

Influencia de las condiciones de contratación en el absentismo laboral

Iván Fernández-Suárez, María Botey, and Fermín Torrano*

Universidad Internacional de La Rioja (España)

Resumen: El propósito de esta investigación es analizar las relaciones entre las condiciones de contratación y el absentismo laboral en una muestra de 5524 trabajadores, con el fin de identificar qué segmentos (por tipo de contrato y jornada, tiempo contratado, antigüedad en la empresa y bajas por enfermedad común ocurridas en los tres últimos años) están más relacionados con la posibilidad de sufrir un proceso de incapacidad temporal en el año 2017. Se realizaron análisis descriptivos, la prueba chi-cuadrado para tablas de contingencia con dos muestras independientes y los árboles de decisión, basados en el algoritmo CHAID (*Chi-squared Automatic Interaction Detection*), para detectar las variables más importantes en la identificación de perfiles con una mayor probabilidad de sufrir una incapacidad temporal derivada de contingencias comunes. Los resultados ponen de manifiesto la existencia de diferencias entre las variables estudiadas. Se considera la modalidad de contratación un factor de riesgo importante del absentismo laboral.

Palabras clave: Absentismo laboral; Contrato; Segmentación jerárquica; CHAID.

Title: Influence of contracting conditions on absenteeism.

Abstract: The aim of this paper is to analyze the relationship between hiring conditions and work absenteeism in a sample of 5524 workers in order to identify which segments (by type of contract and workday, time hired, seniority in the company and sick leaves occurred in the last three years) are more related to the possibility of suffering a temporary incapacity in 2017. Descriptive analyzes, the chi-square test for contingency tables with two independent samples, and the decision trees based on the CHAID algorithm (*Chi-squared Automatic Interaction Detection*) were carried out to detect the most important variables in the identification of profiles with a greater probability of suffering a temporary incapacity for common contingencies. The results show the existence of differences between the variables studied. The hiring modality is considered an important risk factor for work absenteeism.

Keywords: Work absenteeism; Contract; Hierarchical segmentation; CHAID.

Introducción

Actualmente el absentismo laboral es una de las principales preocupaciones en los planos económico, social y empresarial, debido a que la ausencia del trabajador de su puesto de trabajo, en sus diferentes formas y causas, provoca pérdidas productivas, altera el clima organizacional y genera un elevado coste al Sistema público de Seguridad Social (Hassard, Teoh, Visockaite, Dewe y Cox, 2018; López-Alonso, Ibarrodo-Dávila y Rubio-Gámez, 2015; Mazzoni et al., 2019). En España el absentismo por incapacidad temporal (por contingencia profesional y común) supone 13.09 días perdidos por cada trabajador en activo, una de las más elevadas a nivel internacional (Blasco de Luna et al., 2018). En cuanto al absentismo por enfermedad común, los datos son aún más negativos. Según el VII Informe Adecco sobre absentismo (cfr. Blasco de Luna et al., 2018), en 2017 el 86.4% de los procesos de incapacidad temporal corresponden a procesos de enfermedad común, con una duración media de 17.21 días de baja. Esto ha supuesto un aumento, respecto al 2016, de un 6%. El elevado coste que supone el absentismo, sumado a las repercusiones personales y sociales que acarrea en las organizaciones, ha significado que su control sea uno de los principales objetivos estratégicos de cualquier empresa, región o estado (Martín-Fumadó, Martí Amengual, Puig Bausili y Arimany-Manso, 2014; Martín-Román y Moral, 2017).

El absentismo es considerado por diversos autores como un proceso multidimensional (Altunkaynak, 2018; Khashaba,

El-Helaly, El-Gilany, Motawei y Foda, 2018; Thurston y Glendon, 2018), que ha sido explicado tanto por factores individuales, como la edad (Fernández-García, 2013), el sexo (Campos-Serna, Ronda-Pérez, Artazcoz y Benavides, 2012), la nacionalidad del trabajador (Casey, Riseborough y Krauss, 2015) o el consumo de alcohol (Luján, Santana, Alemán y Sánchez, 2014), como por las consecuencias de las diferentes condiciones sociales, económicas y organizativas a las que el trabajador está sometido en su entorno laboral (Boada i Grau, De Diego Vallejo, Agulló Tomás y Mañas Rodríguez, 2005). Respecto de las condiciones laborales, un gran número de investigaciones ponen de relieve la relación entre el absentismo y distintas variables organizacionales, como el salario, la estabilidad laboral y el tipo de contrato (de Menezes y do Monte, 2013). En concreto, se observa que las personas que presentan un peor estado de salud y, por tanto, son más propensas a ausentarse, son aquellos trabajadores con contratos temporales y expuestos a peores condiciones laborales (Virtanen et al., 2005).

La precariedad laboral, caracterizada por la inseguridad, la incertidumbre y la falta de garantías en las condiciones de trabajo, ha sido utilizada como un indicador válido para analizar el absentismo, ya que ha sido identificada como uno de sus factores de riesgo (Jiang y Lavaysse, 2018) y del presentismo laboral (Bogaerts, De Witte, Verlinden y Vermeulen, 2017). Al mismo tiempo, las diferencias en las condiciones de contratación pueden ser también determinantes en el absentismo (García, Green y Navarro Paniagua, 2018; Restrepo y Salgado, 2013). Específicamente, los estudios atribuyen una especial relevancia al tipo de contrato, ya que se ha evidenciado mayores índices de absentismo en los trabajadores con contratos eventuales que en el caso de los indefinidos (Zaballa et al., 2016). De igual forma, estas investigaciones apuntan a que la duración de la jornada establecida en el

* Correspondence address [Dirección para correspondencia]:
Fermín Torrano. Universidad Internacional de La Rioja. Avenida de la Paz
137, 26006 Logroño (España). E-mail: fermin.torrano@unir.net
(Artículo recibido: 04-04-2019; revisado: 07-01-2020; aceptado: 17-02-2020)

contrato laboral muestra una estrecha relación con los niveles de absentismo y la salud de los trabajadores (Benavides, Benach, Diez-Roux y Roman, 2000). Por ejemplo, los trabajadores con jornadas parciales temporales presentan mayores valores de insatisfacción laboral y presencia de síntomas de trastornos musculoesqueléticos (Benavides y Benach, 1999).

Objetivos e hipótesis

Bajo el marco de referencia descrito anteriormente, el objetivo de la presente investigación es analizar las variables de carácter contractual que tienen influencia en el absentismo laboral. Se pretende de esta manera describir qué factores de contratación (tipo de contrato y jornada, tiempo contratado, antigüedad y enfermedades sufridas en los últimos años y duración) podrán influir en la propensión a ausentarse en el ámbito laboral.

Este objetivo principal es detallado a través de diferentes hipótesis de trabajo:

- ¿Son los trabajadores con contrato temporal los más propensos a sufrir incapacidades temporales derivadas de procesos de contingencia común?
- ¿Influye la jornada de trabajo en la probabilidad de sufrir más incapacidades temporales por contingencia común?
- ¿Existe relación entre la antigüedad en el puesto de trabajo y la posibilidad de sufrir incapacidades temporales por contingencia común?
- ¿Son las incapacidades temporales por contingencia común sufridas en los últimos años una variable predictora potencial del absentismo actual?

Método

Participantes

La muestra, de carácter incidental y no probabilístico, está formada por la plantilla completa de trabajadores contratados en el año 2017, 5.524 personas, en una empresa española del sector servicios, dedicada a tareas de mantenimiento y limpieza. En la Tabla 1 se muestran las características de la muestra en cuanto a sus diferentes modalidades de contratación.

Tabla 2. Variables objeto de estudio y codificación.

Variable	Descripción	Posibles valores
TContrato	Tipo Contrato	0 = contrato fijo 1 = contrato eventual
Jornada	Tipo Jornada	0 = jornada completa 1 = jornada parcial
TpoContratado	Tiempo Contratado en 2017	1 = de 0 a 2 meses 2 = de 2 meses a 8 meses 3 = más de 8 meses
DíasContrato	Duración del Contrato 2017 en días	De 0 a 360
Antig	Antigüedad	1 = Menos de 1 año 2 = Entre 1 y 3 años 3 = Entre 4 y 15 años 4 = Más de 15 años

Tabla 1. Características contractuales de la muestra.

Variable y grupo	N	%
Sexo		
Hombres	814	14.7%
Mujeres	4710	85.3%
Tipo de contrato		
Eventual	2653	48.03%
Fijo	2871	51.97%
Tipo de Jornada		
Completa	2130	38.56%
Parcial	3394	61.44%
Días de contrato 2017		
De 0 a 2 meses	1392	25.20%
De 2 meses a 8 meses	1331	24.09%
Más de 8 meses	2801	50.71%
Antigüedad		
Menos de 1 año	1432	25.92%
De 1 año a 3 años	1201	21.74%
De 4 años a 15 años	1651	29.89%
Más de 15 años	1240	22.45%

Al mismo tiempo, para cada trabajador, se obtuvo información relativa a su absentismo por enfermedad común en el año 2017. En concreto, se consideró el número de procesos por enfermedad común que había sufrido y los días de absentismo en ese año.

En el desarrollo del estudio se equipara el término de enfermedad común con el concepto de enfermedad.

Variables de estudio

Teniendo en cuenta lo anterior, las variables que se han utilizado corresponden (ver Tabla 2), por un lado, a los datos de contratación del trabajador en el año 2017 (tipo de contrato y de jornada, tiempo contratado en 2017 y antigüedad), y por otro, al número de bajas por enfermedad sufridas por cada trabajador en los 3 últimos años (2015, 2016 y 2017), así como su duración. También, se calculó el ratio de absentismo por enfermedad, resultante de dividir el número de días de absentismo por el número de días de contrato en el año 2017.

Variable	Descripción	Posibles valores
Procesos_EnfXX	Número de procesos por enfermedad que han ocurrido en el año 20XX (XX = 17. 16. 15)	de 0 a 10
Proc_XXenf01	Variable dicotómica asociada a Procesos_EnfXX	0 = no ha sufrido ningún proceso por enfermedad en el año XX 1 = ha sufrido algún proceso por enfermedad en el año XX
Durac_EnfXX	Duración procesos por enfermedad en el año 20XX (XX =17. 16. 15)	desde 0 hasta 360 días
Ratio_Enfermedad	Durac_EnfXX/DíasContrato	desde 0 hasta 1
Ratio_Enf01	Variable dicotómica asociada a Ratio_Enfermedad	0 = si Ratio_Enfermedad < .5 1 = si Ratio_Enfermedad ≥ .5

Procedimiento

El procedimiento de obtención de los datos se ha realizado a través del programa ERP (*Enterprise Resource Planning*) implantado en la organización objeto de estudio. El control estadístico del absentismo laboral se ha desarrollado a través de las aplicaciones informáticas Navision y Meta4, en las que se han introducido todos los datos relativos a las ausencias por contingencias comunes y profesionales de los trabajadores. De igual forma, en ambos aplicativos se han incluido todas las variables sociodemográficas y contractuales que han formado parte del presente estudio.

La introducción de la información relativa al absentismo ha sido efectuada por el departamento de Recursos Humanos de dicha organización. Asimismo, desde el departamento de Prevención de Riesgos Laborales se ha realizado la investigación de todos los procesos de accidente de trabajo y enfermedad profesional y el análisis estadístico del absentismo derivado de ambos procesos.

Análisis estadísticos

En primer lugar, se realizó un análisis descriptivo de los datos (frecuencias y porcentajes) a través de tablas de contingencia. Por otro lado, se empleó la prueba chi-cuadrado para tablas de contingencia con dos muestras independientes, con objeto de conocer la existencia de diferencias significativas entre los grupos considerados. Finalmente, se utilizó el método CHAID (*Chi-squared Automatic Interaction Detection*) (Kass, 1980), que, como se sabe, es una técnica no paramétrica de análisis de segmentación de variables, adecuada para grandes tamaños muestrales y para variables dependientes categóricas, como es el caso, que permite agrupar elementos descritos por una variable dependiente en función de los valores de otras variables independientes (predictores) (Ortega, 2015).

El método CHAID analiza todos los valores de cada variable predictora potencial a través del estadístico de contraste chi-cuadrado, el cual refleja cuan similares o asociadas están las variables. A partir de aquí, selecciona el predictor más significativo para formar la primera partición en el árbol de decisión, de tal forma que cada nodo está conformado por

aquellas categorías similares de la variable seleccionada. El proceso continúa sucesivamente hasta que el árbol queda completado (Sanz y Ponce de León, 2010).

En este tipo de modelos inductivos resulta fundamental evaluar cómo se comporta la estructura del árbol para generalizar los datos disponibles en una muestra de datos diferente. La técnica más utilizada para esta validación consiste en dividir los datos en una muestra de entrenamiento y una muestra de validación o test. Si el modelo generado con los datos de entrenamiento demuestra una capacidad predictiva similar con los datos de validación, entonces las reglas se podrían generalizar. El árbol se evalúa con la tabla de confusión que muestra la proporción de casos clasificados correctamente por el algoritmo respecto al total de la muestra clasificada y la proporción de incorrectos.

La selección del árbol óptimo está relacionado con el problema del sobreajuste. Normalmente, a mayor complejidad mayor sobreajuste. En el método aplicado a este estudio se definieron parámetros de “autopoda” que evitan el sobreajuste y generan árboles con poca complejidad (sólo tres niveles de división y nodos “hijo” con más de 100 individuos), que proporcionaron un modelo estable, con pocas diferencias entre la muestra de entrenamiento y la de test.

La elección de esta técnica es debido a sus buenos resultados tanto en población normal (Peñalva-Vélez y López-Goñi, 2014), como clínica (López-Goñi, Fernández-Montalvo y Arteaga, 2012).

Todos los análisis fueron realizados con el paquete IBM®-SPSS Statistics®.

Resultados

En primer lugar, se examinaron las interrelaciones entre el número de procesos de enfermedad ocurridos en los tres últimos años (años 2015, 2016 y 2017) y las variables tipo de contrato, tipo de jornada y tiempo contratado (ver Figura 1). Para analizar estos datos de forma comparativa, se dividió el tamaño de cada grupo por el número total de individuos. También, se tuvo en cuenta el número de meses contratado para eliminar en lo posible el efecto tiempo de exposición, de manera que los datos son relativos a un mes de trabajo.

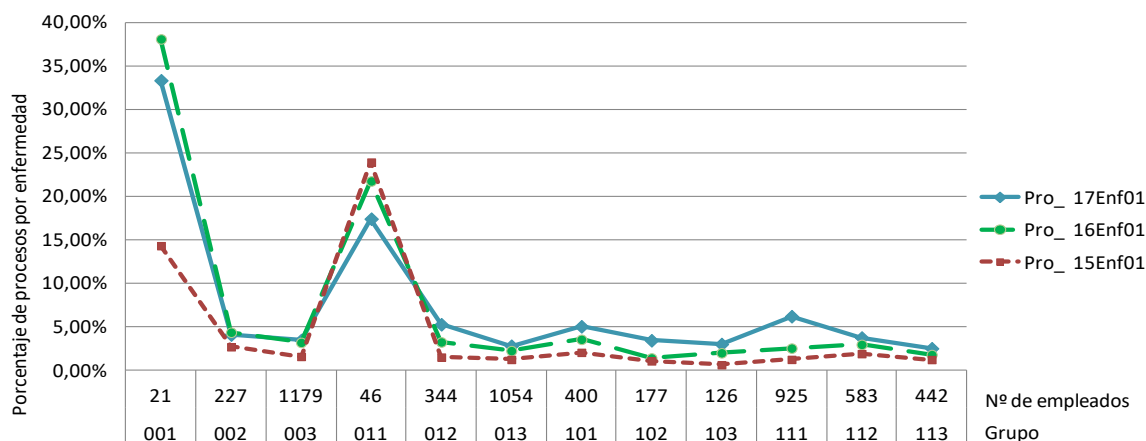


Figura 1. Porcentaje de procesos por enfermedad en cada grupo por mes trabajado.

Nota: Grupo = ABC, dónde A es el Tipo de contrato (0 = Fijo, 1 = Eventual), B es el Tipo de Jornada (0 = Completa, 1 = Parcial) y C son los Meses trabajados en el año (1 <= 2 meses, 2 <= 8 meses, 3 > 8 meses).

Como se puede observar, el número de procesos por enfermedad en el año 2017 mantiene la misma distribución entre grupos que en los dos años anteriores. Destaca la diferencia existente entre los valores altos de la primera mitad del gráfico (contrato fijo) frente a valores bajos en la segunda mitad (contrato eventual). También, se observa que los grupos asociados a contratos de duración inferior a dos meses presentan mayor proporción de personas con algún proceso por enfermedad que los de duración más elevada.

Si se comparan estos resultados con la media de ratio por enfermedad en cada grupo, se observan ciertas variaciones en lo relativo a los grupos de contrato inferior a dos meses. Es decir, como se puede observar en la Figura 2, los grupos con contratos de corta duración presentan una proporción superior de personas con algún proceso por enfermedad en 2017, pero con una ratio de días de absentismo inferior que los grupos de tiempo de contratación superior.

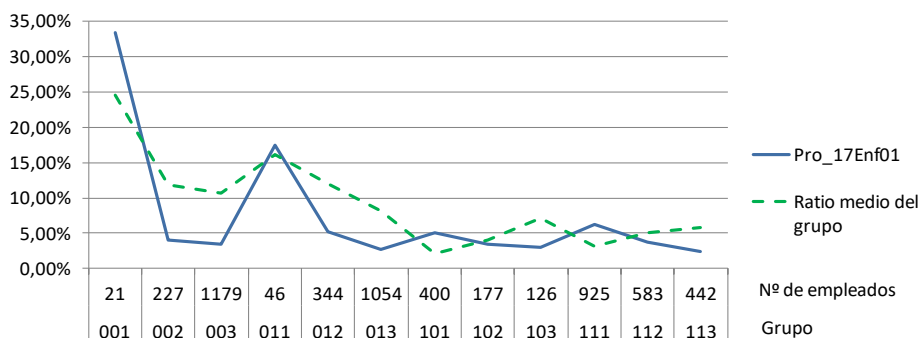


Figura 2. Porcentaje de procesos y ratio por enfermedad en 2017.

En segundo lugar, se realizaron cuatro pruebas chi-cuadrado de la variable Grupo de las Figuras 1 y 2 frente a cada una de las variables consideradas (Pro_17Enf01

, Pro_16Enf01, Pro_15Enf01 y Ratio_Enf01). En todos los casos se obtuvieron diferencias significativas entre los diferentes grupos considerados (Tabla 3).

Tabla 3. Pruebas chi-cuadrado de la variable Grupo frente a las variables Pro_17Enf01, Pro_16Enf01, Pro_15Enf01 y Ratio_Enf01.

	gl	Pro_17Enf01		Pro_16Enf01		Pro_15Enf01		Ratio_Enf01	
		Valor	p.	Valor	p.	Valor	p.	Valor	p.
Chi-cuadrado de Pearson	11	403.832 ^a	.000	462.085 ^b	.000	199.587 ^c	.000	94.297 ^d	.000
Razón de verosimilitudes	11	454.306	.000	528.448	.000	237.862	.000	89.507	.000
Nº de casos válidos		5524		5524		5524		5524	

a. 1 casillas (4.2%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es 4.90.

b. 1 casillas (4.2%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es 3.92.

c. 2 casillas (8.3%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es 2.09.

d. 2 casillas (8.3%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es 1.23.

Por último, se utilizaron los árboles de decisión para identificar qué tipo de perfiles presentan una mayor probabilidad de sufrir una enfermedad en el año 2017. En concreto, se analizaron dos árboles de decisión, uno por cada variable considerada (procesos enfermedad 2017 y ratio enfermedad).

Inicialmente, para el primer árbol, se dividió aleatoriamente la muestra en dos conjuntos: muestra entrenamiento (60%) y muestra test (40%). A continuación, se seleccionaron

las variables explicativas, tipo de contrato, tipo de jornada, tiempo contratado, antigüedad y número de procesos por enfermedad en el año 2016 y 2015. Se definieron algunos parámetros de autopoda del árbol: tres niveles de división y nodos hijo con más de 100 individuos, con el objetivo de garantizar la obtención de un árbol no sobreajustado. Los resultados obtenidos se muestran en la Figura 3:

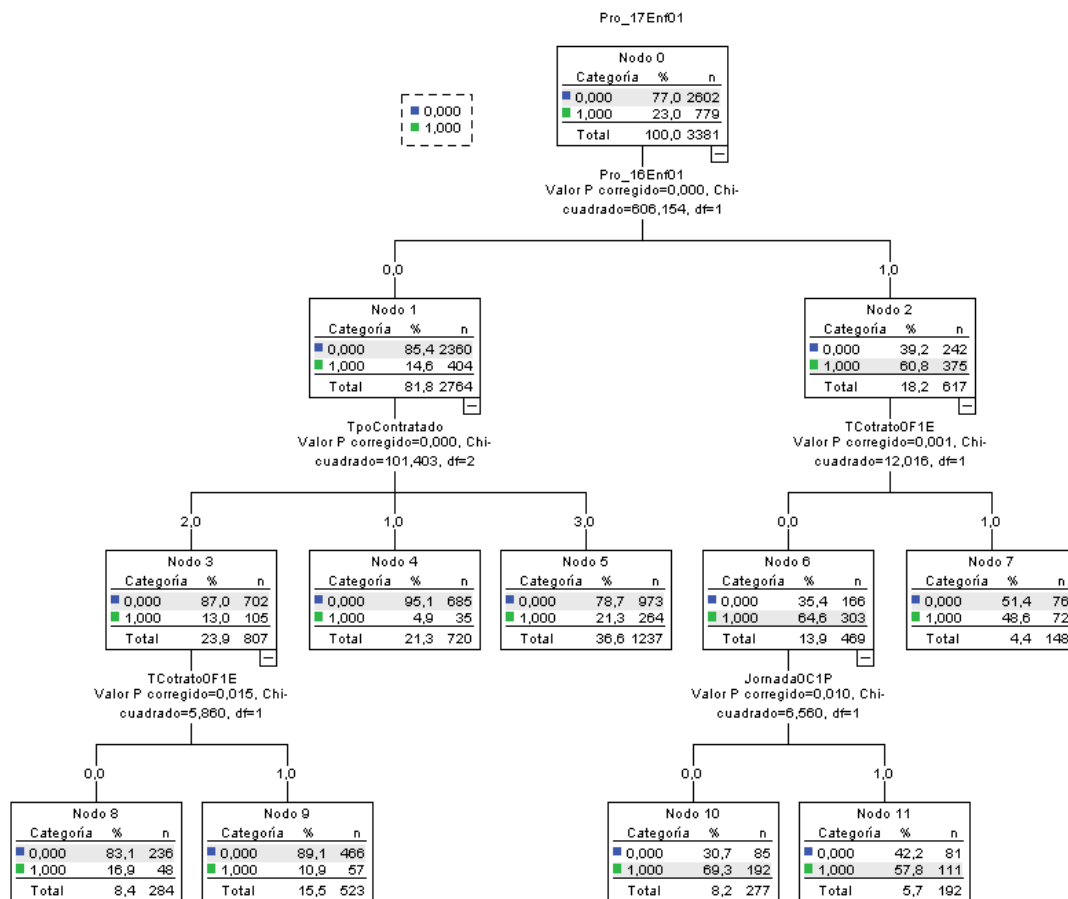


Figura 3. Árbol de decisión para la variable procesos enfermedad 2017.

El árbol obtenido proporciona para cada nodo un porcentaje de personas sin absentismo y otro complementario de personas con absentismo. El algoritmo utilizado para determinar cuál es el valor de predicción para todas las personas de un nodo es el siguiente: se predice 0 si su probabilidad asociada es superior a un valor fijado en la frontera de decisión y 1 cuando es inferior.

Para interpretar correctamente el árbol, se tuvo en cuenta la importancia o peso de cada variable independiente en el árbol, la cual depende de con qué profundidad aparece la variable en el árbol y el número de veces que está presente. La variable de mayor importancia fue la existencia o no de procesos por enfermedad en el año anterior (con un 100%); la segunda variable, pero con una menor relevancia, son los

procesos por enfermedad correspondientes a los dos años anteriores (con un 20%); el resto de variables presentan una influencia inferior al 15%.

La matriz de confusión obtenida para un valor umbral (*threshold* o frontera de decisión) igual a .50 proporciona un porcentaje de aciertos del 81% en la muestra de entrenamiento y del 79.6% en la muestra test, lo que indica que es un árbol estable (la variación entre ambas muestras es muy baja) y con un porcentaje de aciertos muy elevado.

La evaluación del rendimiento de clasificadores binarios se realizó a través de la curva ROC (*Receiver Operating Characteristic*) y PRC (*Precision-Recall*) (Figura 4). Si bien existen varios indicadores o medidas para evaluar un predictor para un

clasificador binario, se tomó como referencia el trabajo de Saito y Rehmsmeier (2015), como se puede ver en la Tabla 4.

Tabla 4. Descripción de medidas de evaluación de predictores para clasificadores binarios de Saito y Rehmsmeier (2015).

Medida	Observed	Predicted	
		0-negative 1-positive	TN-True Negative FN-False Negative
ACC	Fórmula		
ERR		$(TP+TN)/(TP+TN+FN+FP)$	$(FP+FN)/(TP+TN+FN+FP)$
SN, TPR, REC		$TP/(TP+FN)$	
SP		$TN/(TN+FP)$	
FPR		$FP/(TN+FP)$	
PREC, PPV		$TP/(TP+FP)$	
MCC		$(TP*TN-FP*FN)/((TP+FP)*(TP+FN)*(TN+FP)*(TN+FN))^{1/2}$	
F _{0,5}		$1,5*PREC*REC/(0,25*PREC+REC)$	
F ₁		$2*PREC*REC/(PREC+REC)$	
F ₂		$5*PREC*REC/(4*PREC+REC)$	

ACC: accuracy; ERR: error rate; SN: sensitivity; TPR: true positive rate; REC: recall; SP: specify; FPR: false positive; PREC: precision; PPV: positive value; MCC: Matthews correlation coefficient; F: F score; TP: true positive; FP: false positives; FN: false negatives

La curva ROC muestra la figura que se obtiene al representar la exactitud (ACC = *accuracy* o proporción de aciertos)

frente a la sensibilidad (SN = *sensitivity*). La curva asociada al problema de investigación proporciona un área del .693.

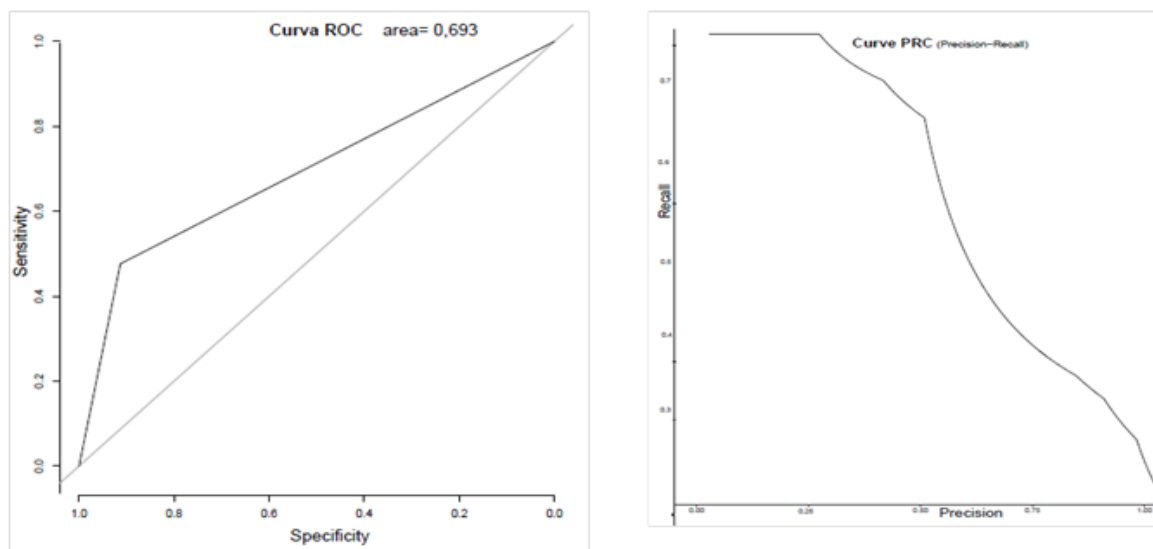


Figura 4. Curva ROC y curva PRC.

La muestra de este estudio presenta una variable objetivo desbalanceada (contiene sólo un 23% de empleados con algún proceso de enfermedad en 2017). Según Saito y Rehmsmeier (2015), para la evaluación del predictor en muestras desbalanceadas, resulta más conveniente utilizar la curva PRC, que representa la precisión (PREC = precisión) frente a la sensibilidad (SN = *sensibilidad* o REC = *recall*).

El objetivo de la presente investigación es identificar el máximo número de empleados con algún proceso por enfermedad, a pesar de perder precisión, es decir, maximizando la sensibilidad. Por tanto, según la curva PRC, se decide seleccionar un umbral igual a .18, que proporciona una matriz

de confusión con una sensibilidad igual a .82 y una precisión de .34.

A partir de lo anterior, se determina que los nodos 5, 7, 10 y 11 tienen un valor de predicción 1 (probabilidad de que existan procesos por enfermedad en 2017 > .18), por tanto, los trabajadores clasificados en estos nodos son el objetivo del presente estudio, pues su pronóstico es que sufren algún proceso por enfermedad en 2017.

La segmentación obtenida de la población de trabajadores es la siguiente:

- Segmento 1 (nodo 10): trabajadores con algún proceso por enfermedad en 2016, tipo de contrato fijo y jornada

completa (probabilidad .693, n° de aciertos = 111, n° errores = 81).

- Segmento 2 (nodo 11): trabajadores con algún proceso por enfermedad en 2016, tipo de contrato fijo y jornada parcial (probabilidad .578, n° de aciertos = 192, n° errores = 85).
- Segmento 3 (nodo 7): trabajadores con algún proceso por enfermedad en 2016, tipo de contrato eventual (probabilidad .486, n° de aciertos = 72, n° errores = 76).
- Segmento 4 (nodo 5): trabajadores sin ningún proceso por enfermedad en 2016, tiempo contratado superior a 8 meses (probabilidad .213, n° de aciertos = 264, n° errores = 973).

Por otro lado, a partir del segundo árbol de decisión se intentó identificar aquellos segmentos de trabajadores que

presentan un ratio por enfermedad superior al 50%, teniendo en cuenta las variables asociadas al contrato. En primer lugar, se seleccionaron los trabajadores que presentan algún proceso por enfermedad en 2017, obteniendo un total de 1290 trabajadores. Se calculó el ratio por enfermedad, dividiendo el número de días de absentismo por enfermedad entre el número de días contratado en 2017. A continuación, se dividió la muestra en dos grupos: muestra test (30%) y entrenamiento (70%) En esta ocasión, el grupo objetivo (trabajadores con ratio superior al 50%) supone un 25% de la muestra, por lo que, al igual que anteriormente, este árbol también es una muestra desbalanceada. Se seleccionaron las variables explicativas, tipo de contrato, tipo de jornada, tiempo contratado y antigüedad. Los resultados obtenidos se muestran en el árbol de decisión recogido en la Figura 5:

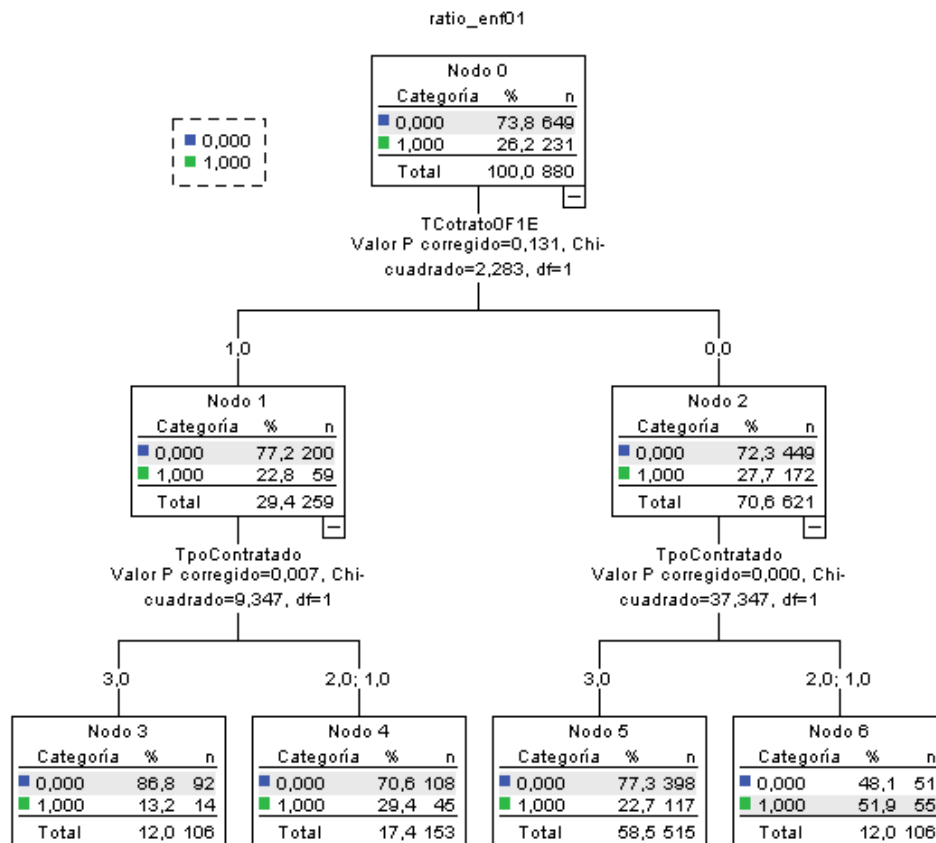


Figura 5. Árbol de decisión para la variable dependiente ratio_enf01.

Para interpretar este árbol, se dispone, al igual que antes, del indicador importancia o peso de cada variable independiente en el árbol, siendo la variable de mayor importancia el tiempo de contrato (con un 100%), seguido del tipo de contrato y la antigüedad (cada una con un 50%). Finalmente, el tipo de jornada no presenta importancia en este árbol.

La matriz de confusión obtenida para un valor umbral igual a .50 proporciona un porcentaje de aciertos del 74.2%

en la muestra de entrenamiento y del 76.6% en la muestra de test. Esto indica que es un árbol estable (la variación entre muestra de entrenamiento y de test es muy baja), con un porcentaje de aciertos muy elevado.

El umbral utilizado en esta matriz de confusión es de .5 y proporciona una sensibilidad igual a .23 y una precisión igual a .52. Por ser un árbol desbalanceado, se utilizó la curva PRC y se seleccionó el valor umbral igual a .25, que mejora la sen-

sitividad igual a .43, perdiendo cierto nivel de precisión (= .38).

Finalmente, los segmentos de población pronosticados cómo *ratio_enf01* superior al .5 fueron los siguientes:

- Segmento 1 (nodo 6): trabajadores con tipo de contrato fijo y tiempo contratado superior a 8 meses (probabilidad .52, n° de aciertos = 55, n° errores = 51).
- Segmento 2 (nodo 4): trabajadores con tipo de contrato eventual y tiempo contratado inferior a 8 meses. (probabilidad .29, n° de aciertos = 45, n° errores = 108).

Discusión

Los resultados obtenidos revelan la existencia de relaciones entre la modalidad de contratación y el absentismo laboral. En concreto, respecto de la primera hipótesis, si los trabajadores con contrato temporal son los más propensos a sufrir enfermedades, los resultados obtenidos no permiten verificar dicha hipótesis, ya que se evidencia que los trabajadores con este tipo de contrato (temporal) manifiestan un porcentaje menor de procesos por enfermedad común que los trabajadores con contrato fijo. Estos resultados contradicen distintos estudios previos, como el de Virtanen et al. (2005); no obstante, deben considerarse los efectos específicos de la actividad de la limpieza y el envejecimiento de la población, además de la mayor carga horaria asignada a los contratos indefinidos frente a los temporales, como variables mediadoras de los resultados, lo cual deberá ser estudiado por futuras investigaciones en este campo.

Por otro lado, se examinó si la jornada de trabajo influye en el absentismo. En relación con esta cuestión, los segmentos 1 y 2 del primer árbol CHAID ponen de relieve que la jornada parcial proporciona menor probabilidad de padecer algún proceso por enfermedad común que la jornada completa, lo que puede ser explicado por el tiempo de exposición, que es uno de los factores más relevantes en los problemas de salud física. Teniendo en cuenta que el absentismo laboral en el sector de limpieza está íntimamente relacionado con una etiología musculoesquelética, el resultado es perfectamente comprensible en coherencia con las investigaciones de Estry-Béhar y Van Der Heijden (2012), Gallis (2013), Moreno, Moncada, Llorens y Carrasquer (2010), y Vercruyssen y Van de Putte (2013).

En tercer lugar, se examinaron las relaciones entre la antigüedad y la posibilidad de sufrir enfermedades, evidenciándose que la antigüedad es una de las variables con mayor importancia en el caso de la *ratio* de enfermedad. Estas conclusiones deben tomarse con cautela, dada la peculiaridad y especificidad de la muestra, y la relación entre el tipo de contrato y la edad media de los trabajadores.

Por último, los hallazgos obtenidos permiten verificar la última hipótesis, ya que en los análisis efectuados la variable procesos por enfermedad común en los tres años de referencia presenta el mismo comportamiento. Asimismo, en el primer árbol CHAID, en el que participan las variables procesos por enfermedad común en el año 2015 y en el 2016, ambas son las que mayor importancia tienen. Y es que, a nuestro juicio, los antecedentes previos de absentismo son indicadores claros de absentismo futuro. La influencia sobre la salud de procesos anteriores puede ser una de las justificaciones, si bien se recomienda la inclusión de variables individuales, como la personalidad y la salud autopercebida, en futuros estudios.

Entre las limitaciones del estudio se destaca la falta de control sobre el pluriempleo, muy habitual en el sector afectado por procesos de subrogación, como es el sector de limpieza, lo cual puede afectar a las jornadas totales de trabajo de la población analizada. Por otro lado, el tipo de muestra no probabilística (por conveniencia) empleada provoca que los resultados no puedan ser ni generalizables ni tampoco concluyentes. De igual forma, la edad de los trabajadores, como señalan distintos estudios previos (p. ej., Fernández-García, 2013), también parece influir en los procesos de absentismo, por lo que en coherencia con el envejecimiento de la población trabajadora que está sufriendo España y Europa, se recomienda la inclusión de esta variable en estudios ulteriores.

Finalmente, si bien existe un gran número de métodos estadísticos para realizar las tareas propias de la clasificación (p. ej., métodos de regresión logística, análisis discriminante, redes neuronales, inducción de reglas, redes bayesianas, etc.), los árboles de decisión presentan ciertas ventajas, ya que permiten plantear de manera completa todas las posibles soluciones; además su representación gráfica facilita la interpretación (Chien, Wang y Cheng, 2007). Asimismo, Fleiss et al. (1972) destacaron los beneficios de esta técnica frente a otras en la generación de diagnósticos. Esta técnica de decisión y clasificación ha sido ampliamente empleada en el ámbito de la psicología cognitiva, la psicometría y la psicología de la creatividad (p. ej., Batchelder y Riefer, 1999; Chun et al., 2013). Sin embargo, al mismo tiempo, los árboles de decisión poseen distintas limitaciones, como la gran cantidad de datos que necesita (disponibles en este estudio) o el hecho de que, si se asignan costes y ganancias a cada una de las acciones, es posible aplicar como criterio de decisión maximizar el valor esperado, el cual exige que el decisor pueda soportar el riesgo de ruina. En la presente investigación el criterio de decisión ha sido el de identificar el máximo número de empleados con algún proceso por enfermedad (con un error moderado), sin evaluar los beneficios económicos que podrían aportar, cuestión que deberá ser objeto de estudio por parte de futuras investigaciones en este campo.

References

- Altunkaynak, B. (2018). A statistical study of occupational accidents in the manufacturing industry in Turkey. *International Journal of Industrial Ergonomics*, 66, 101-109. doi:10.1016/j.ergon.2018.02.012
- Batchelder, W.H., & Riefer, D.M. (1999). Theoretical and empirical review of multinomial process tree modeling. *Psychonomic Bulletin & Review*, 6, 57-86.
- Benavides, F., & Benach, J. (1999). Types of employment and health: Analysis of the Second European Survey on Working Conditions. *Gaceta Sanitaria*, 13(6), 425-430. doi:10.1016/S0213-9111(99)71402-0
- Benavides, F.G., Benach, J., Diez-Roux, A.V., & Roman, C. (2000). How do types of employment relate to health indicators? Findings from the Second European Survey on Working Conditions. *Journal of Epidemiology and Community Health*, 54(7), 494-501. doi:10.1136/jech.54.7.494
- Blasco de Luna, F.J., Guada, J., Sánchez-Robles, B., Blázquez Agudo, E.M., Checa Martín, J.L.,..., Suárez, A. (2018). *VII Informe Adecco sobre absentismo*. Madrid: Adecco. Retrieved from <https://www.adecco.institute.es/descarga-el-vii-informe-adecco-sobre-absentismo-laboral/>
- Boada i Grau, J., De Diego Vallejo, R., Agulló Tomás, E.A., & Mañas Rodríguez, M.M. (2005). Absenteeism from work as consequent of organizational variables. *Psicothema*, 17(2), 212-218.
- Bogaerts, Y., De Witte, H., Verlinden, H., & Vermeulen, V. (2017). Job insecurity and presenteeism: An indirect effect through burnout. *Gedrag en Organisatie*, 30(1), 3-24. Retrieved from <https://www.researchgate.net/publication/317758911>
- Campos-Serna, J., Ronda-Pérez, E., Artázcoz, L., & Benavides, F.G. (2012). Desigualdades de género en salud laboral en España. *Gaceta Sanitaria*, 26(4), 343-351. doi:10.1016/j.gaceta.2011.09.025
- Casey, T.W., Riseborough, K.M., & Krauss, A.D. (2015). Do you see what I see? Effects of national culture on employees' safety-related perceptions and behavior. *Accident Analysis and Prevention*, 78, 173-184. doi:10.1016/j.aap.2015.03.010
- Chien, C., Wang, W., & Cheng, J. (2007). Data mining for yield enhancement in semiconductor manufacturing and an empirical study. *Expert Systems with Applications*, 33(1), 192-198. doi:10.1016/j.eswa.2006.04.014
- Chun Fu, L., Yu-chu, Y., Yu, H.H., & Ray, I.C. (2013). Data mining for providing a personalized learning path in creativity: An application of decision trees. *Computers & Education*, 68, 199-210. doi:10.1016/j.compedu.2013.05.009
- de Menezes, T.A., & do Monte, P.A. (2013). Relationship between the type of contract work and dedication of workers in employment: Evidence for metropolitan regions. *Nova Economia*, 23(1), 51-72.
- Estryn-Béhar, M., & Van Der Heijden, B.I.J.M. (2012). Effects of extended work shifts on employee fatigue, health, satisfaction, work/family balance, and patient safety. *Work*, 41(SUPPL.1), 4283-4290. doi:10.3233/WOR-2012-0724-4283
- Fernández-García, R. (2013). ¿Por qué los jóvenes se accidentan tanto? *Gestión práctica de riesgos laborales: Integración y desarrollo de la gestión de la prevención*, 101, 42-47.
- Fleiss, J.L., Spitzer, R.L., Cohen, J., & Endicott, J. (1972). Three computer diagnosis methods compared. *Archives of General Psychiatry*, 27(5), 643-649. doi:10.1001/archpsyc.1972.01750290057011
- Gallis, C. (2013). Increasing productivity and controlling of work fatigue in forest operations by using prescribed active pauses: A selective review. *Croatian Journal of Forest Engineering*, 34(1), 103-112. Retrieved from <https://hrcak.srce.hr/116733>
- García Mainar, I., Green, C.P., & Navarro Paniagua, M. (2018). The effect of permanent employment on absenteeism: Evidence from labor reform in Spain. *ILR Review*, 71(2), 525-549. doi:10.1177/0019793917717226
- Hassard, J., Teoh, K.R.H., Visockaite, G., Dewe, P., & Cox, T. (2018). The cost of work-related stress to society: A systematic review. *Journal of Occupational Health Psychology*, 23(1), 1-17. doi:10.1037/ocp0000069
- Jiang, L., & Lavayssé, L. M. (2018). Cognitive and affective job insecurity: A meta-analysis and a primary study. *Journal of Management*, 44(6), 2307-2342. doi:10.1177/0149206318773853
- Kass, G.V. (1980). An exploratory technique for investigating large quantities of categorical data. *Applied Statistics*, 29(2), 119-127. Retrieved from <https://www.jstor.org/stable/2986296?seq=1>
- Khashaba, E., El-Helaly, M., El-Gilany, A. H., Motawei, S.M., & Foda, S. (2018). Risk factors for non-fatal occupational injuries among construction workers: A case-control study. *Toxicology and Industrial Health*, 34(2), 83-90. doi:10.1177/0748233717733853
- López-Alonso, M., Ibarrondo-Dávila, M.C., & Rubio-Gámez, C. (2015). Analysis of the cost of prevention in construction sites. *Informes de la Construcción*, 67(537). doi:10.3989/ic.13.062
- López-Goñi, J.J., Fernández-Montalvo, J., & Arteaga, A. (2012). Predictive validity of the EuropAsi: Clinical diagnosis or composite scoring? *Journal of Substance Abuse Treatment*, 42(4), 392-399. Retrieved from [https://academic.elsevier.com/locate/S0893-9782\(12\)00000-0](https://academic.elsevier.com/locate/S0893-9782(12)00000-0)
- Luján, L.M.B., Santana, N.R., Alemán, S.T., & Sánchez, J.A.S. (2014). Consumption of alcohol in canary islands: Trends and the most important sociodemographic factors associated (1997-2009). *Revista Española de Nutrición Comunitaria*, 20(2), 57-66. doi:10.7400/RENC.2014.01.2.5012
- Martín-Fumadó, C., Martí Amengual, G., Puig Bausilí, L., & Arimany-Manso, J. (2014). Temporary disability and its legal implications. *Medicina Clínica*, 142(SUPPL. 2), 37-42. doi:10.1016/S0025-7753(14)70070-3
- Martín-Román, A., & Moral, A. (2017). A methodological proposal to evaluate the cost of duration moral hazard in workplace accident insurance. *European Journal of Health Economics*, 18(9), 1181-1198. doi:10.1007/s10198-017-0878-6
- Mazzoni, C.F., Campos, M.V.G., Sirqueira, C.A.G., Krause, F., de Kraker, H., Könemann, R., & Douwes, M. (2019). Developing a methodology for a participatory ergonomics evaluation process: Human performance and productivity cycle. *Proceedings of the 20th Congress of the International Ergonomics Association*. IEA 2018. Advances in Intelligent Systems and Computing, vol 824. Springer, Cham. doi:10.1007/978-3-319-96071-5_190
- Moreno, N., Moncada, S., Llorens, C., & Carrasquer, P. (2010). Double presence, paid work, and domestic-family work. *New solutions: A journal of environmental and occupational health policy* : NS, 20(4), 511-526. doi:10.2190/NS.20.4.h
- Ortega, J.L. (2015). Diferencias y evolución del impacto académico en los perfiles de Google Scholar Citations: una aplicación de árboles de decisión. *Revista Española de Documentación Científica*, 38(4), e102. doi:10.3989/redc.2015.4.1225
- Peñalva-Vélez, A., & López-Goñi, J.J. (2014). Competencias ciudadanas en el alumnado de magisterio: la competencia intercultural personal. *Revista Electrónica Interuniversitaria de Formación del Profesorado (REIFOP)*, 17(2), 139-153. Retrieved from <https://www.redalyc.org/pdf/2170/217031054011.pdf>
- Restrepo, C., & Salgado, E. (2013). Types of contracts and worker absenteeism in Colombia. *Journal of Business Research*, 66(3), 401-408. doi:10.1016/j.jbusres.2012.04.006
- Saito, T., & Rehmsmeier, M. (2015). The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLoS ONE*, 10(3), e0118432. doi:10.1371/journal.pone.0118432
- Sanz, E., & Ponce de León, A. (2010). Claves en la aplicación del algoritmo Chaid. Un estudio del ocio físico deportivo universitario. *Revista de Psicología del Deporte*, 19(2), 319-333. Retrieved from <https://www.redalyc.org/pdf/2351/235116352011.pdf>
- Thurston, E., & Glendon, A. I. (2018). Association of risk exposure, organizational identification, and empowerment, with safety participation, intention to quit, and absenteeism. *Safety Science*, 105, 212-221. doi:10.1016/j.ssci.2018.02.012
- Vercruyssen, A., & Van de Putte, B. (2013). Work-family conflict and stress: indications of the distinctiveness of role combination stress for Belgian working mothers. *Community, Work and Family*, 16(4), 351-371. doi:10.1080/13668803.2013.776515
- Virtanen, M., Kivimäki, M., Joensuu, M., Virtanen, P., Elovainio, M., & Vahtera, J. (2005). Temporary employment and health: A review. *International Journal of Epidemiology*, 34(3), 610-622. doi:10.1093/ije/dyi024
- Zaballa, E., Martínez, J. M., Duran, X., Albertí, C., Gimeno Ruiz de Porras, D., & Benavides, F.G. (2016). Incidence of sickness absence by type of employment contract: One year follow-up study in Spanish salaried

workers. *Archives of Public Health*, 74(1). doi:10.1186/s13690-016-0152-x.
Taken from

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5037859/>