

FUNDACION BBV

**ANÁLISIS DE LA INCIDENCIA DE LAS
AUSENCIAS DEL TRABAJO EN DOS
COHORTES DE TRABAJADORES DE
TRANSPORTE URBANO**

Marc Sáez

María Antonia Barceló

CENTRO CIENCIA, TECNOLOGIA Y SOCIEDAD

Documento de Trabajo
de la Fundación BBV



FUNDACION BBV

**ANÁLISIS DE LA INCIDENCIA DE LAS
AUSENCIAS DEL TRABAJO EN DOS
COHORTES DE TRABAJADORES DE
TRANSPORTE URBANO**

Marc Sáez

María Antonia Barceló

Marzo 1999

**ANÁLISIS DE LAS INCIDENCIA DE LAS AUSENCIAS DEL
TRABAJO EN DOS COHORTES DE TRABAJADORES DE
TRANSPORTE URBANO**

PRESENTACIÓN

Al igual que en el resto de países desarrollados, en España el sistema sanitario, pese a haber experimentado importantes reformas en la década de los 80, se enfrenta a graves problemas estructurales, originados por el crecimiento de la demanda sanitaria, la presión de los costes y la incorporación de nuevas tecnologías, entre otras razones.

Las ausencias del trabajo, además de constituir un problema sanitario, también son un problema socioeconómico ya que conllevan una serie de prestaciones sanitarias y económicas financiadas por los presupuestos de la Seguridad Social.

La Fundación BBV, continuando con su propósito de contribuir científicamente a las cuestiones relacionadas directa o indirectamente con el empleo, presenta ahora esta investigación, que analiza tres teorías explicativas de las ausencias del trabajo: un enfoque de trabajo-ocio; la versión de “disciplina laboral” de la teoría del salario de eficiencia; y las “normas de grupos de trabajo”. Estas tres teorías intentan explicar la ausencia voluntaria, aquella atribuible al deseo del trabajador, frente a la involuntaria, consecuencia de la imposibilidad de asistir al trabajo.

El estudio ha sido realizado por los profesores Marc Sáez y María Antonia Barceló, investigadores destacados del Centre de Investigació en Economia y Salud (CRES), de la Universidad Pompeu Fabra de Barcelona, a quienes la Fundación BBV agradece su aportación a la búsqueda de alternativas para adoptar de mayor eficiencia y equidad al sistema sanitario español.

Fundación BBV



CENTRE DE RECERCA
EN ECONOMIA I SALUT · CRES
UNIVERSITAT POMPEU FABRA

ANÁLISIS DE LA INCIDENCIA DE LAS AUSENCIAS DEL TRABAJO EN DOS COHORTES DE TRABAJADORES DE TRANSPORTE URBANO

Marc Sáez

Departament d'Economia, Universitat de Girona
Miembro del CRES

María Antonia Barceló

Departament d'Economia, Universitat de Girona

RESUMEN

Introducción: Las tres teorías explicativas de las ausencias del trabajo: un enfoque de trabajo-ocio; la versión de 'disciplina laboral' de la teoría del salario de eficiencia; y las 'normas de grupos de trabajo', intentan explicar la ausencia 'voluntaria', aquella atribuible al deseo del trabajador, frente a la 'involuntaria', consecuencia de la imposibilidad de asistir al trabajo. Los datos, sin embargo, no permiten distinguirlas, implicando un escaso poder explicativo en los modelos empíricos. Desde el punto de vista estadístico, se omiten variables relevantes: relacionadas con la salud de los individuos y otras variables socio-demográficas (sexo, edad, nivel de estudios, etc) cuyo efecto ha sido menos fundamentado teóricamente. En este trabajo contrastamos la hipótesis que la incidencia de las ausencias depende de factores socioeconómicos, en particular los relacionados con el puesto de trabajo.

Métodos: Utilizamos una muestra formada por todos los trabajadores contratados de forma ininterrumpida, entre el 1 de Enero de 1994 y el 31 de Diciembre de 1996, en una de las dos compañías constitutivas de una empresa de transporte urbano. El análisis lo realizamos separadamente por compañía y contingencia: accidente y enfermedad. La variable dependiente analizada la constituyen el número de los episodios de Incapacidad

Temporal (IT). El análisis estadístico se ha basado en la aproximación de Andersen-Gill de la regresión de riesgos proporcionales de Cox. Se intentó corregir la desventaja de tal aproximación utilizando diversos modelos marginales y condicionales, en particular el modelo condicional AG-Autoregresivo.

Resultados: La ocupación ha resultado ser la variable con mayor capacidad explicativa de la incidencia de las ausencias del trabajo. Los mecánicos, los operarios de vías, los operarios de mantenimiento y los conductores presentan los mayores riesgos, incluso tras controlar las características personales y ocupacionales de los trabajadores, sus conductas respecto a la salud y sus antecedentes sanitarios. Las mujeres, aquellos trabajadores mayores de cuarenta años, con una antigüedad en la compañía menor a nueve años, los trabajadores casados con más de un hijo, aquéllos que trabajan en turno más incómodo, aquéllos con menor nivel de estudios, ex-fumadores o fumadores y los que padecen alguna una enfermedad, son los que presentan mayor probabilidad de padecer un episodio de IT.

Conclusiones: Debido a la estructura de protección social de nuestro país, es bastante probable que algunos episodios de enfermedad se correspondan en realidad con accidentes y/o enfermedades profesionales. Esto nos impide separar desde un punto de vista práctico las ausencias 'voluntarias' de aquellas 'involuntarias'. La interpretación de los riesgos estimados, por tanto, no podrá ser completamente fundamentada atendiendo a una de las tres teorías económicas explicativas de las ausencias del trabajo. Creemos que nuestros hallazgos respecto a ocupación podrían ser explicados recurriendo principalmente a la comodidad del puesto de trabajo, o más rigurosamente a unas condiciones de trabajo más adversas, sobre todo desde un punto de vista físico y no tanto desde una perspectiva psicológica.

Palabras clave: episodios de Incapacidad Temporal, ocupación, Andersen-Gill, aproximaciones marginal y condicional.

ÍNDICE

PRESENTACIÓN.	
1. INTRODUCCIÓN.	1
2. MÉTODOS.	3
2.1. ANÁLISIS DESCRIPTIVO.	4
2.2. REGRESIÓN DE POISSON.	6
2.3. APROXIMACIÓN DE ANDERSEN-GIL CON ESTIMACIÓN ROBUSTA DE LA MATRIZ DE COVARIANZAS.	10
2.4. EL MÉTODO WLW.	15
2.5. EL MÉTODO PWP.	18
3. APROXIMACIONES CONDICIONALES. EL MODELO AG-AUTORREGRESIVO.	23
4. RESULTADOS.	27
5. CONCLUSIONES.	31
6. REFERENCIAS BIBLIOGRÁFICAS.	35
7. ANEXO METODOLÓGICO.	39

1.- INTRODUCCIÓN.

Las tres teorías económicas más utilizadas en la explicación de las ausencias del trabajo: un enfoque de trabajo-ocio; la versión de 'disciplina laboral' de la teoría del salario de eficiencia; y las 'normas de grupos de trabajo' (véase Saez y Barceló (1998a) para una revisión del tema) intentan explicar la ausencia 'voluntaria', aquella atribuible al deseo del trabajador, frente a la 'involuntaria', consecuencia de la imposibilidad de asistir al trabajo. Los datos, sin embargo, no permiten distinguir tales categorías de las ausencias, lo que suele implicar que los modelos empíricos presenten un escaso poder explicativo. Desde el punto de vista estadístico, los modelos omiten variables relevantes: relacionadas con la salud de los individuos y otras variables socio-demográficas (sexo, edad, nivel de estudios, etc) con un efecto menos fundamentado teóricamente.

En este trabajo pretendemos contrastar, como principal hipótesis, que la incidencia de las ausencias de trabajo depende de factores socioeconómicos, en particular los relacionados con la ocupación. Entendemos por incidencia el número de episodios de ausencia del trabajo por trabajador, en un periodo de tiempo determinado, generalmente el correspondiente al periodo de seguimiento de tal trabajador. Nuestro primer objetivo es pues el de contrastar si cuanto más baja es la categoría laboral del trabajador y peores son sus condiciones de trabajo, mayor es el número de episodios de ausencias del trabajo.

Por otra parte, la mayoría del análisis de las ausencias del trabajo ha prestado poca atención al hecho de que los episodios de ausencia no se presentan de manera única para todos los trabajadores, sino que algunos de ellos pueden presentar recurrencias, es decir, más de un episodio en el periodo de seguimiento. Este hecho podría invalidar gran parte de las conclusiones derivadas de aquellos análisis, podríamos decir más estándar. Por este motivo otro de los objetivos de este trabajo es el de investigar el efecto de las recurrencias en las inferencias realizadas y, al mismo tiempo, el de comparar diversas estrategias de análisis alternativas.

2.- MÉTODOS.

Utilizamos las cohortes formadas por todos los trabajadores contratados de forma ininterrumpida, entre el 1 de Enero de 1994 y el 31 de Diciembre de 1996, en una de las dos compañías (la Compañía 1 y la Compañía 2) constitutivas de una empresa de transporte urbano. El periodo completo de seguimiento suma 1096 días y el número total de sujetos incluidos en el estudio fue de 5456 trabajadores (4990 hombres y 466 mujeres); 2885 trabajadores en la compañía 1 (2784 hombres y 101 mujeres) y 2571 en la compañía 2 (2206 hombres y 365 mujeres).

En nuestro país los registros fiables de las ausencias del trabajo se basan en un concepto de salud laboral como es el de la declaración de Incapacidad Temporal (IT). La IT se define como la situación en la que se encuentra el trabajador que no puede realizar su actividad por razón de una enfermedad común, un accidente no laboral, una enfermedad profesional o un accidente laboral (Ley 42/1994 de 30 de Diciembre de medidas fiscales, normativas y de orden social. Reforma normativa referente a la contingencia de incapacidad temporal. BOE 313, de 31 de Diciembre de 1994). El sistema de Seguridad Social garantiza el derecho a una prestación económica mientras dura la situación de IT. El documento administrativo que reconoce dicha prestación son los partes de IT cuyo control corresponde a un médico de atención primaria del sistema de Seguridad Social. Así pues, la variable dependiente que analizamos la constituyen tanto el número como la duración de los episodios de Incapacidad Temporal (definida tal y como recoge la normativa legal).

Como variables explicativas disponemos para cada individuo de información sobre su categoría laboral, sexo, edad, antigüedad en la empresa, número de hijos, estado civil, turno de trabajo al que pertenece, estudios que ha realizado y número de enfermedades crónicas que padece en la actualidad. Además de estas variables, disponemos de valoraciones sobre su estado de salud (auto-percepción de la salud), actividad física en el tiempo libre, hábito tabáquico y patrón de consumo de alcohol. Tenemos variables explicativas fijas en el tiempo (sexo), otras tiempo-dependientes, pero que se recogieron una sola vez durante el periodo de seguimiento (número de hijos, estado civil y estudios) y finalmente, variables tiempo-dependientes, tanto externas como internas. Las variables tiempo-dependientes externas son aquéllas que pueden ser predichas por un observador externo sin necesidad de que el individuo sea observado (edad y antigüedad, en particular). Las variables internas dependen de la observación del individuo en el periodo de seguimiento (ocupación, turno de trabajo, auto-percepción de la salud, actividad física en el tiempo libre, hábito tabáquico, patrón de consumo de alcohol y enfermedades crónicas, en nuestro caso). Las variables explicativas fueron categorizadas (véase por ejemplo **Tabla 1**).

Los análisis se realizaron separadamente para la Compañía 1 y la Compañía 2 y según el tipo de contingencia: enfermedad común (enfermedad a partir de ahora) o accidente laboral o no laboral y enfermedad profesional (accidente a partir de ahora). Nótese que los estratos son independientes. Ningún trabajador perteneció simultáneamente a las dos compañías. Tampoco se produjeron traslados de una a otra compañía en el periodo considerado. Por lo que respecta al tipo de contingencia, mientras los episodios de IT por enfermedad son notificados por un médico de atención primaria perteneciente al sistema de Seguridad Social, los partes por accidente laboral son extendidos por el servicio médico de la empresa o por el de una mutua o compañía de seguros de accidentes y enfermedad. Es de esperar por tanto, que, condicionados a sus variables explicativas (cuyos efectos es de esperar que no sean necesariamente los mismos) los episodios de accidentes sean independientes de los episodios de IT por enfermedad.

2.1.- ANÁLISIS DESCRIPTIVO.

Como instrumento descriptivo previo y con el objetivo de medir la frecuencia de episodios de IT, se calculó la tasa de IT por 100.000 personas-día para cada una de las variables consideradas en el estudio (véase **Tabla 1**). Hemos de señalar que en el cálculo de dichos descriptivos fijamos el valor de las variables tiempo-dependientes como el del primero observado para esa variable. Por eso creemos que estas tasas no deben ser interpretadas de forma rigurosa, mostrándose aquí con un propósito meramente descriptivo.

En la **Tabla 1** presentamos las tasas de IT por 100.000 trabajadores según diferentes variables estratificando por compañías (Compañía 1 y Compañía 2). La tasa total de incidencia de las ausencias del trabajo es casi el doble en la Compañía 2 que en la Compañía 1 (215,20 vs 129,60). Las diferentes categorías de la variable ocupación fueron ordenadas por sueldo y estatus socio-laboral. Las tasas de IT parecen aumentar en relación inversa a la categoría ocupacional, siendo máximas entre 'Mecánicos' (ambas compañías) y 'Motoristas' y 'Personal de estación' (Compañía 2, categorías inexistentes en la Compañía 1) y mínimas entre 'Directivos' (ambas compañías). Es de destacar el elevado número de episodios (no tanto la tasa) entre los 'Mandos' en la Compañía 2. Excepto aquellos trabajadores que declararon no tener estudios, las tasas disminuyen conforme aumenta el nivel de los mismos en la Compañía 2, presentando un comportamiento extraño en la Compañía 1. Igualmente, las tasas son mayores en aquellos turnos de trabajo más incómodos: 'Nocturno' sobre 'Diurno', siendo mayores en 'Partido/sin turno' y mayores las de Compañía 2 que las de la Compañía 1.

Las mujeres presentan tasas mayores en ambas compañías. Debe recordarse que los episodios de IT no incluyen baja maternal. Parece existir cierto gradiente en las tasas de IT según estado civil, 'Casado/a' - 'Soltero/a' - 'Separado/Divorciado', que no se cumple en el caso de aquellos trabajadores viudos. Como era de esperar las tasas aumentan conforme lo hace el número de hijos (excepto cuando se tienen dos hijos). Las relaciones entre edad y antigüedad y episodios de IT, no parecen ser lineales, presentando máximos en las categorías inferiores.

La relación entre episodios de IT y actividad física, tabaco, enfermedades crónicas y auto-percepción de la salud, es la que podría esperarse a priori. Por ejemplo, las tasas de IT aumentan con el número de enfermedades y con una percepción mala de la propia salud. Nótese, sin embargo, que aquéllos que declaran consumir alcohol diariamente presentan menores tasas.

Tabla 1. Distribución de la frecuencia de los episodios de incapacidad temporal según diferentes variables (al principio del periodo de seguimiento 1994-1996).

	Trabajadores		Episodios		Casos /100.000 trabajadores	
	Comp. 1	Comp. 2	Comp. 1	Comp. 2	Comp. 1	Comp. 2
TOTAL	2885	2571	3690	5456	129.60	215.2
OCUPACION						
Directivos	103	89	39	55	35.80	58.5
Mandos	148	211	152	328	96.12	149.4
Técnicos	183	97	189	131	97.80	129.1
Administrativos	52	136	40	208	72.41	143.67
Mecánicos	231	199	368	505	163.20	253.4
Electricistas	57	228	75	449	128.00	201.6
Conductores	1979	16	2646	37	139.20	214.5
Pers. Auxiliar	132	125	159	307	120.90	244.3
Pers. Estación ¹		904		2157		249.3
Motoristas ¹		435		965		255.1
Operarios vías ¹		98		247		225.6
Mantenimiento ¹		33		67		195.2
ESTUDIOS						
Sin estudios	556	557	647	1232	112.70	214.14
Primarios	1011	771	1398	1774	137.40	222.60
Secundarios	884	642	1182	1187	144.10	199.03
Universitarios	143	177	76	235	51.70	127.91
TURNO						
Diurno	669	1905	718	3988	103.60	200.37
Nocturno	126	271	171	597	135.30	216.10
Partido/sin turno	2090	395	2779	871	137.50	251.80
SEXO						
Hombres	2784	2206	3500	4487	127.90	201.30
Mujeres	101	365	168	969	166.00	282.70
EST. CIVIL						
Soltero	442	425	508	789	135.50	211.90
Casado	2356	2061	3024	4418	127.20	209.10
Sep/divorciado	70	58	119	185	174.70	325.80
Viudo	17	28	17	64	101.24	218.83
NUM. HIJOS						
1 hijo	639	419	819	948	129.90	221.80
2 hijos	1046	1011	1339	2052	124.80	195.80
Más 2 hijos	532	500	746	1196	138.10	232.85
Ninguno	668	641	764	1260	128.60	216.00
EDAD						
<30	448	404	555	778	152.60	224.90
31-40	837	278	1191	592	145.50	213.80
41-50	597	915	726	2068	115.50	217.86
51-60	821	858	1060	1836	127.30	208.21
>60	182	116	136	182	70.22	154.42
ANTIGUEDAD						
<7	992	515	1276	964	147.00	215.80
7-19	638	650	956	1495	142.41	220.31
20-26	579	938	740	1941	123.30	199.63
>26	676	468	696	1056	99.98	222.90
ACT. FISICA						
Ninguna	1434	1383	1902	3156	131.30	224.50
Regularmente	1205	889	1473	1654	127.80	189.10
ALCOHOL						
Diario	767	956	927	1879	121.30	195.40
Semanal	1069	1330	995	1898	132.70	202.80
No bebe	805	1098	393	1039	135.20	270.50
TABACO						
No fuma	1272	1043	1521	1862	119.80	178.60
Ex-fumador	242	285	333	585	143.00	200.90
Fumador	1130	951	1533	2379	138.90	249.90
ENFERMEDADES CRÓNICAS						
0	1543	1662	1697	3299	114.50	200.80
1	517	515	746	1105	146.90	208.60
Más de 1	825	394	1225	1052	144.50	264.13
AUTO-PERCEPCIÓN DE SALUD						
Buena/Muy B.	2610	2229	3300	4625	128.10	206.60
Mala/Regular	27	37	78	180	334.23	534.19

¹ Estas categorías no existen en la compañía 1

2.2.- REGRESIÓN DE POISSON.

En la **Tabla 2** mostramos el número de episodios de ausencia por trabajador en ambas compañías. Como se observará las ausencias son recurrentes en algunos de los trabajadores. El 35.5% de los mismos en la Compañía 1 y el 49% en la Compañía 2 presentaron más de un episodio en el periodo de seguimiento. Por otra parte, el número de episodios de IT puede ser considerado una variable aleatoria Poissoniana, es decir una variable discreta que toma únicamente valores enteros positivos y cuyo rango no es muy amplio. El 98.2% de los trabajadores de la Compañía 1 y el 91.5% de los de la Compañía 2 tuvieron como máximo 5 episodios de IT.

Tabla 2. Número y porcentaje de número de episodios de Incapacidad Temporal por trabajador, en la Compañía 1 y en la Compañía 2 (1994-96).

Número de episodios	Compañía 1		Compañía 2	
	Número	%	Número	%
0	1059	36,7	712	27,7
1	831	28,8	598	23,3
2	518	18,0	450	17,5
3	244	8,5	255	9,9
4	114	4,0	198	7,7
5	68	2,4	116	4,5
6	21	0,7	93	3,6
7	9	0,3	52	2,0
8	9	0,3	30	1,2
9	5	0,2	15	0,6
10	2	0,1	22	0,9
11 ó más	5	0,2	30	1,2
Total	2885	100	2571	100

El número de episodios IT registrado por cada individuo, Y_i (denotando por i al individuo, $i=1, \dots, 2909$ en la compañía 1 y $i=1, \dots, 2572$ en la compañía 2) se generará por una distribución de Poisson con parámetro μ_i (el número esperado de episodios de IT para cada individuo). La probabilidad que un individuo registre exactamente y_i episodios de IT (siendo $y_i = 0, 1, 2, \dots$) puede expresarse:

$$\text{Pr ob}(Y_i = y_i) = \frac{e^{-\mu_i} \mu_i^{y_i}}{y_i!}$$

En nuestro caso, el número (medio) de episodios de IT, μ_i , dependerá de p variables explicativas, x_{ij} ($j=1,2,\dots,p$). De esta forma, estimaremos un modelo de regresión de Poisson:

$$\text{Ln}(\mu_i) = \beta_0 + \sum_{j=1}^p \beta_j x_{ij}$$

en el cual se asume que la esperanza condicionada es igual a la varianza condicionada, es decir, $E(y_i | x_{ij}) = \text{Var}(y_i | x_{ij}) = \mu_i = e^{\beta_0 + \sum \beta_j x_{ij}}$.

Cuando el periodo de seguimiento no sea el mismo para todos los individuos (como en nuestro caso, puesto que algunos individuos permanecieron 1096 días en observación y otros 1096 menos los días en los que estuvieron en situación de IT) el modelo debe reparametrizarse expresando, $\mu_i = \lambda_i t_i$, siendo λ_i el número esperado de episodios de IT por unidad de tiempo y t_i el período de seguimiento. De hecho, se trata simplemente de introducir $\ln(t_i)$ como offset en la regresión de Poisson. Resumiendo, la variable dependiente de la regresión será el número de episodios de IT, introduciendo como offset el logaritmo del período de seguimiento en días (1096 menos el número de días en situación de baja laboral por IT).

$$\text{Ln}(\mu_i) = \beta_0 + \sum_{j=1}^p \beta_j x_{ij} + \text{Ln}(t_i)$$

En la **Tabla 3** mostramos los resultados de las estimaciones de los modelos de Poisson, los cuales, en términos generales, son acordes con los descriptivos previos.

Tabla 3a. Resultados de la estimación de modelos de regresión de Poisson.

Accidentes	Compañía 1			Compañía 2		
	RR	95% C.I.		RR	95% C.I.	
Termino independiente	0,0000036	0,0000003	0,0000389	0,0000102	0,0000023	0,0000355
Sexo (hombre)	1,738	1,045	2,890	1,548	1,248	1,921
Edad (<=30)						
31-40	1,274	0,961	1,688	1,483	1,196	1,839
41-50	1,355	0,852	2,154	1,727	1,258	2,370
51-60	1,575	0,942	2,633	1,719	1,223	2,418
>60	1,621	0,733	3,587	1,337	0,773	2,313
Antigüedad (<9)						
10-20	0,765	0,524	1,118	0,648	0,501	0,837
21-25	0,954	0,624	1,459	0,707	0,528	0,944
>25	0,686	0,428	1,099	0,692	0,511	0,937
Numero de hijos (1 hijo)						
2 hijos	1,321	1,060	1,645	1,144	0,983	1,330
Más de 2	1,218	0,924	1,607	1,056	0,884	1,263
Estado civil (soltero/a)						
Casado/a	1,200	0,841	1,710	1,203	0,959	1,508
Separ/divorciado	1,051	0,586	1,886	1,280	0,875	1,874
Viudo/a	0,572	0,106	3,087	0,836	0,437	1,597
Ocupación (Directivos)						
Mandos	29,075	2,556	330,775	12,305	3,708	40,839
Técnicos	7,678	0,775	76,098	4,003	1,159	13,820
Administrativos	16,547	1,338	204,659	5,628	1,621	19,537
Mecánicos	81,741	7,285	917,182	25,025	7,519	83,285
Electricistas	50,613	4,358	587,779	15,253	4,532	51,339
Conductores	35,381	3,103	403,376	13,236	3,917	44,726
Personal auxiliar	14,728	1,247	173,899	11,881	3,537	39,916
Personal estación				9,553	2,861	31,903
Motoristas				10,419	3,103	34,984
Operarios vías				20,498	5,877	71,489
Operarios manten.				25,256	7,055	90,410
Turno (diurno)						
Nocturno	1,013	0,723	1,420	0,885	0,722	1,085
Partido	1,478	0,513	4,262	0,660	0,464	0,939
Rotativo/sin turno	0,836	0,574	1,217	0,628	0,488	0,809
Estudios (universitarios)						
Sin estudios	1,001	0,399	2,508	1,429	0,883	2,312
Primarios	1,259	0,515	3,078	1,747	0,868	2,472
Secundarios	1,578	0,652	3,822	1,567	0,988	2,485
Perc. Salud (MB/B/R)						
Mala/muy mala	1,298	0,713	2,365	1,759	1,297	2,384
Actividad física (No)						
Habitualmente	1,047	0,888	1,235	1,094	0,978	1,225
Tabaco (no fuma)						
Ex-fumador	0,955	0,697	1,307	1,118	0,917	1,363
Fumador	1,229	1,035	1,459	1,295	1,150	1,458
Alcohol (no bebe)						
Semanal/espórádico	0,923	0,755	1,130	1,041	0,901	1,202
Diariamente	0,928	0,755	1,142	0,942	0,812	1,094
Enfermedades (ninguna)						
Una	1,097	0,919	1,311	1,024	0,913	1,150
Más de una	1,308	0,842	2,032	1,411	1,049	1,899
Deviance (dí)	2563,21 (2868)			5410,705 (5436)		
Parametro dispersión ϕ (p)	1,3898 (0,000)			1,4083 (0,000)		

En negrita test de Wald mayor que 1, aunque no significativo

Tabla 3b. Resultados de la estimación de modelos de regresión de Poisson.

Enfermedad	Compañía 1			Compañía 2		
	RR	95% C.I.		RR	95% C.I.	
Término independiente	0.0000623	0.0000355	0.0001093	0.000273	0.000192	0.000389
Sexo (hombre)	1.770	1.419	2.208	1.556	1.389	1.742
Edad (<30)						
31-40	1.486	1.285	1.717	1.294	1.140	1.470
41-50	1.374	1.075	1.756	1.254	1.044	1.505
51-60	1.629	1.240	2.139	1.260	1.033	1.536
>60	1.701	1.118	2.590	1.297	0.960	1.753
Antigüedad (<9)						
10-20	0.619	0.508	0.754	0.750	0.648	0.867
21-25	0.984	0.788	1.227	0.937	0.793	1.106
>25	0.521	0.406	0.669	0.750	0.628	0.895
Número de hijos (1 hijo)						
2 hijos	0.900	0.797	1.017	0.913	0.833	1.000
Más de 2	1.085	0.938	1.256	1.062	0.955	1.180
Estado civil (soltero/a)						
Casado/a	1.006	0.850	1.191	1.035	0.911	1.175
Separ/divorciado	1.094	0.823	1.453	1.324	1.075	1.630
Viudo/a	0.931	0.507	1.709	0.879	0.618	1.251
Ocupación (Directivos)						
Mandos	6.519	3.606	11.781	3.224	2.384	4.360
Técnicos	2.708	1.739	4.216	2.034	1.500	2.758
Administrativos	6.773	3.624	12.658	2.183	1.580	3.016
Mecánicos	7.063	3.953	12.620	2.372	1.723	3.267
Electricistas	5.292	2.709	10.336	2.842	2.045	3.949
Conductores	10.222	5.722	18.261	3.700	2.670	5.127
Personal auxiliar	7.035	3.920	12.626	2.982	2.169	4.101
Personal estación				3.266	2.403	4.439
Motoristas				3.669	2.678	5.025
Operarios vías				3.768	2.564	5.536
Operarios manten.				2.389	1.438	3.968
Turno (diurno)						
Nocturno	1.221	0.985	1.513	0.950	0.829	1.088
Partido	4.398	3.026	6.393	1.176	1.011	1.369
Rotativo/sin turno	0.766	0.627	0.936	0.550	0.472	0.639
Estudios (universitarios)						
Sin estudios	1.265	0.880	1.818	1.135	0.925	1.393
Primarios	1.321	0.935	1.867	1.217	1.002	1.479
Secundarios	1.311	0.933	1.841	1.163	0.963	1.405
Perc. Salud (MB/B/R)						
Mala/muy mala	1.688	1.250	2.280	1.567	1.283	1.913
Actividad física (No)						
Habitualmente	0.971	0.886	1.065	0.921	0.859	0.987
Tabaco (no fuma)						
Ex-fumador	1.160	0.987	1.364	1.271	1.127	1.433
Fumador	1.140	1.035	1.257	1.474	1.369	1.587
Alcohol (no bebe)						
Semanal/espórádico	0.931	0.832	1.041	0.870	0.799	0.949
Diariamente	0.968	0.861	1.089	0.868	0.794	0.950
Enfermedades (ninguna)						
Una	1.384	1.254	1.528	1.224	1.140	1.313
Más de una	1.914	1.521	2.409	1.581	1.318	1.896
Deviance (df)	3835.778 (2868)			9352.727 (5436)		
Parametro dispersión ϕ (p)	1.3639 (0.000)			1.8548 (0.000)		

En negrita test de Wald mayor que 1, aunque no significativo

En la estimación de un modelo de regresión de Poisson estándar, sin embargo, se suelen cometer importantes errores de especificación. Estos se manifiestan principalmente como sobre-dispersión, es decir las varianzas condicionales son mayores que las nominales (véanse por ejemplo los trabajos de Taylor y Pocock, 1992; North et al., 1993; Marmot et al., 1995; North et al., 1996). De hecho se incumplen las hipótesis básicas del modelo, a saber: independencia de las observaciones de la variable dependiente y un número esperado de sucesos que no varía durante el periodo de seguimiento.

En nuestro caso es fácil suponer que el número de episodios de IT que pueda padecer un individuo no dependa del que pueda padecer otro individuo distinto (siempre y cuando se considere *condicionado* a las variables explicativas, es decir no exista ninguna variable explicativa omitida que pueda provocar tal dependencia). Sin embargo, se incumple claramente la segunda hipótesis. El periodo de seguimiento es demasiado largo (1994-1996) como para suponer que el riesgo (el número esperado de episodios de IT por periodo de tiempo) haya permanecido constante. Además y en relación a esta hipótesis podemos señalar, al menos, dos limitaciones más. Nótese que, con independencia a su duración, la regresión de Poisson pondera de la misma forma todos los episodios de IT que pueda haber padecido un trabajador. Se podría esperar que episodios de muy distinta duración pudiesen tener factores explicativos diferentes. Por otra parte no se aprovecha la naturaleza longitudinal del análisis. De hecho se tratan todas las variables explicativas, incluso las tiempo-dependientes, como fijas en un punto del tiempo predeterminado (en nuestro caso el que corresponde a la primera observación) lo que sin duda, podría sesgar las estimaciones.

La significación del contraste de Breslow (1996), sugiere que los modelos estimados presentan una elevada sobre-dispersión, síntoma de la existencia de importantes errores de especificación. Este hecho nos impide interpretar los resultados.

2.3.- APROXIMACIÓN DE ANDERSEN-GILL CON ESTIMACIÓN ROBUSTA DE LA MATRIZ DE COVARIANZAS.

Disponemos de un *diseño longitudinal* (un *panel de datos*), observaciones para cada individuo y repetidas en el tiempo, sobre la variable dependiente y las explicativas. Aprovechando este tipo de diseño, hemos optado por analizar la incidencia de las ausencias del trabajo recurriendo al *análisis de supervivencia* o en términos econométricos *análisis de duración* (véanse Barmby et al., 1991; van Ours y Ridder, 1991; McCall, 1994). En nuestro contexto, dicho análisis intentaría determinar los factores explicativos del *tiempo de supervivencia* de los episodios de IT, el que transcurre desde el primer hasta el último día de baja por IT (suceso de interés).

Como es sabido, los métodos convencionales de dicho análisis se basan en el modelo (o regresión) de riesgos proporcionales de Cox (1972) (véanse los libros de Kalbfleisch y Prentice (1980) y Cox y Oakes (1984) o el capítulo 22.5 de Greene (1993)). Tal análisis no es directamente aplicable en nuestro caso. Los métodos estándar suponen por un lado que no existen *recurrencias*, es decir que cada individuo experimenta, como máximo, una ocurrencia del suceso de interés durante el periodo de seguimiento (claramente no es nuestro caso). Por otra parte supone que los tiempos de supervivencia de diferentes

individuos son independientes entre sí. Además, puesto que la mayoría de variables explicativas son dependientes en el tiempo, los riesgos no son proporcionales, lo que nos impide utilizar los métodos estándar. Por último, nosotros pretendemos no tanto determinar los factores explicativos de la duración de los episodios de ausencia sino aquéllos que pudiesen explicar la incidencia de las mismas.

Todas estas razones obligan a utilizar de forma alternativa alguna generalización del análisis que permita la recurrencia de los sucesos de interés. Utilizaremos la aproximación de Andersen-Gill (1982) al modelo de riesgos proporcionales de Cox como proceso contador (*AG* a partir de ahora). En la aproximación *AG* cada individuo es tratado como una observación de un proceso de Poisson (muy lento). Un individuo censurado por la derecha (un trabajador que no experimentó ningún episodio de IT en nuestro caso) no se considera un *dato incompleto* como en el enfoque tradicional, sino un individuo cuyo contador del suceso de interés es todavía cero.

De hecho, los modelos de procesos contadores, las regresiones logísticas y las regresiones de Poisson están muy relacionados. Todos estos métodos intentan estimar la probabilidad de que se produzca el resultado de interés en cada uno de un conjunto de experimentos de Bernoulli (variables aleatorias discretas que toman únicamente dos valores: éxito o fracaso). Que se escoja un método u otro en el caso de recurrencias, dependerá en gran medida de la frecuencia de los sucesos recurrentes. Cuando el número de sucesos (recurrentes) por individuo sea relativamente pequeño y el riesgo (o probabilidad) de que el suceso se produzca, varíe en los periodos entre recurrencias (como es presumiblemente nuestro caso) deberían utilizarse los métodos de análisis de supervivencia apropiados (procesos contadores, por ejemplo). Por otra parte, cuando los sucesos sean relativamente frecuentes en los individuos y el riesgo sea más o menos constante entre recurrencias, deberían utilizarse modelos de Poisson (Clayton, 1994).

Sin embargo, la hipótesis clave de *AG* es que las múltiples observaciones de un mismo individuo son independientes (condicionado a las variables explicativas), hipótesis denominada de *incrementos independientes*. De otra forma, para un individuo determinado el riesgo de la ocurrencia de un suceso no viene determinado por la ocurrencia de sucesos anteriores en el mismo individuo, a no ser que tal dependencia sea incluida explícitamente en el modelo como variable explicativa (por ejemplo variables dependientes en el tiempo que pudiesen inducir tal dependencia). Esta hipótesis es demasiado restrictiva. En la práctica esta limitación puede conducir a estimadores sesgados e ineficientes. Además, se suele sobreestimar la precisión de los estimadores puesto que, habitualmente, las observaciones de un mismo individuo suelen estar positivamente correlacionadas.

La dependencia de las observaciones puede entenderse como un 'estorbo', como algo a eliminar o, en todo caso, controlar sus efectos en el análisis, pero no como objetivo en sí. Por el contrario, la dependencia puede ser también de interés por sí misma. Por ejemplo podría desearse estimar en que cuantía se incrementa la probabilidad de padecer un episodio de IT cuando ya se padeció uno previo (o quizás más de uno). Mientras el primer argumento conduce a los modelos denominados *marginales*, el último se aproxima mediante los modelos llamados *condicionales* (Clayton, 1994).

A grandes rasgos los modelos marginales estiman el modelo ignorando la posible dependencia entre las observaciones, aunque corrigen posteriormente la varianza 'naive' (sesgada) mediante estimadores jackknife, bootstrap o 'sandwich'. Los modelos 'condicionales' estiman la dependencia (o la heterogeneidad) especificando explícitamente la distribución de probabilidad de la misma e incorporándola en el modelo.

La primera aproximación marginal que utilizamos es la de la estimación *robusta* de la matriz de covarianzas. Partimos del hecho que la aproximación AG proporciona estimadores consistentes de los parámetros asociados a las variables explicativas. Aún así, si existe dependencia, el método proporciona varianzas sesgadas. Se trataría, por tanto, de estimar de forma *robusta* la matriz de covarianzas, es decir de corregirla adecuadamente (Therneau y Hamilton, 1997).

En la **Tabla 4** mostramos los resultados de la estimación por la aproximación AG-estimación *robusta* de la matriz de covarianzas.

El problema de esta aproximación es que las estimaciones de los parámetros serán consistentes siempre y cuando el modelo marginal (según el cual se cumple la hipótesis de riesgos proporcionales para *cada una de las observaciones*, no necesariamente para cada uno de los individuos) esté correctamente especificado. Aún así, siguiendo a Therneau y Hamilton (1997), e incluso en el caso de que se omita erróneamente alguna variable relevante, esta aproximación proporciona los estimadores con menor sesgo respecto a otros enfoques marginales y condicionales alternativos.

Tabla 4a. Resultados de la estimación. AG- estimación robusta de la varianza.

Accidente	Compañía 1			Compañía 2		
	RR	95% C.I.		RR	95% C.I.	
Sexo (hombre)	2.090	1.178	3.709	1.360	1.046	1.769
Edad (<30)						
31-40	1.432	1.098	1.867	1.838	1.283	2.632
41-50	1.804	1.155	2.817	2.022	1.315	3.111
51-60	2.103	1.273	3.472	1.748	1.105	2.764
>60	1.863	0.708	4.901	1.215	0.492	2.996
Antigüedad (<9)						
10-20	0.612	0.398	0.942	0.381	0.261	0.557
21-25	0.939	0.598	1.476	0.427	0.284	0.642
>25	0.488	0.303	0.788	0.600	0.391	0.922
Numero de hijos (1 hijo)						
2 hijos	1.299	1.055	1.598	0.965	0.786	1.185
Más de 2	1.278	0.984	1.659	0.964	0.762	1.218
Estado civil (soltero/a)						
Casado/a	1.207	0.844	1.726	1.182	0.861	1.621
Separ/divorciado	1.018	0.568	1.827	1.538	0.838	2.821
Viudo/a	0.478	0.117	1.948	1.035	0.532	2.014
Ocupación (Directivos)						
Mandos	43.116	6.087	305.408	8.283	1.682	40.789
Técnicos	8.242	1.447	46.934	3.483	0.702	17.268
Administrativos	24.676	3.359	181.272	4.422	0.879	22.227
Mecánicos	121.047	16.856	869.280	14.474	2.922	71.696
Electricistas	66.683	8.963	496.085	9.082	1.827	45.155
Conductores	65.259	9.134	466.262	14.249	2.671	76.033
Personal auxiliar	20.356	2.788	148.613	10.864	2.163	54.573
Personal estación				5.714	1.1492	28.415
Motoristas				6.357	1.279	31.571
Operarios vías				12.608	2.473	64.274
Operarios manten.				13.410	2.609	68.923
Turno (diurno)						
Nocturno	0.973	0.686	1.381	0.790	0.613	1.018
Partido	2.786	0.827	9.389	0.392	0.233	0.659
Rotativo/sin turno	0.628	0.456	0.864	0.224	0.144	0.349
Estudios (universitarios)						
Sin estudios	1.029	0.475	2.233	1.531	0.791	2.962
Primarios	1.378	0.663	2.865	1.629	0.855	3.104
Secundarios	1.790	0.883	3.627	1.426	0.753	2.701
Perc. Salud (MB/B/R)						
Mala/muy mala	1.547	0.872	2.743	2.329	1.546	3.510
Actividad física (No)						
Habitualmente	1.056	0.899	1.241	1.069	0.917	1.246
Tabaco (no fuma)						
Ex-fumador	1.007	0.750	1.350	1.331	1.042	1.701
Fumador	1.208	1.015	1.438	1.470	1.250	1.728
Alcohol (no bebe)						
Semanal/espórádico	0.843	0.684	1.039	1.078	0.859	1.353
Diariamente	0.949	0.775	1.163	0.891	0.708	1.121
Enfermedades (ninguna)						
Una	1.117	0.932	1.339	1.004	0.859	1.173
Más de una	4.889	2.211	10.814	5.143	1.742	15.183
Deviance (df)	13697.37 (10161)			17444.89 (13259)		
Test de Razón de Verosimilitud (p)	389.00 (0.000)			517.000 (0.000)		

En negrita test de Wald mayor que 1, aunque no significativo

Tabla 4b. Resultados de la estimación. AG- estimación robusta de la varianza.

Enfermedad	Compañía 1			Compañía 2		
	RR	95% C.I.		RR	95% C.I.	
Sexo (hombre)	1.658	1.231	2.233	1.394	1.199	1.621
Edad (<30)						
31-40	1.372	1.180	1.595	0.954	0.846	1.299
41-50	1.357	1.055	1.746	0.978	0.759	1.258
51-60	1.563	1.190	2.053	0.955	0.725	1.257
>60	1.914	1.242	2.951	0.996	0.655	1.514
Antigüedad (<9)						
10-20	0.636	0.513	0.788	0.770	0.620	0.956
21-25	1.006	0.794	1.274	0.755	0.597	0.955
>25	0.480	0.371	0.621	0.976	0.765	1.245
Numero de hijos (1 hijo)						
2 hijos	0.949	0.839	1.073	0.886	0.772	1.019
Más de 2	1.193	1.031	1.381	1.000	0.851	1.175
Estado civil (soltero/a)						
Casado/a	1.050	0.858	1.285	0.988	0.818	1.194
Separ/divorciado	1.128	0.835	1.524	1.387	1.048	1.835
Viudo/a	0.805	0.470	1.378	0.845	0.487	1.468
Ocupación (Directivos)						
Mandos	5.703	3.286	9.899	2.167	1.511	3.108
Técnicos	2.630	1.713	4.037	1.860	1.249	2.771
Administrativos	6.154	3.370	11.237	1.705	1.177	2.470
Mecánicos	6.856	3.992	11.775	1.726	1.172	2.542
Electricistas	4.536	2.531	8.129	1.783	1.199	2.652
Conductores	8.896	5.205	15.203	2.063	1.109	3.838
Personal auxiliar	6.027	3.486	10.419	2.216	1.491	3.293
Personal estación				1.617	1.118	2.339
Motoristas				1.927	1.327	2.799
Operarios vías				2.230	1.434	3.369
Operarios manten.				1.480	0.874	2.506
Turno (diurno)						
Nocturno	1.354	1.087	1.686	0.817	0.681	0.981
Partido	4.969	3.421	7.216	0.613	0.500	0.753
Rotativo/sin turno	0.790	0.645	0.966	0.256	0.198	0.330
Estudios (universitarios)						
Sin estudios	1.231	0.870	1.740	0.988	0.758	1.287
Primarios	1.455	1.048	2.019	1.071	0.829	1.383
Secundarios	1.488	1.077	2.056	1.052	0.822	1.346
Perc. Salud (MB/B/R)						
Mala/muy mala	1.533	1.075	2.184	1.733	1.357	2.215
Actividad física (No)						
Habitualmente	0.939	0.852	1.035	0.909	0.822	1.005
Tabaco (no fuma)						
Ex-fumador	1.076	0.912	1.269	1.278	1.097	1.488
Fumador	1.047	0.946	1.159	1.593	1.434	1.770
Alcohol (no bebe)						
Semanal/espórádico	0.878	0.776	0.993	0.796	0.702	0.903
Diariamente	0.950	0.841	1.073	0.759	0.665	0.867
Enfermedades (ninguna)						
Una	1.368	1.236	1.513	1.170	1.059	1.293
Más de una	5.882	3.652	9.476	6.390	3.238	12.614
Deviance (df)	44713.62 (10161)			66589.41 (13259)		
Test de Razón de Verosimilitud (p)	614.00 (0.000)			1055.00 (0.000)		

En negrita test de Wald mayor que 1, aunque no significativo

2.4.- EL MÉTODO WLW.

El método WLW (Wei et al., 1989) es una aproximación marginal que no impone ninguna estructura particular para la dependencia entre las distintas observaciones de un mismo individuo. Cada recurrencia del suceso (o tipo distinto de suceso) se modeliza como un estrato diferente. Dentro de cada estrato se utilizan los datos de modo *marginal*, es decir lo que resultaría si el analista a la hora de recoger los datos ignorase todo excepto el tipo de suceso determinado. El resultado es que (todos) los individuos aparecen en *todos* los estratos.

En el artículo original de Wei et al. (1989) cada individuo con $k=1,2,\dots,K$ recurrencias es dividido en sub-muestras (a las que denominan *sub-historias*), la primera desde el principio de su historia (o de su seguimiento) hasta la primera recurrencia, la segunda hasta la segunda recurrencia, etc. Como el resultado final de cada una de estas sub-historias es un único suceso, se puede utilizar el modelo de riesgos proporcionales estándar para estimar, en cada k sub-historia asociada a las distintas recurrencias del suceso de interés, los parámetros asociados a las variables explicativas, digamos β_k . Wei et al. (1989) demuestran que la distribución conjunta de los β_k ($k=1,\dots,K$) es aproximadamente normal con una covarianza que puede ser estimada consistentemente. Los parámetros para cada una de las variables explicativas pueden ser obtenidos combinando linealmente los diferentes β_k .

En la **Tabla 5** hemos mostrado los resultados de las estimaciones de los modelos WLW.

Tabla 5a. Resultados de la estimación. WLW.

Accidentes	Compañía 1			Compañía 2		
	RR	95% C.I.		RR	95% C.I.	
Sexo (hombre)	0.747	0.614	0.910	0.978	0.829	1.154
Edad (<30)						
31-40	0.938	0.806	1.090	1.112	0.892	1.388
41-50	0.961	0.746	1.240	1.065	0.776	1.461
51-60	0.932	0.697	1.250	1.016	0.729	1.416
>60	0.796	0.447	1.420	0.784	0.420	1.461
Antigüedad (<9)						
10-20	1.428	1.132	1.800	1.121	0.850	1.478
21-25	0.804	0.597	1.080	0.987	0.720	1.351
>25	1.154	0.867	1.540	0.910	0.661	1.254
Número de hijos (1 hijo)						
2 hijos	1.132	1.006	1.270	1.033	0.896	1.192
Más de 2	1.073	0.924	1.250	1.096	0.935	1.284
Estado civil (soltero/a)						
Casado/a	0.979	0.806	1.190	1.126	0.910	1.393
Separ/divorciado	0.833	0.656	1.060	1.158	0.736	1.822
Viudo/a	0.926	0.516	1.660	1.347	0.849	2.136
Ocupación (Directivos)						
Mandos	0.897	0.092	8.700	0.411	0.302	0.561
Técnicos	0.597	0.268	1.330	0.633	0.364	1.102
Administrativos	0.885	0.093	8.450	0.510	0.359	0.725
Mecánicos	1.092	0.113	10.510	0.560	0.404	0.777
Electricistas	1.052	0.111	9.940	0.526	0.379	0.731
Conductores	0.894	0.087	9.170	0.395	0.250	0.623
Personal auxiliar	1.051	0.107	10.350	0.601	0.433	0.835
Personal estación				0.499	0.367	0.678
Motoristas				0.463	0.338	0.633
Operarios vías				0.506	0.341	0.753
Operarios manten.				0.577	0.412	0.807
Turno (diurno)						
Nocturno	0.821	0.619	1.090	0.952	0.781	1.159
Partido	1.470	0.248	8.690	0.822	0.608	1.113
Rotativo/sin turno	0.936	0.622	1.410	1.043	0.758	1.435
Estudios (universitarios)						
Sin estudios	0.946	0.519	1.720	1.066	0.651	1.745
Primarios	1.065	0.587	1.930	0.959	0.589	1.561
Secundarios	1.132	0.627	2.040	0.962	0.593	1.559
Perc. Salud (MB/B/R)						
Mala/muy mala	1.130	0.888	1.440	1.203	0.916	1.579
Actividad física (No)						
Habitualmente	0.967	0.866	1.080	1.029	0.925	1.145
Tabaco (no fuma)						
Ex-fumador	0.926	0.793	1.080	0.917	0.782	1.074
Fumador	1.045	0.950	1.150	0.757	0.679	0.843
Alcohol (no bebe)						
Semanal/espórádico	0.955	0.841	1.080	1.115	0.961	1.292
Diariamente	0.976	0.859	1.110	1.160	1.005	1.339
Enfermedades (ninguna)						
Una	0.924	0.842	1.010	0.859	0.775	0.953
Más de una	1.208	0.797	1.830	0.866	0.477	1.571
Deviance (df)	11588.35 (10161)			15757.74 (13259)		
Test de Razon de Verosimilitud (p)	52.60 (0.071)			60.30 (0.000)		

En negrita test de Wald mayor que 1. aunque no significativo

Tabla 5b. Resultados de la estimación. WLW.

Enfermedad	Compañía 1			Compañía 2		
	RR	95% C.I.		RR	95% C.I.	
Sexo (hombre)	1.337	1.101	1.624	1.026	0.928	1.134
Edad (<30)						
31-40	0.845	0.771	0.927	0.888	0.772	1.021
41-50	0.887	0.731	1.075	0.787	0.651	0.951
51-60	0.851	0.690	1.050	0.740	0.607	0.903
>60	0.702	0.543	0.907	0.701	0.537	0.916
Antigüedad (<9)						
10-20	1.145	0.976	1.343	1.575	1.360	1.823
21-25	0.783	0.650	0.943	1.333	1.125	1.579
>25	0.942	0.770	1.153	1.204	1.010	1.434
Número de hijos (1 hijo)						
2 hijos	0.979	0.914	1.049	0.917	0.839	1.003
Más de 2	1.149	1.054	1.253	1.029	0.923	1.148
Estado civil (soltero/a)						
Casado/a	1.036	0.925	1.160	1.066	0.949	1.198
Separ/divorciado	1.084	0.926	1.269	1.120	0.924	1.359
Viudo/a	0.929	0.646	1.335	1.306	0.895	1.907
Ocupación (Directivos)						
Mandos	0.997	0.753	1.319	0.985	0.809	1.200
Técnicos	1.255	1.022	1.542	1.209	1.002	1.460
Administrativos	0.988	0.742	1.317	1.060	0.868	1.294
Mecánicos	0.867	0.659	1.141	0.963	0.771	1.203
Electricistas	0.810	0.585	1.122	1.068	0.856	1.333
Conductores	0.891	0.667	1.190	1.041	0.726	1.492
Personal auxiliar	1.106	0.833	1.470	0.968	0.786	1.192
Personal estación				1.155	0.942	1.416
Motoristas				1.065	0.862	1.316
Operarios vías				1.042	0.811	1.338
Operarios manten.				0.844	0.634	1.122
Turno (diurno)						
Nocturno	1.076	0.928	1.248	0.893	0.796	1.002
Partido	0.833	0.658	1.055	0.951	0.830	1.088
Rotativo/sin turno	1.155	0.990	1.348	1.298	1.119	1.506
Estudios (universitarios)						
Sin estudios	1.031	0.843	1.261	0.949	0.795	1.131
Primarios	1.003	0.830	1.211	0.942	0.796	1.115
Secundarios	0.951	0.789	1.145	0.844	0.720	0.990
Perc. Salud (MB/B/R)						
Mala/muy mala	0.980	0.751	1.278	1.172	0.958	1.432
Actividad física (No)						
Habitualmente	0.933	0.878	0.992	0.940	0.884	0.999
Tabaco (no fuma)						
Ex-fumador	0.976	0.895	1.064	0.990	0.900	1.088
Fumador	0.904	0.851	0.960	0.957	0.892	1.026
Alcohol (no bebe)						
Semanal/espórádico	0.974	0.904	1.050	0.918	0.846	0.995
Diariamente	0.946	0.881	1.015	0.949	0.867	1.038
Enfermedades (ninguna)						
Una	1.033	0.973	1.096	1.035	0.971	1.104
Más de una	1.304	1.060	1.604	1.410	1.098	1.809
Deviance (df)	39871.87 (10161)			59191.81 (13259)		
Test de Razón de Verosimilitud (p)	91.40 (0.000)			162.00 (0.000)		

En negrita test de Wald mayor que 1, aunque no significativo

El método WLW presenta importantes desventajas. Cook y Lawless (1997) señalan que el modelo WLW es válido únicamente bajo la posiblemente muy restrictiva hipótesis de censura independiente (restrictiva como consecuencia de las recurrencias). Esta desventaja, sin embargo, no tiene implicaciones prácticas de importancia, excepto en el caso de que alguna recurrencia fuese terminal (no es nuestro caso). La desventaja más importante en nuestro contexto, es que el método WLW supone que en el instante t todos los individuos (observados en t) constituyen el conjunto de riesgo para la recurrencia correspondiente, independientemente del número de recurrencias previas que tuviese el individuo (Cook y Lawless, 1997; Lipschutz y Snapinn, 1997). Esto no representaría ningún problema si se tratase de diferentes sucesos de un tipo totalmente distinto. Pero si se trata de recurrencias del mismo suceso (como en nuestro caso) podría ocurrir que algún hecho emprendido con el fin de reducir el número de tales recurrencias (control por parte de la empresa, por ejemplo) redujese el número de recurrencias a partir de su implementación (nótese que ocurriría lo mismo en el caso de *harvesting*). Como consecuencia aumentaría la correlación entre los estimadores de los parámetros.

2.5.- EL MÉTODO PWP.

Las limitaciones del método WLW pueden ser superadas, en parte, utilizando la aproximación de Prentice, William y Peterson (1981), PWP. Se trata de un modelo marginal respecto a la estimación de los parámetros, pero condicional en relación a la construcción del conjunto de individuos en riesgo. El modelo PWP permite que el riesgo basal varíe entre diferentes recurrencias. Se trata, por tanto, de un modelo de riesgos proporcionales con estratos tiempo-dependientes, en el que la dependencia entre recurrencias se controla estratificando por el número previo de ocurrencias del suceso de interés. Así, a diferencia del método WLW, en el que todos los individuos están en riesgo en todo momento, el modelo PWP incluye en el conjunto de riesgo para la recurrencia k únicamente a los individuos que han experimentado $k-1$ recurrencias.

Tabla 6a. Resultados de la estimación. PWP.

Accidente	Compañía 1			Compañía 2		
	RR	95% C.I.		RR	95% C.I.	
Sexo (hombre)	0.819	0.592	1.133	0.926	0.731	1.173
Edad (<30)						
31-40	1.159	0.897	1.496	1.020	0.748	1.393
41-50	1.377	0.904	2.097	1.261	0.863	1.844
51-60	1.252	0.794	1.977	1.157	0.779	1.720
>60	1.707	0.848	3.434	0.950	0.463	1.950
Antigüedad (<9)						
10-20	0.552	0.406	0.751	0.642	0.463	0.890
21-25	0.830	0.589	1.169	0.755	0.525	1.085
>25	0.629	0.428	0.927	0.886	0.608	1.290
Numero de hijos (1 hijo)						
2 hijos	1.214	1.009	1.461	0.898	0.754	1.068
Más de 2	1.165	0.923	1.469	0.878	0.720	1.071
Estado civil (soltero/a)						
Casado/a	1.116	0.787	1.581	0.802	0.624	1.030
Separ/divorciado	0.876	0.529	1.453	0.736	0.495	1.094
Viudo/a	3.069	1.327	7.097	0.615	0.374	1.013
Ocupación (Directivos)						
Mandos	0.693	0.176	2.733	0.603	0.349	1.041
Técnicos	1.544	0.646	3.694	0.357	0.157	0.815
Administrativos	0.973	0.235	4.037	0.782	0.452	1.352
Mecánicos	0.598	0.156	2.283	0.714	0.412	1.237
Electricistas	0.604	0.153	2.380	0.767	0.434	1.354
Conductores	0.517	0.125	2.139	0.676	0.329	1.391
Personal auxiliar	0.608	0.149	2.475	0.842	0.473	1.496
Personal estación				0.792	0.458	1.368
Motoristas				0.686	0.391	1.203
Operarios vías				0.812	0.418	1.580
Operarios manten.				0.620	0.327	1.177
Turno (diurno)						
Nocturno	1.116	0.803	1.551	1.051	0.815	1.355
Partido	0.810	0.314	2.089	1.073	0.835	1.379
Rotativo/sin turno	1.054	0.701	1.583	1.290	0.866	1.921
Estudios (universitarios)						
Sin estudios	1.706	0.878	3.315	1.136	0.760	1.697
Primarios	1.882	0.970	3.649	1.206	0.813	1.788
Secundarios	1.670	0.877	3.181	1.293	0.888	1.884
Perc. Salud (MB/B/R)						
Mala/muy mala	0.695	0.492	0.982	0.925	0.692	1.235
Actividad física (No)						
Habitualmente	1.005	0.856	1.180	1.100	0.956	1.265
Tabaco (no fuma)						
Ex-fumador	1.191	0.941	1.507	1.050	0.858	1.285
Fumador	1.048	0.908	1.208	0.940	0.821	1.076
Alcohol (no bebe)						
Semanal/espórádico	0.986	0.816	1.190	0.925	0.772	1.107
Diariamente	0.876	0.725	1.057	0.913	0.761	1.095
Enfermedades (ninguna)						
Una	0.952	0.823	1.100	1.007	0.891	1.138
Más de una	10.929	5.102	23.412	6.188	2.182	17.551
Deviance (df)	8473.776 (10161)			11245.71 (13259)		
Test de Razon de Verosimilitud (p)	88.80 (0.000)			70.20 (0.005)		

En negrita test de Wald mayor que 1. aunque no significativo

Tabla 6b. Resultados de la estimación. PWP.

Enfermedad	Compañía 1			Compañía 2		
	RR	95% C.I.		RR	95% C.I.	
Sexo (hombre)	0.968	0.829	1.130	1.017	0.923	1.119
Edad (<30)						
31-40	0.972	0.866	1.090	0.951	0.807	1.122
41-50	0.932	0.773	1.123	0.975	0.809	1.174
51-60	0.980	0.800	1.201	1.030	0.848	1.253
>60	1.142	0.771	1.691	1.031	0.751	1.417
Antigüedad (<9)						
10-20	0.822	0.707	0.956	0.661	0.561	0.779
21-25	1.187	0.983	1.433	0.691	0.578	0.825
>25	0.809	0.670	0.978	0.706	0.585	0.853
Número de hijos (1 hijo)						
2 hijos	0.965	0.879	1.058	1.041	0.960	1.130
Más de 2	0.976	0.879	1.083	1.050	0.958	1.150
Estado civil (soltero/a)						
Casado/a	0.926	0.811	1.056	0.966	0.857	1.088
Separ/divorciado	0.874	0.713	1.073	0.929	0.771	1.119
Viudo/a	0.922	0.632	1.344	0.964	0.770	1.208
Ocupación (Directivos)						
Mandos	0.508	0.327	0.789	1.001	0.748	1.339
Técnicos	0.751	0.547	1.030	0.895	0.679	1.178
Administrativos	0.746	0.464	1.199	1.075	0.812	1.422
Mecánicos	0.648	0.418	1.006	0.986	0.733	1.327
Electricistas	0.589	0.350	0.993	1.052	0.779	1.422
Conductores	0.460	0.294	0.722	1.119	0.737	1.699
Personal auxiliar	0.412	0.264	0.643	1.115	0.833	1.492
Personal estación				1.187	0.885	1.594
Motoristas				1.126	0.840	1.509
Operarios vías				1.123	0.798	1.579
Operarios manten.				1.191	0.780	1.820
Turno (diurno)						
Nocturno	1.436	1.200	1.718	0.999	0.869	1.149
Partido	0.841	0.589	1.200	1.054	0.908	1.224
Rotativo/sin turno	1.352	1.134	1.611	1.344	1.085	1.664
Estudios (universitarios)						
Sin estudios	1.164	0.905	1.498	0.928	0.766	1.124
Primarios	1.230	0.968	1.562	1.000	0.830	1.205
Secundarios	1.268	1.006	1.599	1.079	0.905	1.287
Perc. Salud (MB/B/R)						
Mala/muy mala	0.908	0.759	1.087	0.898	0.780	1.033
Actividad física (No)						
Habitualmente	0.972	0.898	1.052	1.037	0.971	1.108
Tabaco (no fuma)						
Ex-fumador	1.067	0.943	1.207	1.022	0.928	1.126
Fumador	0.954	0.887	1.026	1.130	1.058	1.207
Alcohol (no bebe)						
Semanal/espórádico	1.032	0.944	1.129	1.010	0.932	1.094
Diariamente	1.046	0.953	1.148	1.014	0.934	1.101
Enfermedades (ninguna)						
Una	0.949	0.882	1.022	1.042	0.980	1.108
Más de una	3.460	2.197	5.449	4.226	2.451	7.289
Deviance (df)	31368.91 (10161)			44546.15 (13259)		
Test de Razón de Verosimilitud (p)	201.00 (0,000)			264.00 (0,000)		

En negrita test de Wald mayor que 1. aunque no significativo

Pero precisamente aquí radica su principal desventaja. En el primer estrato, la distribución de frecuencias de las variables explicativas del suceso de interés entre aquellos individuos que padecieron dicho suceso y aquellos que no lo padecieron, será más o menos aleatoria. En el segundo estrato, sin embargo, la probabilidad de ocurrencia del suceso no será la misma para todos los individuos, sino que presumiblemente será mayor para aquellos que ya padecieron el suceso en el primer estrato. Como consecuencia la distribución de frecuencias de las variables explicativas entre los individuos en el segundo estrato no será aleatoria. Este fenómeno se conoce como *pérdida de aleatorización* (Therneau y Hamilton, 1997). Nótese que esta *pérdida* será mayor conforme vayamos moviéndonos a estratos superiores (y cuanto más estratos se consideren). En estos casos la consistencia de las estimaciones depende, quizás esencialmente, de la inclusión de *todas* las variables relevantes.

Aunque los métodos marginales cumplen (*marginamente*) la hipótesis de riesgos proporcionales, suponen explícita (PWP) o implícitamente (WLW) que la función de riesgo basal varía de una recurrencia a otra (Clayton, 1994). Como consecuencia son extremadamente sensibles al cumplimiento (en realidad) de la hipótesis de riesgos proporcionales (Therneau y Hamilton, 1997). Finalmente, y respecto a las aproximaciones condicionales, los métodos marginales no son totalmente eficientes (Oakes, 1992 y 1997). La ineficiencia aumentará de forma proporcional al número de estratos (recurrencias) e inversamente proporcional al número de observaciones en cada estrato. En analogía a los modelos lineales es razonable suponer que la ineficiencia también aumentará conforme lo hace la correlación entre recurrencias. Es por tanto de esperar que WLW sea menos eficiente que PWP.

Como habíamos comentado, los modelos marginales WLW y PWP son muy ineficientes. Los intervalos de confianza de los riesgos relativos son muy amplios, conteniendo en la mayoría de las ocasiones a la unidad (**Tabla 5 y 6**). Pero además, las estimaciones de tales riesgos no parecen ser consistentes. Prácticamente todas las variables presentan unos riesgos relativos de significado opuesto al de las tasas crudas. Nótese que ocurre incluso con aquellas variables fijas en el tiempo, como sexo, en las que las tasas crudas y las ajustadas deberían coincidir.

Estas inconsistencias pueden ser debidas al incumplimiento de la hipótesis de riesgos proporcionales. Contrastamos dicha hipótesis, tanto global como individualmente, utilizando los residuos de Schoenfeld (véase en este sentido Venables y Ripley (1997)). En todos los casos se rechaza la hipótesis de riesgos proporcionales tanto globalmente (para WLW $p=0,00000$ y para PWP $p=0,00008$) como para la mayoría de las variables explicativas.

En el caso del método PWP se produce, además, el fenómeno de *pérdida de aleatorización* que comentamos más arriba. Obsérvese en la **Tabla 7** como los 'Directivos' (categoría de referencia), por ejemplo, notifican seis episodios de enfermedad (estratos) como máximo, no existiendo prácticamente observaciones a partir del segundo estrato.

Tabla 7. Número de episodios de enfermedad según estratos y ocupación en la Compañía 2.

	Estrato (Número de orden del episodio)									
	1	2	3	4	5	6	7	8	9	10
Directivos	39	14	6	3	2	2				
Mandos	268	200	121	79	47	32	18	9	7	5
Técnicos	59	36	22	14	5	4	1			
Administrativos	93	43	24	12	8	6	2	2	1	
Mecánicos	131	83	48	24	14	8	6	4	1	1
Electricistas	131	76	39	24	15	9	8	4	4	2
Conductores	9	6	6	4	2	1				
Personal auxiliar	114	76	42	27	20	13	9	1	1	1
Personal estación	459	268	169	115	71	43	27	20	15	13
Motoristas	311	189	112	67	40	22	14	11	10	7
Operarios de vías	60	37	28	18	11	9	1			
Operarios de mantenimiento	20	12	6	3	1					

Como señalamos ante este fenómeno la consistencia de las estimaciones depende de la inclusión de *todas* las posibles variables explicativas, bastante difícil en este caso.

3.- APROXIMACIONES CONDICIONALES. EL MODELO AG-AUTORREGRESIVO.

Entre los llamados métodos condicionales utilizamos un modelo autoregresivo o Markoviano. Se trata de introducir en el modelo AG, variables explicativas tiempo-dependientes a fin de capturar la posible estructura de dependencia, en concreto el número de recurrencias previas. Aunque es un modelo muy general, y por tanto bastante apropiado en la mayoría de las ocasiones, supone que la dependencia entre recurrencias decae exponencialmente. Que sepamos no se han explorado hasta el momento otros esquemas de dependencia alternativos.

En la **Tabla 8** presentamos las estimaciones del modelo condicional autoregresivo. Hemos señalado en negrita aquellos riesgos relativos (RR) cuyo valor del test de Wald es mayor que la unidad aunque no es estadísticamente significativo ($p > 0,05$). La posible existencia de una elevada multicolinealidad, nos impide concluir que tales riesgos relativos no sean efectivamente estadísticamente significativos (Saez y Barceló, 1998b).

El riesgo relativo se define como la razón entre las tasas de incidencia de la categoría correspondiente (grupo expuesto) y la categoría de referencia (grupo no expuesto) (véase Kahn y Sempos, 1989). En nuestro caso, por ejemplo, que las mujeres presenten un riesgo relativo estimado de 1,309 en accidentes en la Compañía 1 (**Tabla 8a**) se interpretaría que éstas tienen más riesgo o probabilidad (en particular un 30,9% más) de padecer un episodio de IT por accidente que los hombres (categoría de referencia).

Aunque los riesgos relativos son muy similares, la aproximación AG condicional parece ser 'mejor' que la simple estimación robusta de la varianza. En este sentido, y utilizando un contraste de razón de verosimilitud, rechazamos las hipótesis que permitirían pasar del modelo general (AG autoregresivo) al modelo restringido (estimación robusta de la varianza). Interpretaremos por tanto los resultados de las estimaciones de los modelos condicionales (**Tabla 8**).

Tabla 8a. Resultados de la estimación. AG-Autoregresivo.

Accidente	Compañía 1			Compañía 2		
	RR	95% C.I.		RR	95% C.I.	
Sexo (hombre)	1.309	0.861	1.989	1.169	0.938	1.458
Edad (<30)						
31-40	1.313	1.037	1.662	1.275	0.931	1.746
41-50	1.677	1.113	2.528	1.452	0.993	2.123
51-60	1.587	1.012	2.490	1.416	0.948	2.115
>60	1.403	0.636	3.092	0.971	0.445	2.120
Antigüedad (<9)						
10-20	0.588	0.426	0.812	0.513	0.372	0.706
21-25	0.854	0.582	1.255	0.566	0.399	0.805
>25	0.585	0.389	0.880	0.689	0.477	0.996
Número de hijos (1 hijo)						
2 hijos	1.143	0.945	1.383	0.971	0.812	1.162
Más de 2	1.163	0.915	1.478	0.899	0.732	1.104
Estado civil (soltero/a)						
Casado/a	1.227	0.906	1.661	1.080	0.840	1.387
Separ/divorciado	0.986	0.599	1.626	1.089	0.697	1.703
Viudo/a	0.683	0.163	2.870	0.924	0.504	1.695
Ocupación (Directivos)						
Mandos	41.517	5.118	336.788	6.287	1.925	20.535
Técnicos	9.483	1.308	68.736	3.127	0.877	11.155
Administrativos	28.489	3.284	247.159	3.855	1.134	13.097
Mecánicos	62.269	7.752	500.153	7.814	2.377	25.683
Electricistas	50.487	6.117	416.735	6.913	2.092	22.839
Conductores	61.144	7.561	494.431	7.119	1.900	26.667
Personal auxiliar	23.364	2.800	194.996	6.983	2.114	23.061
Personal estación				4.532	1.384	14.842
Motoristas				4.839	1.470	15.927
Operarios vías				7.316	2.142	24.995
Operarios manten.				8.370	2.408	29.091
Turno (diurno)						
Nocturno	0.821	0.610	1.103	0.828	0.657	1.045
Partido	2.804	1.186	6.632	0.496	0.349	0.705
Rotativo/sin turno	0.562	0.421	0.750	0.294	0.207	0.417
Estudios (universitarios)						
Sin estudios	1.164	0.533	2.539	1.650	1.011	2.694
Primarios	1.302	0.608	2.790	1.679	1.038	2.715
Secundarios	1.408	0.662	2.995	1.545	0.966	2.471
Perc. Salud (MB/B/R)						
Mala/muy mala	1.220	0.753	1.976	0.978	0.709	1.349
Actividad física (No)						
Habitualmente	1.051	0.913	1.211	1.039	0.910	1.186
Tabaco (no fuma)						
Ex-fumador	0.994	0.768	1.285	1.207	0.980	1.486
Fumador	1.072	0.926	1.242	1.224	1.063	1.409
Alcohol (no bebe)						
Semanal/espórádico	0.901	0.758	1.069	1.049	0.876	1.256
Diariamente	0.963	0.807	1.149	0.930	0.772	1.121
Enfermedades (ninguna)						
Una	1.101	0.947	1.280	1.011	0.889	1.150
Más de una	8.501	4.614	15.665	7.704	3.383	17.545
Deviance (df)	13036.64 (10157)			16447.09 (13254)		
Test de Razon de Verosimilitud (p)	1050.00 (0.000)			1515.000 (0.000)		

En negrita test de Wald mayor que 1. aunque no significativo

Tabla 8b. Resultados de la estimación. AG- Autoregresivo.

Enfermedad	Compañía 1			Compañía 2		
	RR	95% C.I.		RR	95% C.I.	
Sexo (hombre)	1.050	0.867	1.273	1.153	1.044	1.273
Edad (<30)						
31-40	1.156	1.023	1.306	0.971	0.819	1.150
41-50	1.054	0.852	1.304	0.933	0.766	1.137
51-60	1.093	0.864	1.382	0.971	0.788	1.197
>60	1.352	0.919	1.990	0.964	0.686	1.354
Antigüedad (<9)						
10-20	0.845	0.711	1.005	0.774	0.652	0.918
21-25	1.025	0.835	1.258	0.759	0.630	0.915
>25	0.739	0.593	0.921	0.864	0.711	1.050
Número de hijos (1 hijo)						
2 hijos	0.980	0.883	1.088	0.994	0.908	1.088
Más de 2	1.056	0.932	1.197	0.995	0.899	1.101
Estado civil (soltero/a)						
Casado/a	1.049	0.908	1.211	0.975	0.861	1.105
Separ/divorciado	1.005	0.787	1.282	1.061	0.866	1.300
Viudo/a	0.942	0.560	1.586	0.816	0.598	1.115
Ocupación (Directivos)						
Mandos	2.476	1.533	3.999	1.372	1.028	1.832
Técnicos	1.458	1.003	2.118	1.272	0.943	1.715
Administrativos	2.875	1.726	4.787	1.337	0.994	1.797
Mecánicos	3.024	1.899	4.815	1.156	0.853	1.568
Electricistas	2.528	1.466	4.360	1.268	0.934	1.722
Conductores	3.511	2.195	5.616	1.232	0.767	1.977
Personal auxiliar	2.417	1.492	3.917	1.474	1.094	1.985
Personal estación				1.194	0.893	1.595
Motoristas				1.322	0.985	1.776
Operarios vías				1.387	0.983	1.958
Operarios manten.				1.328	0.873	2.020
Turno (diurno)						
Nocturno	1.262	1.052	1.514	0.898	0.789	1.023
Partido	3.363	2.471	4.578	0.795	0.686	0.920
Rotativo/sin turno	0.869	0.734	1.029	0.409	0.342	0.490
Estudios (universitarios)						
Sin estudios	1.228	0.897	1.681	1.016	0.849	1.217
Primarios	1.409	1.045	1.901	1.094	0.921	1.298
Secundarios	1.427	1.066	1.912	1.122	0.952	1.322
Perc. Salud (MB/B/R)						
Mala/muy mala	1.061	0.816	1.380	1.010	0.843	1.209
Actividad física (No)						
Habitualmente	0.965	0.892	1.045	0.982	0.916	1.063
Tabaco (no fuma)						
Ex-fumador	1.072	0.935	1.229	1.124	1.007	1.255
Fumador	0.985	0.906	1.070	1.202	1.118	1.293
Alcohol (no bebe)						
Semanal/ esporádico	0.923	0.839	1.016	0.939	0.862	1.024
Diariamente	0.988	0.894	1.092	0.931	0.851	1.020
Enfermedades (ninguna)						
Una	1.177	1.081	1.282	1.064	0.995	1.138
Más de una	3.864	2.831	5.275	5.766	3.985	8.344
Deviance (df)	43014.43 (10153)			63192.90 (13248)		
Test de Razón de Verosimilitud (p)	2314.00 (0.000)			4420.00 (0.000)		

En negrita test de Wald mayor que 1. aunque no significativo

4.- RESULTADOS.

Los riesgos relativos, en especial los correspondientes a la variable ocupación, son mucho mayores en la Compañía 1 que en la Compañía 2, aunque en el caso de los episodios de enfermedad las diferencias se reducen mucho. Estos resultados no se corresponden con los descriptivos previos presentados en la **Tabla 1** (tasa *cruda* total de incidencia igual a 129,6 en la Compañía 1 vs 215,2 en la Compañía 2). Sin embargo, los resultados coinciden cuando se realiza el análisis estratificado. La tasa *cruda* total de ausencia por accidentes fue de 97,0 en la Compañía 1 y de 168,5 en la Compañía 2, mientras que la misma tasa por enfermedad fue de 30,7 y de 45,2, respectivamente. Nótese que todos los riesgos relativos, en particular nuevamente los que se refieren a ocupación, son mayores entre los accidentes que entre los episodios de enfermedad.

Por lo que se refiere a accidentes, y con respecto a la principal variable de interés, las categorías de ocupación con mayor riesgo relativo (tomando como categoría de referencia 'Directivos') son las de 'Mecánicos' (62,269, intervalo de confianza al 95% 7,752-500,153) y 'Conductores' (61,144, intervalo de confianza al 95% 7,561-494,431) en la Compañía 1 y 'Operarios de Mantenimiento' en la Compañía 2 (8,370, intervalo de confianza 2,408-29,091). Muy próximos se encuentran los riesgos relativos asociados a 'Electricistas' (50,487, intervalo 6,117-416,735) (Compañía 1) y 'Mecánicos' (7,814, intervalo 2,377-25,683), 'Operarios de vías' (7,316, intervalo 2,142-24,995) y 'Conductores' (7,119, intervalo 1,900-26,667) (Compañía 2). Los menores riesgos relativos los encontramos en la categoría de 'Técnicos' en ambas compañías (9,483, intervalo 1,308-68,736 en la Compañía 1), aunque en la Compañía 2 el riesgo no es estadísticamente significativo al 95% (3,127, intervalo 0,877-11,155).

Por lo que se refiere a los episodios de enfermedad, aunque menores, los riesgos relativos de la variable ocupación se comportan de forma similar. Es de destacar, no obstante, la falta de significación estadística al 95 % de la mayoría de los riesgos en la Compañía 2. La posible presencia de una elevada multicolinealidad únicamente nos permite asegurar la no significación estadística de los riesgos relativos en el caso de 'Mecánicos' (1,156, intervalo 0,853-1,568), 'Electricistas' (1,268, intervalo 0,934-1,722) y 'Conductores' (1,232, intervalo 0,767-1,977).

Por lo que respecta al resto de variables explicativas, las mujeres, aquellos trabajadores entre cuarenta y sesenta años, con una antigüedad en la compañía menor a nueve años, los trabajadores casados con más de un hijo, aquéllos que trabajan en turno partido en la Compañía 1 y en turno diurno en la Compañía 2, aquéllos con menor nivel de estudios (únicamente en la Compañía 2) y los que padecen más de una enfermedad crónica, son los que presentan mayor probabilidad de padecer un episodio de accidente. Por lo que se refiere a los episodios de enfermedad, no existe diferencia significativa según el sexo, aquellos trabajadores mayores de sesenta años (Compañía 1), con una antigüedad en la compañía menor a nueve años, los que trabajan en turno partido en la Compañía 1 y en turno diurno en la Compañía 2, aquéllos con menor nivel de estudios (Compañía 1), ex-fumadores o fumadores y aquéllos con alguna enfermedad son los que presentan mayor probabilidad de padecer un episodio de IT.

En la **Figura 1** representamos los riesgos relativos de padecer un episodio de IT en función del número de episodios previos de IT (entre aquellos trabajadores con un mínimo de un episodio). Nótese que los perfiles de los riesgos son muy similares en ambas compañías, aunque el máximo riesgo se produce algo más tarde en la Compañía 2. Como era de esperar, el riesgo de que un trabajador padezca un episodio de IT aumenta conforme lo hace el número de episodios previos. Existe un momento, sin embargo, en el que el riesgo disminuye (a partir del tercer episodio en la Compañía 1 y del cuarto en la Compañía 2 por lo que se refiere a accidentes y del quinto episodio y del séptimo respectivamente, por lo que se refiere a los episodios de enfermedad). El perfil del riesgo de accidentes es más cóncavo que el de los episodios de enfermedad, sugiriendo mayor persistencia y, por tanto, una mayor dependencia entre las recurrencias.

Figura 1a. Riesgo relativo de padecer un episodio de accidente en función del número de episodios de accidente previos.

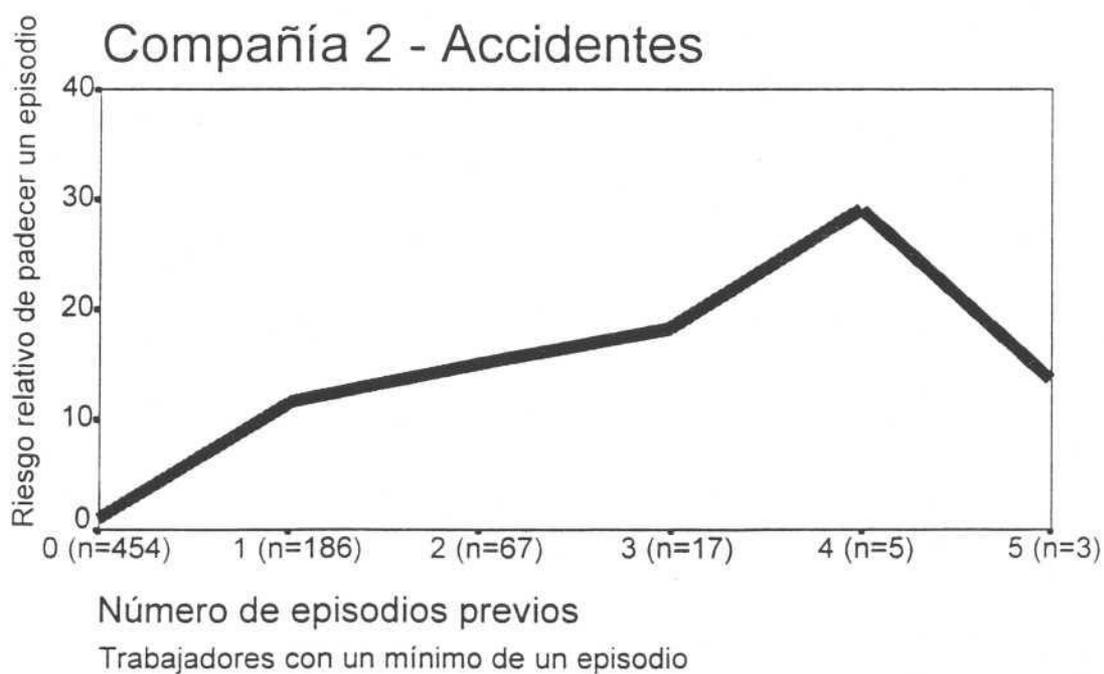
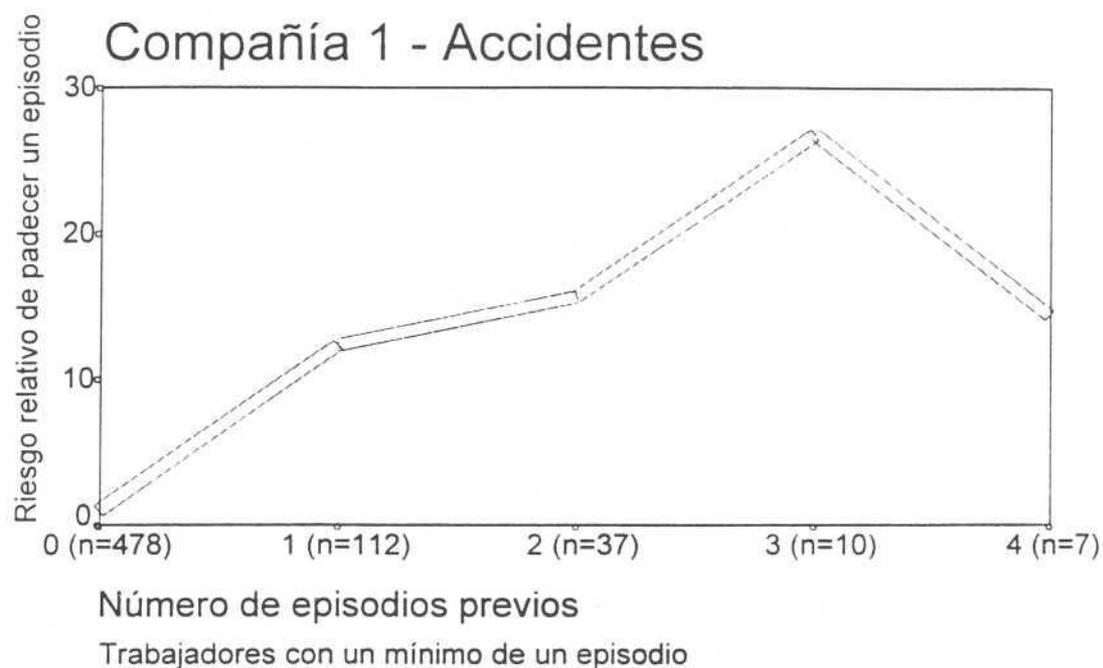
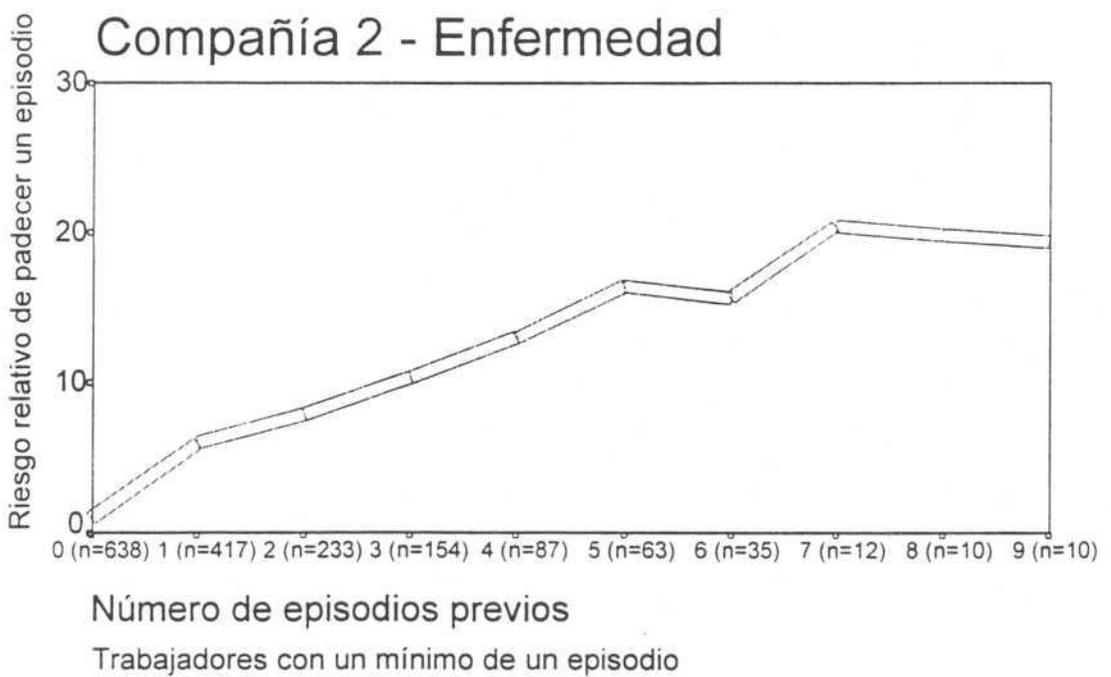
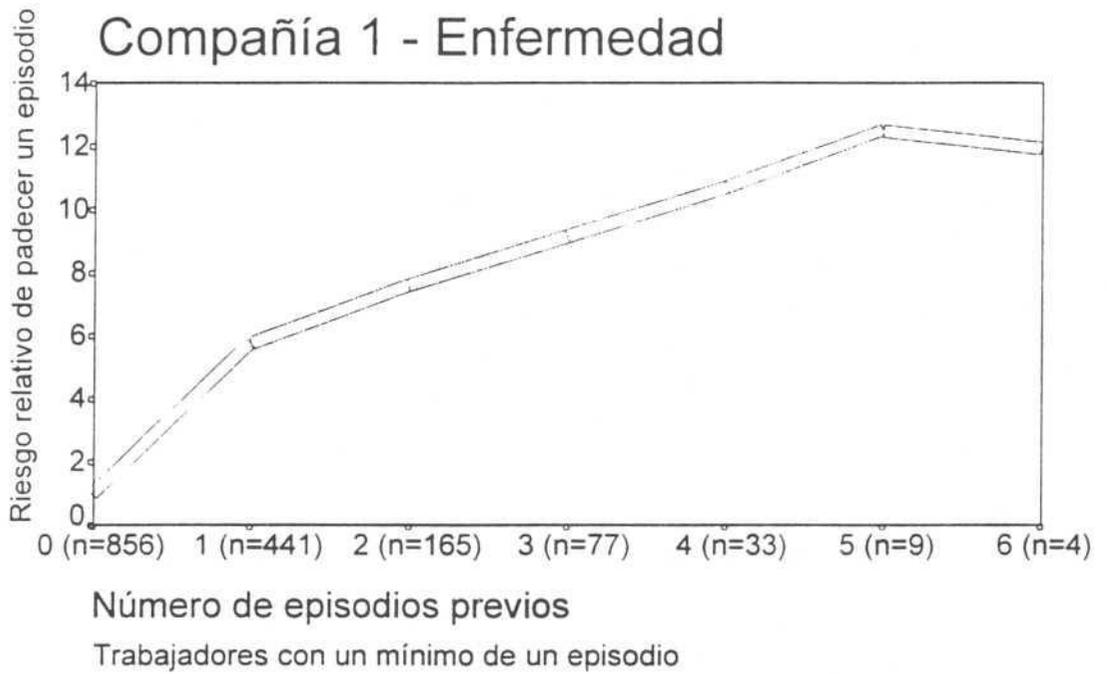


Figura 1b. Riesgo relativo de padecer un episodio de enfermedad en función del número de episodios de enfermedad previos.



5.- CONCLUSIONES.

La ocupación ha resultado ser la variable con mayor capacidad explicativa de la incidencia de las ausencias del trabajo, en especial por lo que se refiere a los episodios de accidente. Es de destacar, sin embargo, que la ocupación también está relacionada con los episodios de enfermedad. Creemos que la principal explicación de esta relación se encuentra en la estructura de seguridad social existente en nuestro país. Las ausencias por enfermedad son pagadas por la Seguridad Social, mientras que las ausencias por accidente las pagan las mutuas y compañías de seguros de enfermedad y accidente. Es bastante posible que algunos episodios de enfermedad se correspondan en realidad con accidentes laborales y/o enfermedades profesionales. Precisamente, las diferencias en los riesgos entre ambas compañías, siempre mayores en la Compañía 1 y una relación entre ocupación y episodios de enfermedad prácticamente inexistente en la Compañía 2, podrían ser la consecuencia de un diferente esquema de cobertura de las contingencias por incapacidad en la Compañía 1 (alcanzado por convenios colectivos específicos).

Nótese que, también en nuestro caso, estos razonamientos nos impiden separar desde un punto de vista práctico las ausencias 'voluntarias' de aquellas 'involuntarias'. De hecho todos nuestros modelos son muy sobre-dispersos (deviance mayor que los grados de libertad), o varianzas estimadas mayores que las nominales. La interpretación de los riesgos estimados, por tanto, no podrá ser completamente fundamentada atendiendo a ninguna de las tres teorías económicas explicativas de las ausencias del trabajo (véase Saez y Barceló, 1998a).

Se detectan importantes diferencias entre los riesgos relativos correspondientes a las distintas categorías de ocupación. Los mecánicos, los operarios de vías, los operarios de mantenimiento y los conductores presentan los mayores riesgos, incluso tras controlar las características personales y ocupacionales de los trabajadores, sus conductas respecto a la salud y sus antecedentes sanitarios. El menor riesgo estimado en aquellas categorías ocupacionales con mayores salarios y estatus socio-laboral ('Directivos' y 'Técnicos') podría ser explicado atendiendo al signo negativo de la relación entre salarios y ausencias del trabajo. Este signo podría ser consecuencia del predominio del efecto sustitución, atendiendo el enfoque trabajo-ocio, o de mayores pérdidas esperadas en la utilidad de los trabajadores en estas categorías, derivadas de un hipotético despido, según argumentan los modelos de 'disciplina laboral'. En este sentido, nuestros hallazgos podrían compararse a los de Kenyon y Dawkins (1989), Drago y Wooden (1992), Barmby et al (1995) y Johansson y Palme (1996). Podría ocurrir, además, que la mayor renta no salarial (enfoque trabajo-ocio), el mayor componente de inversión en capital humano (hipótesis de 'disciplina laboral'), una mayor duración esperada del desempleo (hipótesis de 'disciplina laboral') o un mayor compromiso respecto a la empresa ('normas de grupo de trabajo') en éstas respecto a otras categorías ocupacionales, explicase también su menor riesgo relativo.

Estas argumentaciones, sin embargo, no justificarían el elevado riesgo estimado para la categoría 'Mandos', presumiblemente con elevado salario y similar estatus socio-laboral, ni las diferencias de riesgo entre aquellas categorías con salarios más bajos y estatus inferior. Es posible que los trabajadores en categorías laborales más bajas sean más propensos a los accidentes laborales (Leigh, 1986), tengan menos flexibilidad de ocupar su tiempo en aquellas actividades no laborales que permitan no hacer necesario

ausentarse del trabajo (Chadwick-Jones et al., 1982), o pertenezcan a categorías más peligrosas, peor pagadas y con condiciones de trabajo más adversas (Kenyon y Dawkins, 1989). En este sentido nuestros resultados estarían más cercanos a los de Allen (1996) desde el punto de vista económico o a los trabajos de salud laboral de Marmot et al. (1995) y North et al. (1996) o, en nuestro país, Buitrago et al. (1993). Todos estos autores encuentran una relación inversa entre el nivel de la categoría laboral y las ausencias del trabajo.

Creemos que nuestros hallazgos respecto a ocupación podrían ser explicados recurriendo, principalmente, a la 'adversidad' de las condiciones de trabajo (como argumentan de Marmot et al. (1995) y North et al. (1996); y encuentran empíricamente Johansson y Palme (1996)). Según argumenta la hipótesis de 'disciplina laboral' unas condiciones de trabajo adversas podrían disminuir las pérdidas esperadas en la utilidad de los trabajadores, derivadas de un hipotético despido, lo que incrementaría la probabilidad de las ausencias. La 'adversidad' puede ser tanto desde un punto de vista físico como desde una perspectiva psicológica, más en la línea de la hipótesis de las 'normas de grupos de trabajo' (y, curiosamente, también en la de Marmot et al. (1995) y North et al. (1996)).

En este sentido nótese que las ocupaciones con mayor riesgo de padecer un episodio de accidente ('Mecánicos', 'Electricistas' y 'Conductores') presentan un riesgo de padecer un episodio de enfermedad prácticamente igual a los 'Directivos' (categoría de referencia) en especial en la Compañía 2, aquella menos distorsionada por el esquema de cobertura de las contingencias. Por otra parte, nótese que los turnos más 'incómodos' (partido en la Compañía 1 y diurno en la Compañía 2) son los que presentan mayores riesgos estimados. Nuestros hallazgos están en la línea de la investigación empírica, aunque debamos remontarnos a Vernon et al. (1928 y 1931), quienes al analizar una cohorte de mineros del carbón en Gales, encuentran que las ausencias (salvo las relacionadas con accidentes graves) se relacionan principalmente con la incomodidad del puesto de trabajo (profundidad de la mina, temperatura interior y velocidad del aire).

Como la mayoría de trabajos empíricos (Barmby et al (1991); Drago y Wooden (1992); Barmby et al (1995); Marmot et al. (1995); North et al. (1996); Rizzo et al., (1996) y Allen (1996)) encontramos que, incluso sin incluir las bajas maternales, las mujeres tienen mayor riesgo de ausentarse del trabajo, tanto por accidentes como por episodios de enfermedad (Buitrago et al. (1993), sin embargo, no encuentran ninguna relación entre sexo y episodios de IT). Es muy posible que las dependencias familiares expliquen este mayor absentismo entre las mujeres (Johansson y Palme, 1996).

Precisamente, las dependencias familiares (como en Allen, 1996) podrían explicar el mayor riesgo en trabajadores con más de un hijo, sobretodo por lo que se refiere a los episodios de accidente (véase en este mismo sentido a Drago y Wooden, 1992). Sin embargo, presiones financieras, es decir necesidad de mayores ingresos, explicarían porqué el riesgo es menor entre trabajadores casados. Similares resultados fueron encontrados por Barmby et al (1991); Allen (1996) y Johansson y Palme (1996).

La hipótesis de 'disciplina laboral' podría explicar porque mayor antigüedad se asocia con una menor ausencia (al igual que Fitzgibbons y Moch (1980); Watson (1981), Keller (1983) y Youngblood (1984)). Los trabajadores con mayor antigüedad estarían más comprometidos con la empresa. No descartamos, sin embargo, cierto sesgo de

selección, aquellos trabajadores con mayores ausencias serían los que menos tiempo trabajarían en la empresa. El diseño longitudinal que hemos utilizado, sin embargo, puede haber minimizado este posible sesgo.

Parece existir una relación no lineal entre ausencias y edad, aunque los trabajadores mayores tienen un riesgo mayor (al igual que en Drago y Wooden, (1992) o, más interesante, en nuestro país Buitrago et al. (1993)). Incluso controlando por variables de salud, el deterioro físico de estos trabajadores podría implicar mayores ausencias. La educación parece estar inversamente relacionada con las ausencias (como en van Ours y Ridder (1991) y en Rizzo et al. (1996)) especialmente cuando se ha controlado la educación. La educación puede estar asociada con buenos hábitos laborales, condiciones de trabajo agradables o buena salud (Grossman, 1975), implicando en todos los casos menos ausencias.

La conducta adversa respecto a la salud (fumadores) y condiciones de salud adversas (presencia de enfermedades crónicas) implican lógicamente mayor riesgo de ausencia, particularmente entre los episodios de enfermedad. Resultados similares fueron encontrados por Bartel y Taubman (1986); Drago y Wooden (1992); Marmot et al. (1995); North et al. (1996); Rizzo et al. (1996) y Johansson y Palme (1996).

Finalmente, nos atreveríamos a intentar explicar algunos resultados algo paradójicos. Nótese que el riesgo de padecer un episodio de IT por cualquier contingencia, es mayor entre los que declaran ser no bebedores. Es posible que deba utilizarse un indicador de la conducta respecto al alcohol más informativo, al menos en nuestra cultura, como por ejemplo alguna medida cuantitativa de consumo de alcohol. Nótese que el riesgo relativo asociado a la auto-percepción de la salud, aunque mayor que la unidad (excepto en accidentes en la Compañía 2), no resultó estadísticamente significativo, ni siquiera por lo que se refiere a los episodios de enfermedad. Podría dudarse de la fiabilidad de esta pregunta. Por último, los riesgos relativos asociados a padecer más de una enfermedad son mayores entre los accidentes que entre las enfermedades. Es posible que esta paradoja quede explicada por la imposibilidad práctica de separar las ausencias 'involuntarias' de aquellas 'voluntarias'.

6.- REFERENCIAS

Allen,W.D. (1996): "Family illness and temporary work absence", *Applied Economics*, 28:1177-1180.

Andersen,P.K. (1992): "Repeated assessment of risk factors in survival analysis", *Statistical Methods in Medical Research*, 1:297-315.

Andersen,P.K. y Gill,R.D. (1982): "Cox's regression model for counting processes: A large sample studies". *Annals of Statistics*, 10:1100-1120.

Andersen,P.K; Borgan,O; Gill,R.D. y Keiding,N. (1992): *Statistical Models Based on Counting Processes*, New York: Springer-Verlag.

Barmby,T.A; Orme,C.D. y Treble,J.G. (1991): "Worker absenteeism: an analysis using microdata", *The Economic Journal*, 101:214-229.

Barmby,T.A; Orme,C.D. y Treble,J.G. (1995): "Worker absence histories: a panel data study", *Labour Economics*, 2:53-65.

Bartel,A. y Taubman,P. (1986): "Some economic and demographic consequences of mental illness", *Journal of Labour Economics*, 4(2):243-256.

Breslow,N.E. (1996): "Generalized linear models: checking assumptions and strengthening conclusions", *Statistica Applicata*, 8(1):23-41.

Buitrago,F; Lozano,L; Fernández-Lozano,C; Bonino,F. et al (1992): "Incapacidades Laborales Transitorias de los colectivos profesionales administrativos y de enfermería", *Gaceta Sanitaria*, 7:190-195.

Cain,K.C. y Lange,N.T. (1984): "Approximate case influence for the proportional hazards regression model with censored data", *Biometrics*, 40:493-499.

Chadwick-Jones,J.K; Nicholson,N. y Brown,C. (1982): *Social Psychology of Absenteeism*, New York: Praeger.

Clayton,D. (1994): "Some approaches to the analysis of recurrent event data", *Statistical Methods in Medical Research*, 3:244-262.

Cook,R.J. y Lawless,J.F. (1997):"Discussion of paper by Wei and Glidden", *Statistics in Medicine*, 16:841-843.

Cox,D.R. (1972): "Regression models and life tables (with discussion)", *Journal of the Royal Statistical Society, B*, 74:187-220.

Cox,D.R. (1975): "Partial likelihood", *Biometrika*, 62:269-276.

Cox,D.R. y Oakes, D. (1984): *Analysis of Survival Data*, London: Chapman and Hall.

- Drago,R. y Wooden,M. (1992):** "The determinants of labor absence: economic factors and workgroup norms across countries". *Industrial and Labor Relations Review*, 45(4):764-778.
- Fitzgibbons,D. y Moch,M. (1980):** "Employee absenteeism: a multivariate analysis with replication". *Organizational Behavior and Human Performance*, 26(4):349-372.
- Greene,W.H. (1993):** *Econometric Methods*. Englewood Cliffs. New Jersey: Prentice-Hall.
- Grossman,M. (1975):** "The correlations between health and schooling" en **Terlecky,N.** (ed). *Household Production and Consumption*. New York: Columbia University Press.
- Huber,P.J. (1967):** "The behaviour of maximum likelihood estimates under non-standard conditions". *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 1:221-233.
- Johansson,P. y Palme,M. (1996):** "Do economic incentives affect work absence? Empirical evidence using Swedish micro data". *Journal of Public Economics*, 59:195-219.
- Kahn,H.A. y Sempos,C.T. (1989):** *Statistical Methods in Epidemiology*, Oxford University Press.
- Kalbfleisch,J.D. y Prentice,R.L. (1980):** *The Statistical Analysis of Failure Time Data*, New York: Wiley.
- Keller,R.T. (1983):** "Predicting absenteeism from prior absenteeism, attitudinal factors, and nonattitudinal factors". *Journal of Applied Psychology*, 68(3):536-540.
- Kenyon,P. y Dawkins,P. (1989):** "A time series analysis of labour absence in Australia". *The Review of Economics and Statistics*, 71:232-239.
- Lee,E.W; Wei,L.J. y Amato,D. (1992):** "Cox-type regression analysis for large number of small groups of correlated failure time observations" en **Klein,J.P. y Goel,P.K.** (eds). *Survival Analysis. State of the Art*, pp. 237-247. Dordrech: Kluwer Academic Publishers.
- Leigh,J.P. (1986):** "Correlates of absence from work due to illness", *Human Relations*, 39(1):81-100.
- Liang,K.Y. y Zeger,S.L. (1986):** "Longitudinal data analysis using Generalized Linear Models". *Biometrika*, 73:13-22.
- Liang,K.Y; Zeger,S.L. y Qaqish,B. (1992):** "Multivariate regression analysis for categorical data (with Discussion)", *Journal of the Royal Statistical Society, Series B*, 54:3-40.
- Lin,D.Y. y Wei,L.J. (1989):** "The robust inference for the Cox proportional hazards model", *Journal of the American Statistical Association*, 84:1074-1079.

Lipschutz,K.H. y Snappin,S.M. (1997): "Discussion of paper by Wei and Glidden", *Statistics in Medicine*, 16:846-848.

Marmot,M; Feeney,A; Shipley,M; North,F. y Syme, S.L. (1995): "Sickness absence as a measure of health status and functioning: from the UK Whitehall II Study", *Journal of Epidemiology and Community Health*, 49:124-130.

McCall,B.P. (1994): "Testing the proportional hazards assumption in the presence of unmeasured heterogeneity", *Journal of Applied Econometrics*, 9:321-334.

North,F.M; Syme,S.L; Feeney,A; Head,J; Shipley,M.J. y Marmot,M. (1993): "Explaining socioeconomic differences in sickness absence: The Whitehall II Study", *British Medical Journal*, 306:361-365.

North,F.M; Syme,S.L; Feeney,A; Shipley,M. y Marmot,M. (1996): "Psychosocial work environment and sickness absence among British civil servants. The Whitehall II Study", *American Journal of Public Health*, 86(3):332-340.

Oakes,D. (1992): "Frailty models for multiple events", en **Klein,J.P. y Goel,P.K. (eds).** *Survival Analysis. State of the Art*, pp. 371-379. Dordrech: Kluwer Academic Publishers.

Oakes,D. (1997):"Discussion of paper by Wei and Glidden", *Statistics in Medicine*, 16:843.

Prentice,R.L.;Williams,B.J. y Peterson,A.V. (1981):"On the regression analysis of multivariate failure time data", *Biometrika*, 68:373-379.

Reid,N. y Crépeau,H. (1985): "Influence functions for proportional hazards regression", *Biometrika*, 72:1-9.

Rizzo,J.A; Abbot,T.A. y Pashko,S. (1996): "Labour productivity effects of prescribed medicines for chronically ill workers", *Health Economics*, 5:249-265.

Saez,M. y Barceló,M.A. (1998a): "Una revisión del análisis económico de la incidencia de las ausencias del trabajo", Barcelona: CRES, Universitat Pompeu Fabra.

Saez,M. y Barceló,M.A. (1998b): "Un criterio para omitir variables superfluas en modelos de regresión", *Gaceta Sanitaria* (en prensa).

Schoenfeld,D.A. (1982): "Partial residuals for the proportional hazards regression model", *Biometrika*, 69:239-241.

Taylor,P.J. y Pocock,S.J. (1982): "Sickness absence - its measurement and control" en **Schilling,R.S. (ed).** *Occupational Health Practice*, pág. 339-358. Londres: Butterworths & Co.

Therneau,T.M; Grambsch,P.M. y Fleming,T.R. (1990): "Martingale based residuals for survival models", *Biometrika*, 77:147-160.

- Therneau,T.M. y Hamilton,S.A. (1997):** "rhDNase as an example of recurrent event analysis". *Statistics in Medicine*, 16(18):2029-2047.
- van Ours,J. y Ridder,G. (1991):** "Cyclical variation in vacancy durations and vacancy flows. An empirical analysis". *European Economic Review*, 35:1143-1155.
- Venables,W.N. y Ripley,B.D. (1997):** *Modern Applied Statistics with S-Plus*, New York: Springer-Verlag.
- Vernon,H.M. y Bedford,T. (1928):** "A study of absenteeism in a group of ten collieries", London: Medical Research Council, Industrial Fatigue Research Board, Report No. 51.
- Vernon,H.M. y Warner,C.G. (1931):** "Two studies of absenteeism in coal mines", London: Medical Research Council, Industrial Health Research Board, Report No. 62.
- Watson,C.J. (1981):** "An evaluation of some aspects of the Steers and Rhodes model of employee attendance", *Journal of Applied Psychology*, 66(3):385-389.
- Wei,L.J; Lin,D.Y. y Weissfeld,L. (1989):** "Regression analysis of multivariate incomplete failure time data by modelling marginal distributions", *Journal of the American Statistical Association*, 84(408):1065-1073.
- White,H. (1980):** "A heteroskedasticity-consistent covariance matrix estimate and a direct test for heteroskedasticity", *Econometrica*, 48:817-830.
- White,H. (1982):** "Maximum likelihood estimation of misspecified models", *Econometrica*, 50:1-25.
- Youngblood,S.A. (1984):** "Work, nonwork and withdrawal", *Journal of Applied Psychology*, 69(1):106-117.
- Zeger,S.L. y Liang,K.Y. (1992):** "An overview of methods for the analysis of longitudinal data", *Statistics in Medicine*, 11:1825-1839.

7.- ANEXO METODOLÓGICO.

Tasa de IT por 100.000 trabajadores-día.

$$Tasa_i = \frac{\sum \text{episodios en la categoría } i}{\text{trabajadores en la categoría } i \times 1096 - \sum \text{días de baja en la categoría } i}$$

Estimación máximo-verosímil del modelo de regresión de Poisson.

El logaritmo de la función de verosimilitud puede escribirse:

$$\ln L = \sum_i -\mu_i + y_i \beta' x_i - \ln y_i!$$

siendo x_i el vector de p variables explicativas para cada individuo (suele contener una constante).

Las ecuaciones de verosimilitud se expresan (vector de *scores*):

$$U(\beta) = \frac{\partial \ln L}{\partial \beta} = \sum_i y_i - \mu_i x_i = 0$$

El Hessiano se puede expresar:

$$H(\beta) = \frac{\partial^2 \ln L}{\partial \beta \partial \beta'} = -\sum_i \mu_i x_i x_i' = 0$$

Nótese que el Hessiano está definido negativo para toda x y β .

La maximización del logaritmo de la función de verosimilitud se puede realizar utilizando el algoritmo de Newton-Raphson. Cuando el algoritmo converja, se puede obtener un estimador de la matriz de covarianzas asintótica de los estimadores de los parámetros, V ,

$$V = \left[\sum_i \mu_i x_i x_i' \right]^{-1}$$

Contraste de la sobre-dispersión en un modelo de Poisson.

Planteamos la hipótesis nula que la varianza condicional de la variable dependiente en un modelo de regresión de Poisson, $\text{Var}(y_i | x_{ij})$ viene determinada por completo por su esperanza condicional, $E(y_i | x_{ij})$, es decir: $H_0: \text{Var}(y_i | x_{ij}) = E(y_i | x_{ij}) = \mu_i$. La hipótesis alternativa es que la varianza está relacionada con los regresores y , por tanto, no está determinada totalmente por la esperanza condicional.

Siguiendo a **Breslow (1996)**:

$H_0: \phi=1$, en $\text{Var}(y_i | x_{ij}) = \phi \mu_i$

$$T_{\phi}^2 = \frac{1}{2n} \left\{ \sum_{i=1}^n \frac{y_i - \hat{\mu}_i}{\hat{\mu}_i} \right\}^2$$

Se introducen los *leverages* del modelo, h_i (los valores de la diagonal principal de la matriz *hat*, H , $H=X(X'X)^{-1}X'$) a fin de compensar la diferencia entre la esperanza de la suma de los residuos al cuadrado y la suma de la esperanza de la variable dependiente (distintos a causa de la estimación de los p coeficientes de la regresión).

El estadístico se distribuye como una χ^2 con un grado de libertad.

Conceptos fundamentales del análisis de supervivencia.

Puede suponerse que el tiempo de supervivencia de un individuo (llamémoslo T) es una variable aleatoria (continua) que toma valores no negativos. La distribución de probabilidad de T puede caracterizarse de muchas formas. En el análisis de supervivencia, sin embargo, se utilizan la *función de densidad*, la *función de supervivencia* y la *función de riesgo*.

Denotemos por $f(t)$ a la *función de densidad* de T :

$$f(t) = \lim_{\Delta t \rightarrow 0} \left[\frac{\text{Prob}(t \leq T \leq t + \Delta t)}{\Delta t} \right]$$

La función de distribución de probabilidad acumulada se expresaría:

$$F(t) = \text{Prob}(T \leq t) = \int_0^t f(s) ds$$

La probabilidad de supervivencia hasta t , denominada *función de supervivencia*, $S(t)$ se define como:

$$S(t) = \text{Prob}(T \geq t) = 1 - \text{Prob}(T \leq t) = 1 - F(t)$$

Es decir:

$$f(t) = \lim_{\Delta t \rightarrow 0} \left[\frac{\text{Prob}(t \leq T \leq t + \Delta t)}{\Delta t} \right] = \frac{-dS(t)}{dt}$$

Suponiendo un paciente vivo en t , ¿cuál es la probabilidad que sobreviva el siguiente intervalo de tiempo Δt (considerando éste lo suficientemente pequeño)? la *función de riesgo*, $h(t)$, responde a esta pregunta:

$$h(t) = \lim_{\Delta t \rightarrow 0} \left[\frac{\text{Prob}(t \leq T \leq t + \Delta t | T \geq t)}{\Delta t} \right] = \frac{f(t)}{S(t)}$$

De hecho, $h(t)$ es la función de densidad condicional en t dada la supervivencia hasta t , representando la *tasa de mortalidad instantánea* para un individuo que ha sobrevivido hasta t .

Pueden derivarse otras relaciones útiles entre la función de riesgo, la función de densidad y la función de supervivencia:

$$h(t) = \frac{-d \text{Ln} S(t)}{dt} \quad \text{y} \quad f(t) = S(t)h(t)$$

Otra función de interés es la *función integrada o acumulada de riesgo*, $H(t)$:

$$H(t) = \int_0^t h(s) ds \quad \text{y}$$

$$S(t) = e^{-H(t)} \quad \text{por tanto} \quad H(t) = -\text{Ln} S(t)$$

Aproximación de Andersen-Gill al modelo de Cox. El modelo de regresión de riesgos proporcionales para procesos contadores multivariantes.

Consideremos un proceso contador, $N_{hi}(t)$, en el que el subíndice $i=1,2,\dots,n$ se refiere a individuos y el $h=1,\dots,k$ a distintos tipos de sucesos que un individuo puede experimentar a lo largo de su vida. A cada $N_{hi}(t)$ le corresponde un *proceso de intensidad*, $\lambda_{hi}(t)$ (suponiendo que se satisfacen ciertas condiciones de regularidad). De hecho, en el instante t , $\lambda_{hi}(t)\Delta t$ no es más que la probabilidad condicional de un salto de $N_{hi}(\cdot)$ en el (pequeño) intervalo temporal $[t,t+\Delta t)$, (siendo $\Delta t > 0$), dada toda la historia hasta t , pero sin incluirlo (llamémosla F_{t-}) (Andersen, 1992).

Definiendo el siguiente proceso contador, $N_i(t)=I(X_i \leq t, D_i=1)$. Es decir el proceso (para el individuo i) es igual a 0 mientras $t < X_i$ y cuando $t \geq X_i$ el proceso toma el valor 1 si al individuo le ha ocurrido el suceso ($D_i=1$) y 0 en caso contrario ($D_i=0$). Suponiendo ahora que la función de riesgo se puede expresar como:

$$h_i(t) = h_0(t)e^{\beta \cdot X_i(t)}$$

y que la 'historia' incluye la información de supervivencia, de censura y de posibles variables explicativas para el individuo i en el intervalo desde 0 a t , el proceso de intensidad se puede expresar como:

$$\lambda_i(t) = h_0(t)e^{\beta \cdot X_i(t)} Y_i(t)$$

siendo $Y_i(t)=I(X_i \geq t)$, es decir un indicador (con el valor 1) de si el individuo i se observa en el conjunto de riesgo justo antes de t .

Así, $N_{hi}(t)$ tiene un proceso de intensidad dado por,

$$\lambda_{hi}(t) = Y_{hi}(t)h_0(t)e^{\beta \cdot X_{hi}(t)}$$

En este modelo β se puede estimar como solución a la ecuación:

$$\frac{\partial C(\beta, \infty)}{\partial \beta} = 0$$

en la que:

$$C(\beta, t) = \sum_{h=1}^k \left[\sum_{i=1}^n \int_0^t \beta' X_{hi}(s) dN_{hi}(s) - \int_0^t \log \left\{ \sum_{i=1}^n Y_{hi}(s) e^{\beta' X_{hi}(s)} \right\} dN_h(s) \right]$$

y $N_h = N_{h1} + \dots + N_{hn}$ (**Andersen et al., 1992**).

Las intensidades basales integradas,

$$A_{h0}(t) = \int_0^t h_{h0}(s) ds \quad h = 1, \dots, k$$

Pueden ser estimadas mediante (estimador de Breslow, también denominado de Link, Tsiatis y de Nelson-Aalen):

$$A_{h0}(t) = \int_0^t \left\{ \sum_{i=1}^n Y_{hi}(s) e^{\beta' X_{hi}(s)} \right\}^{-1} dN_h(s)$$

Los estimadores de β y $A_{h0}(t)$ son asintóticamente normales y consistentes (**Andersen et al., 1992**).

Cuando sólo existe un único tipo de suceso, es decir $h=1$, estas expresiones se reducen a la *verosimilitud parcial* del modelo de Cox estándar (**Cox, 1975**),

$$L(\beta) = \prod_{i=1}^n \left(\frac{e^{\beta' X_i}}{\sum_{j \in R} e^{\beta' X_j}} \right)^{D_i}$$

siendo R el conjunto de riesgo y D_i definido previamente.

$$A_0(t) = \sum_{Ti \leq t} \frac{D_i}{\sum_{j \in R} e^{\beta' X_j}}$$

Estimación robusta de la matriz de covarianzas en la aproximación AG.

La implementación práctica de la estimación robusta, tal y como por ejemplo la realiza el programa S-Plus, es bastante sencilla. Llamemos D a la matriz de las variaciones en los estimadores de los parámetros resultante de ir eliminando observación tras observación (es decir primero se elimina una observación, se estima el modelo sin ella, se vuelve a introducir la observación; luego se elimina la siguiente observación, se estima el modelo sin ella, etc). La matriz D no es más que la matriz de los residuos *leverage* (denominados *dfbeta* en S-Plus o SAS) (**Cain y Lange (1984)** y **Reid y Crépeau (1985)**). Éstos se definen como los residuos de Schoenfeld, o de score, L (**Schoenfeld (1982)**, **Therneau et al. (1990)**) escalados por la matriz de covarianzas de los estimadores de los parámetros β , V. Así $D=LV^{-1}$ y,

$$L_{ij} = \int Z_{ij}(s) - \bar{Z}_j(s) d\hat{M}_i(s)$$

siendo M_i los residuos de martingala, la diferencia para cada individuo entre el número de sucesos observado y esperado (según el modelo estimado):

$$\hat{M}_i = N_i(t) - \int e^{\beta'Z_i(s)} Y_i(s) d\hat{A}_0(s)$$

Therneau y Hamilton (1997) señalan que las sumas de las columnas de D, es decir $1'D$, son iguales a las variaciones en los estimadores de β en cada una de las iteraciones del método de Newton-Raphson y, por tanto, $1'D=0$ en la iteración final.

Cuando las observaciones se presentan agrupadas y están correlacionadas dentro de cada grupo, como es nuestro caso, es el grupo (individuo en nuestro caso) y no la observación, el eliminado cada vez. Ello permite definir la matriz D^* cuyas filas serán un estimador del leverage del grupo k. Así pues, $D^{*'}D^*$ no será más que el estimador jackknife agrupado de la matriz de covarianzas (aproximadamente). Así pues, el estimador robusto de la matriz de varianzas sería igual a $D^{*'}D^*$. Se puede demostrar que se trata de un estimador asintóticamente insesgado de la matriz de covarianzas de los estimadores de β (**Lin y Wei (1989)**, **Therneau y Hamilton (1997)**). **Lee et al. (1992)** muestran el uso del estimador robusto en el caso de sucesos del mismo tipo, cuando están correlacionados.

El estimador robusto es totalmente equivalente al estimador de 'independencia de trabajo' de los modelos GEE (**Liang y Zeger, 1986, 1988 y 1992**); al estimador *sandwich* de **Huber (1967)** y al estimador de **White (1980 y 1982)**. El estimador robusto $D'D$ es mencionado de pasada en **Reid y Crépeau (1985)**. Desde otro contexto, **Lin y Wei (1989)** sugieren la utilización de tal estimador para contrastar la hipótesis de riesgos proporcionales.

Método marginal de Wei, Lin y Weissfeld (WLW).

Para cada una de las k recurrencias que pueda padecer el individuo i ($k=1,2,\dots,K$), la función de riesgo puede expresarse como:

$$h_{ki}(t) = h_{k0}(t)e^{\beta_k' X_{ki}(t)}$$

Se trata de estimar, para cada recurrencia k , β_k por máxima verosimilitud parcial. Estos estimadores son consistentes si la función de riesgo está correctamente especificada (**Wei et al., 1989**).

Cuando N es grande (aunque según **Wei et al. (1989)** la aproximación es también bastante precisa para muestras pequeñas y medianas) el vector de estimadores $(\beta_1, \beta_2, \dots, \beta_k)'$ (el subíndice indica recurrencia) es aproximadamente normal con matriz de covarianzas Q . El efecto para una variable explicativa determinada puede ser calculado como una combinación lineal de los parámetros asociados a tal variable en cada una de las recurrencias. En concreto,

$$\beta = \sum_{k=1}^K c_k \beta_k \quad \text{siendo } c = (c_1, c_2, \dots, c_k)' \quad \text{y } c = (e' \psi e)^{-1} \psi^{-1} e$$

$e=(1,\dots,1)'$ y ψ es el estimador de la covarianza de los β_k derivado de Q .

La expresión del estimador de la matriz de covarianzas Q proporcionada por **Wei et al. (1989)** es extremadamente compleja, en concreto:

$$Q = \frac{1}{n} \begin{bmatrix} D_{11}(\beta_1, \beta_1) & \dots & D_{1k}(\beta_1, \beta_k) \\ \dots & \dots & \dots \\ D_{k1}(\beta_k, \beta_1) & \dots & D_{kk}(\beta_k, \beta_k) \end{bmatrix}$$

siendo,

$$D_{kl}(\beta_k, \beta_l) = A_k' \beta_k B_{kl} \beta_l A_l' \beta_l$$

donde,

$$A_k \beta = \frac{1}{n} \sum_{j=1}^n D_{kj} \left[\frac{\sum_{t=1}^n Y_{kt}(X_{kj}) Z_{kt}(X_{kj}) e^{\beta' Z_{kt}(X_{kj})}}{\sum_{t=1}^n Y_{kt}(X_{kj}) e^{\beta' Z_{kt}(X_{kj})}} \right] - \left[\frac{\sum_{t=1}^n Y_{kt}(X_{kj}) Z_{kt}(X_{kj}) e^{\beta' Z_{kt}(X_{kj})}}{\sum_{t=1}^n Y_{kt}(X_{kj}) e^{\beta' Z_{kt}(X_{kj})}} \right]^{\otimes 2}$$

en la que $a^{\otimes 2}$ denota la matriz aa' construida a partir del vector columna a.

$$B_{kl}(\beta_k, \beta_l) = \frac{1}{n} \sum_{j=1}^n W_{kj}(\beta_k) W_{lj}(\beta_l)'$$

Además,

$$W_k(\beta_k) = D_{kj} \left\{ Z_{kj}(X_{kj}) - \frac{S_k^{(1)}(\beta_k; X_{kj})}{S_k^{(0)}(\beta_k; X_{kj})} \right\} - \sum_{m=1}^n \frac{D_{km} Y_{kj}(X_{km}) e^{\beta_k' Z_{kj}(X_{km})}}{n S_k^{(0)}(\beta_k; X_{km})} \times \left\{ Z_{kj}(X_{km}) - \frac{S_k^{(1)}(\beta_k; X_{km})}{S_k^{(0)}(\beta_k; X_{km})} \right\}$$

y en el que,

$$S_k^{(1)}(\beta; t) = \frac{1}{n} \sum_{t=1}^n Y_{kt}(t) Z_{kt}(t) e^{\beta' Z_{kt}(t)} \quad S_k^{(0)}(\beta; t) = \frac{1}{n} \sum_{t=1}^n Y_{kt}(t) e^{\beta' Z_{kt}(t)}$$

Afortunadamente la expresión de dicha matriz es totalmente equivalente a la siguiente (Therneau y Hamilton, 1997):

$$Q = \begin{pmatrix} D'_1 D_1 & D'_1 D_2 & D'_1 D_3 \\ D'_2 D_1 & D'_2 D_2 & D'_2 D_3 \\ D'_3 D_1 & D'_3 D_2 & D'_3 D_3 \end{pmatrix}$$

siendo D_1 la matriz de los residuos $d\beta$ correspondientes al ajuste del modelo en la primera recurrencia, D_2 los del ajuste en la segunda, etc (se han supuesto tres recurrencias).

Pero, además, **Therneau y Hamilton (1997)** muestran que el método WLW original es algebraicamente equivalente a estimar β y calcular el estimador robusto de la matriz de covarianzas, $D^{-1}D^{-1}$, en un modelo que combine *todos* los estratos pero que estratifique por el número de recurrencias del suceso (es muy importante tener en cuenta la organización de los datos vista más arriba).

Método de Prentice, William y Peterson (PWP).

Prentice, William y Peterson (1981) proponen dos modelos semi-paramétricos de riesgos proporcionales:

a) tiempo desde el inicio del estudio (de *tiempo total*):

$$h_i(t) = h_{0s}(t)e^{(\beta s' \cdot X_i(t))}$$

b) tiempo desde la recurrencia inmediatamente precedente (de *brecha de tiempo*):

$$h_i(t) = h_{0s}(t - t_n)e^{(\beta s' \cdot X_i(t))}$$

La particularidad de ambos modelos reside en que se estratifica en función del número de recurrencias previas.

