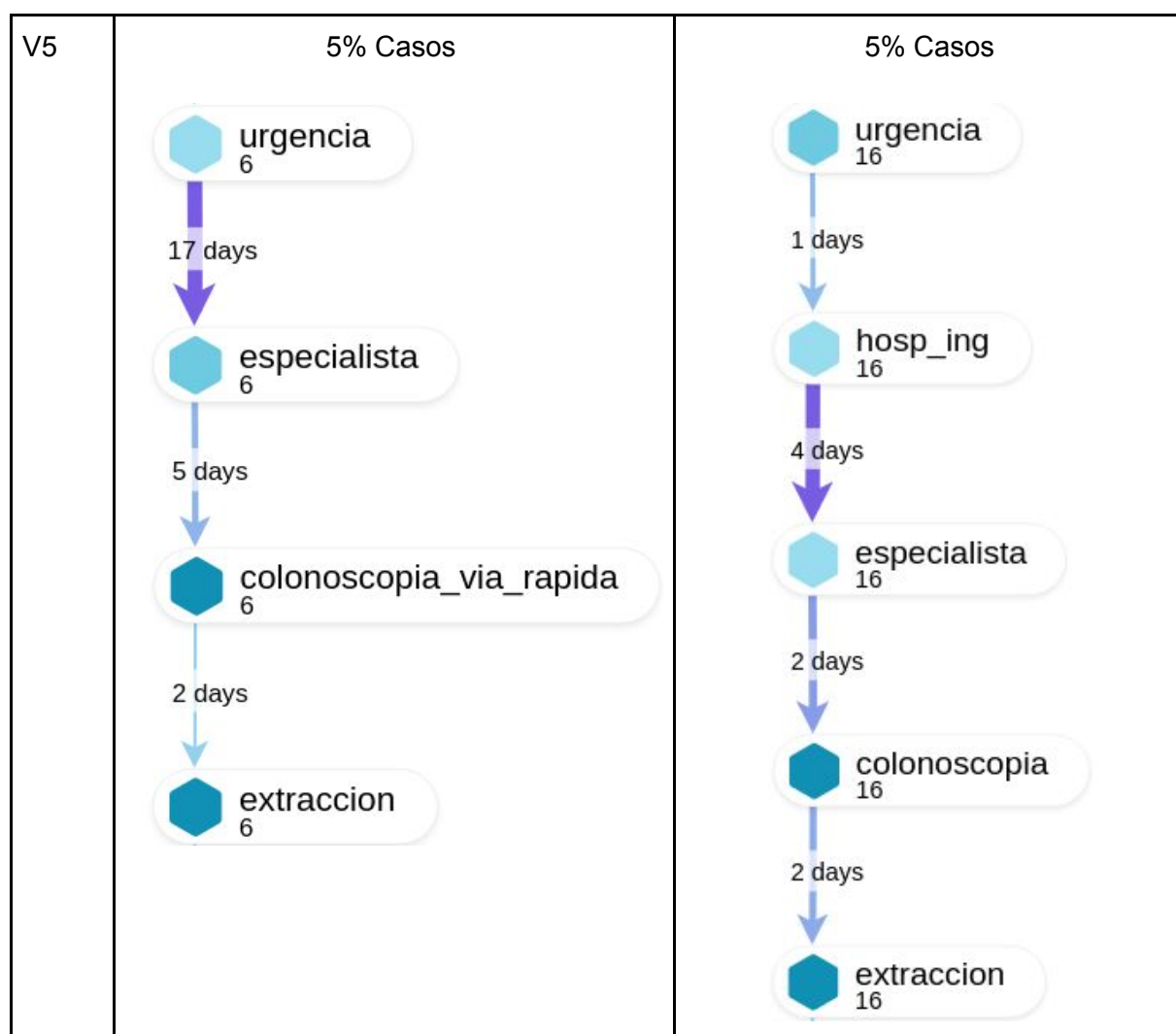


Fig25: Variantes de trazas simplificadas de los procesos de pacientes previos 90 días a la extracción

## 2.14 Términos clínicos seleccionados, descartados y agrupaciones

Se realizará, en los próximos puntos, un **análisis de los términos** empleados en notas clínicas de los pacientes a estudiar.

Dada la heterogeneidad de estos términos, se decidió en un estudio preliminar, hacer una **agrupación** de términos que reflejaban conceptos similares, para facilitar su análisis. Por tanto se procedió a **agrupar los términos** de la siguiente manera. Los siguientes grupos de

términos fueron agrupados de forma que para cualquiera de los términos del grupo se presenta sólo el primer término:

- **alteración\_del\_ritmo\_intestinal**, alternancia\_del\_ritmo\_intestinal, cambio\_en\_ritmo\_intestinal, cambio\_en\_el\_ritmo\_intestinal
- **diarrea**, deposiciones\_blandas, deposiciones\_líquidas
- **dolor\_abdominal**, dolor\_en\_fid, dolor\_en\_fii, dolor\_en\_flanco\_derecho, dolor\_en\_flanco\_izquierdo, dolor\_en\_fosa\_iliaca\_derecha, dolor\_en\_fosa\_iliaca\_izquierda, dolor\_en\_hipogastrio, dolor, dolor\_en\_hemiabdomen\_derecho, dolor\_en\_hemiabdomen\_izquierdo
- **hemorragia\_digestiva\_baja**, hdb, hematoquecia, rectorragia, rectorragias, sangrado\_bajo, sangrado\_digestivo\_bajo, sangrado\_rectal, hemorragias
- **masa\_rectal**, tumoración rectal
- **pérdida de apetito**, hiporexia, anorexia
- **dolor\_rectal**, proctalgia, dolor\_anal
- **masa\_abdominal**, bulto\_abdominal, tumoración\_abdominal

Por otro lado, se descartaron de los posteriores estudios los siguientes términos:

- ccr, ccr\_familiar, estadio\_0, estadio\_i, estadio\_ii, estadio\_iii, estadio\_iv, let, m1, m2, mlh1, msh2, msh6, mx, myh, n0, n1, n2, nx, paf, pms2, sps, t1, t2, t3, t4, tis, tx, síndrome\_de\_lynch

## 2.15 Frecuencia términos agrupados

En primer lugar, se consideró interesante valorar, para los distintos pacientes, la **frecuencia de los términos** empleados en informes del paciente previos a la **extracción**.

Para ello, se lista el **número de apariciones** totales de cada uno de los términos presentes en los pacientes, clasificando estos según su pertenencia o no a la VDR. Para tener datos comparativos de frecuencia, **se divide este número total entre el número de pacientes**, para obtener por tanto una frecuencia por paciente, o cuantas veces por paciente, de media, aparece cada uno de los términos.

VDR			No VDR		
Término	Cantidad	Freq	Término	Cantidad	Freq
hemorragia_digestiva_baja	106	0,80	dolor_abdominal	676	2,17
anemia	99	0,75	anemia	434	1,39
dolor_abdominal	93	0,70	hemorragia_digestiva_baja	338	1,08
diarrea	52	0,39	vómitos	335	1,07
dolor_rectal	47	0,36	diarrea	281	0,90
metástasis	47	0,36	dolor_rectal	236	0,76
vómitos	41	0,31	estreñimiento	180	0,58
estreñimiento	39	0,30	astenia	165	0,53
astenia	28	0,21	metástasis	162	0,52
pérdida de apetito	28	0,21	pérdida de apetito	114	0,37
melenas	18	0,14	masa_abdominal	102	0,33
masa_abdominal	16	0,12	distensión	87	0,28
pérdida_de_peso	14	0,11	pérdida_de_peso	67	0,21
tenesmo	12	0,09	melenas	66	0,21
masa_rectal	9	0,07	ferropenia	47	0,15

Fig26: Tabla de frecuencia de términos agrupados



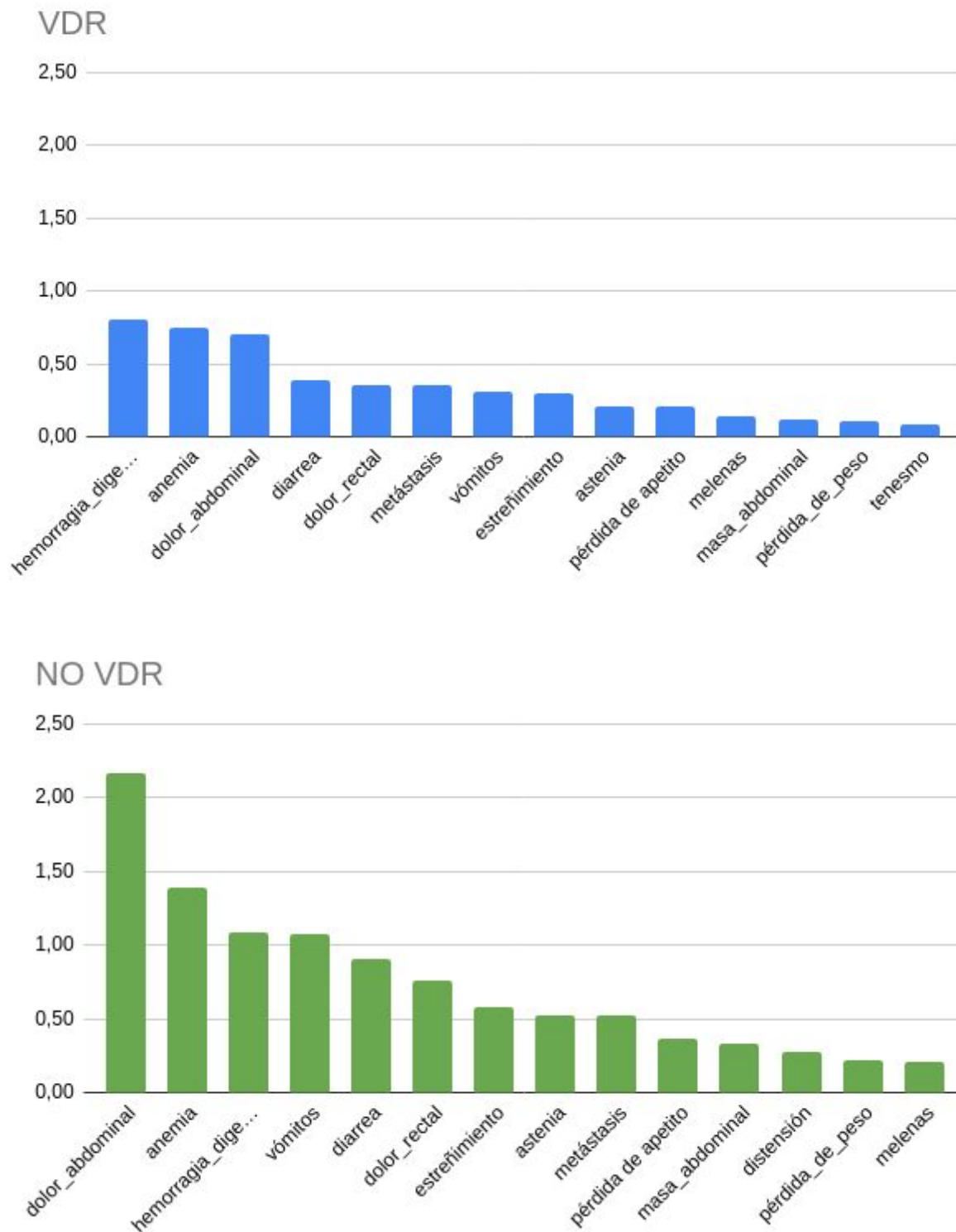


Fig27: Gráfica de frecuencia de términos agrupados

## 2.16 Presencia de términos agrupados

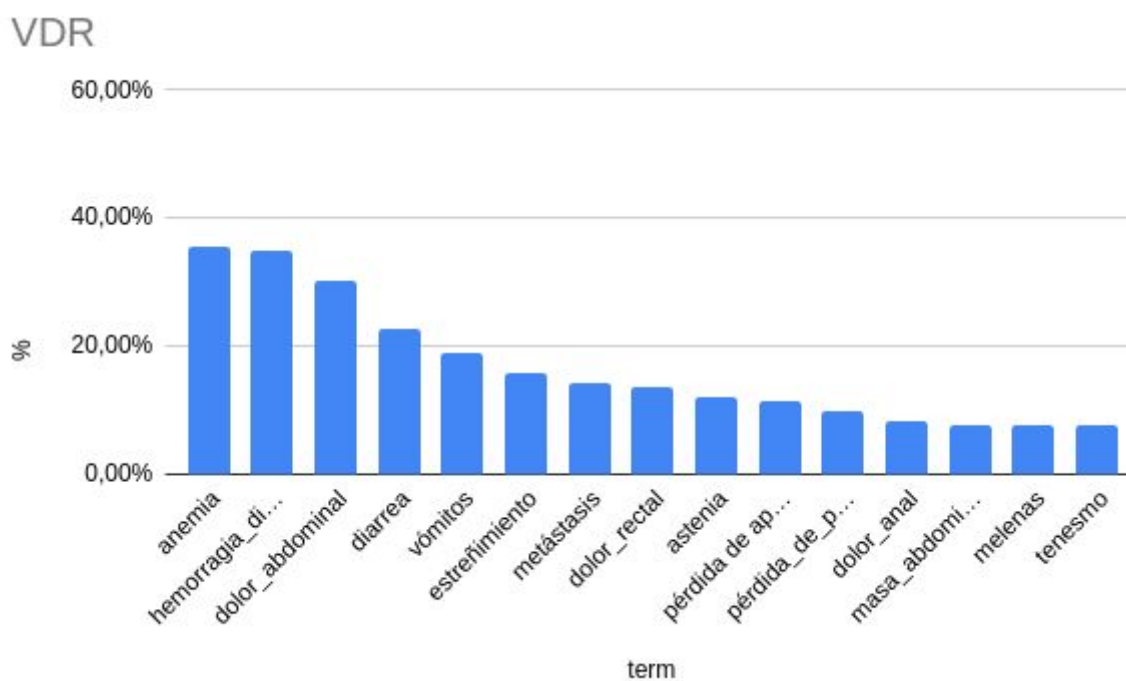
En este punto se realiza un estudio de la presencia de los términos para los distintos pacientes. En este caso, sin valorar la cantidad de veces, se estudia la **presencia** (o no) de los distintos **términos** en pacientes, de forma que se lista la **cantidad de pacientes** para los que se ha registrado la **presencia cierto término en al menos una ocasión** (una o más veces) cierto término. Se presentan diferenciando nuevamente los que están en VDR y los que no. Para hacer los resultados más comparables se presenta el % de pacientes que presentan el término.

A continuación se listan los más destacados:

VDR			No VDR		
Término	Cantidad	%	Término	Cantidad	%
anemia	47	35,61%	dolor_abdominal	174	55,77%
hemorragia_digestiva_baja	46	34,85%	anemia	141	45,19%
dolor_abdominal	40	30,30%	vómitos	137	43,91%
diarrea	30	22,73%	hemorragia_digestiva_baja	121	38,78%
vómitos	25	18,94%	diarrea	107	34,29%
estreñimiento	21	15,91%	estreñimiento	88	28,21%
metástasis	19	14,39%	astenia	82	26,28%
dolor_rectal	18	13,64%	dolor_rectal	76	24,36%
astenia	16	12,12%	metástasis	59	18,91%
pérdida de apetito	15	11,36%	pérdida de apetito	59	18,91%
pérdida_de_peso	13	9,85%	distensión	49	15,71%

dolor_anal	11	8,33%	masa_abdominal	49	15,71%
masa_abdominal	10	7,58%	dolor_anal	46	14,74%
melenas	10	7,58%	pérdida_de_peso	38	12,18%
tenesmo	10	7,58%	melenas	36	11,54%

Fig28: Tabla de aparición de términos agrupados



## NO VDR

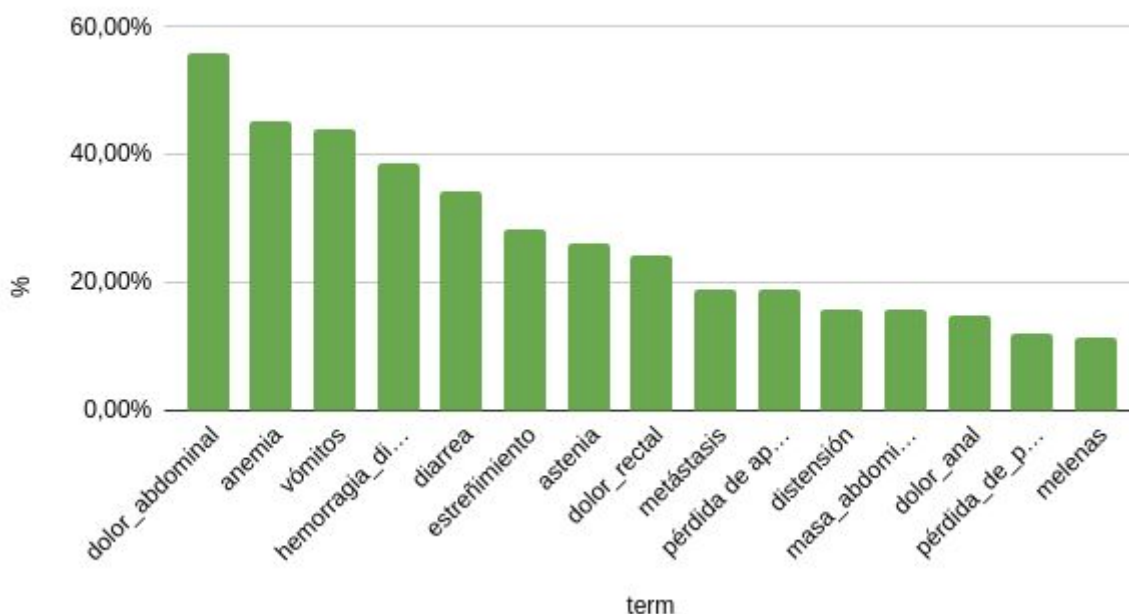


Fig29: Gráfica de aparición de términos agrupados

## 2.17 Temporalidad de términos agrupados

Otro punto a estudiar dentro de los términos registrados para los pacientes, es **el momento de aparición de cada término** dentro del proceso del paciente. ¿Qué términos aparecen antes? ¿Hay diferencias entre los pacientes de VDR y No VDR? Para ello se registra el **tiempo medio de la primera aparición**, dentro de los informes médicos, de cada uno de los **grupos** de términos. También se presenta el número de veces que este grupo aparece, para transmitir una idea del peso de este término dentro del proceso.

Total			VDR		No VDR	
término	número	media (h)	número	media (h)	número	media (h)
alteración_del_ritmo_intestinal	16	170,8	4	45,5	12	212,5
anemia	188	228,8	47,00	193,0	141	240,7
astenia	98	242,6	16	252,1	82	240,7

diarrea	137	351,2	30	355,3	107	350,1
distensión	55	140,8	6	242,7	49	128,3
dolor_abdominal	215	311,1	40	304,0	175	312,7
dolor_rectal	112	216,2	21	291,4	91	198,8
estreñimiento	109	282,6	21	261,5	88	287,6
ferropenia	37	219,5	4	151,5	33	227,8
hemorragia_digestiva_baja	169	230,0	46	175,4	123	250,4
masa_abdominal	62	97,8	10	98,7	52	97,6
masa_rectal	33	60,8	9	7,2	24	80.875,0
melenas	46	158,9	10	146,3	36	162,4
metástasis	78	269,0	19	189,7	59	294,5
obstrucción_intestinal	3	51,7	1	148,0	2	3,5
pérdida de apetito	74	136,7	15	174,1	59	127,2
pérdida_de_apetito	16	203.625,0	5	162,6	11	222,3
pérdida_de_peso	51	244,6	13	280,7	38	232,3
tenesmo	30	298,0	10	349,4	20	272,4
urgencia_defecatoria	3	52,7	1	7,0	2	75,5

Fig30: Tabla de tiempos de aparición de términos agrupados

A continuación se muestra un gráfico de burbujas, capturando en el **eje X** el **momento de aparición** (en media y en días) de cada término, y mostrando el **tamaño** de la burbuja proporcional al **número de pacientes** que lo presentaban.

Código de colores:

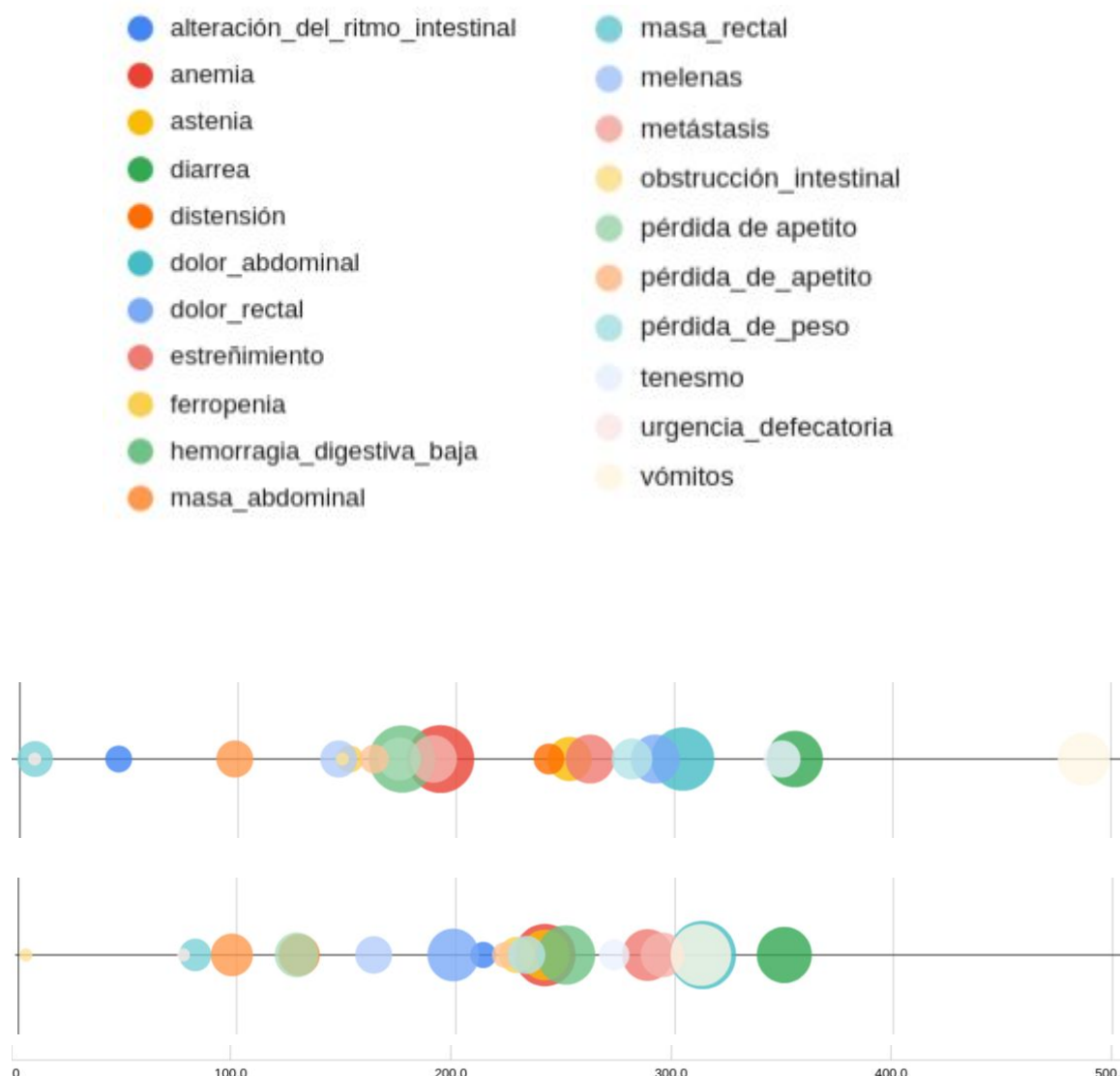


Fig31: Gráfico de burbujas de aparición de términos agrupados

A continuación se presenta otro gráfico, en el que se presentan **gráficos de tipo violín** para todos los términos agrupados, para las dos tipologías de pacientes.

La columna de la izquierda presenta los distintos términos, la central los pacientes VDR y la derecha pacientes no VDR. En las distintas filas, se van representando cada uno de los términos, y las gráficas de tipo violín para los dos tipos de pacientes.

Un diagrama de violín se utiliza para visualizar la **distribución de los datos** y su **densidad de probabilidad**. Este gráfico es una combinación de un diagrama de cajas y bigotes y un diagrama de densidad girado y colocado a cada lado, para mostrar la forma de distribución

de los datos. La barra central representa el intervalo intercuartil, la barra negra fina que se extiende desde ella, representa el 95 % de los intervalos de confianza, y la línea vertical la mediana.

Se puede ver más claramente en el siguiente diagrama:

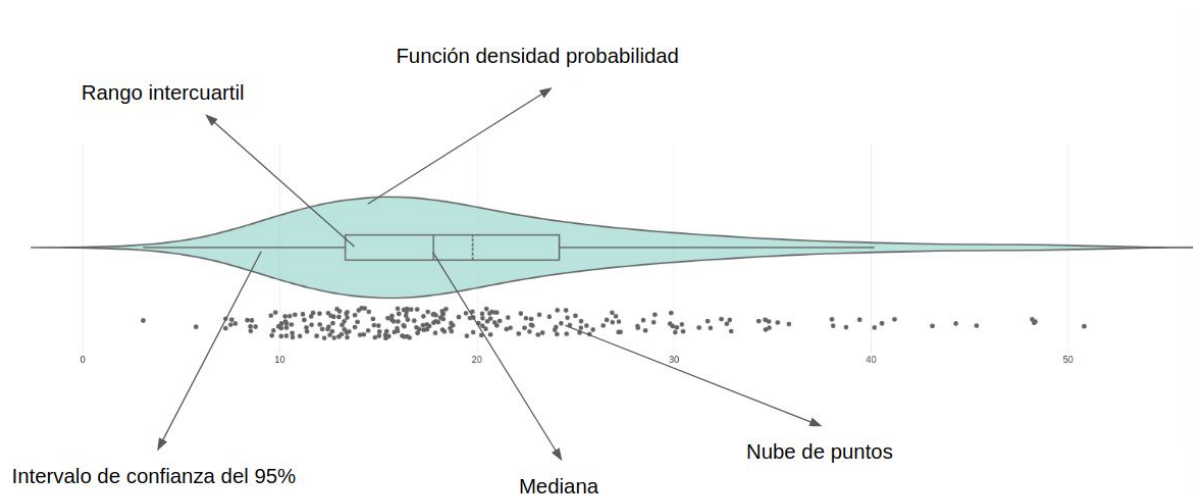


Fig32: Explicación gráfico violín

Se presenta a continuación el gráfico para los términos agrupados:

